VLSI DESIGN

UNIT – I

INTRODUCTION: Introduction to IC Technology – MOS, PMOS, NMOS, CMOS & BiCMOS technologies.

BASIC ELECTRICAL PROPERTIES : Basic Electrical Properties of MOS and BiCMOS Circuits : Ids-Vds relationships, MOS transistor threshold Voltage, gm, gds, figure of merit ωo ; Pass transistor, NMOS Inverter, Various pull ups, CMOS Inverter analysis and design, Bi-CMOS Inverters.

VIDYA SAGAR P

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. 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.

The expansion of VLSI is 'Very-Large-Scale-Integration'. Here, the term 'Integration' refers to the complexity of the Integrated circuitry (IC). An IC is a well-packaged electronic circuit on a small piece of single crystal silicon measuring few mms by few mms, comprising active devices, passive devices and their interconnections. The technology of making ICs is known as 'MICROELECTRONICS'. This is because the size of the devices will be in the range of micro, sub micrometers. The examples include basic gates to microprocessors, op-amps to consumer electronic ICs. There is so much evolution taken place in the field of Microelectronics, that the IC industry has the expertise of fabricating an IC successfully with more than 100 million MOS transistors as of today. ICs are classified keeping many parameters in mind. Based on the transistors count on the IC, ICs are classified as SSI, MSI, LSI and VLSI. The minimum number of transistors on a VLSI IC is in excess of 40,000.

The concept of IC was conceived and demonstrated by JACK KILBY of TEXAS INSTRUMENTS at Dallas of USA in the year 1958.The silicon IC industry has not looked back since then. A lot of evolution has taken place in the industry and VLSI is the result of this. This technology has become the backbone of all the other industries. We will see every other field of science and technology getting benefit out of this.

INTRODUCTION TO IC TECHNOLOGY:

Over the last two decades electronics industry has achieved remarkable growth, mainly due to the advent of **Very-large-scale integration** (**VLSI**). VLSI is the process of creating an integrated circuit (IC) by combining thousands of transistors into a single chip. The number of applications of IC' s is in high performance computing, telecommunications, consumer electronics etc. The required computational power (or the intelligence) of these applications is the driving force the fast development of this field.

As more and more complex functions are required in various data processing and telecommunications devices, the need to integrate these functions in a small system/package is also increasing.

The levels of integration are measured by the no. of logic gates in a monolithic chip. Table 1.1 shows evaluation of logic complexity in integrated circuits

Name	Year	Transistors number	Logic gates number
small-scale integration (SSI)	1964	1 to 10	1 to 12
medium-scale integration (MSI)	1968	10 to 500	13 to 99
large-scale integration (LSI)	1971	500 to 20,000	100 to 9,999
very large scale integration (VLSI)	1980	20,000 to 1,000,000	10,000 to 99,999
Ultra large scale integration (ULSI)	1984	1,000,000 and more	100,000 and more

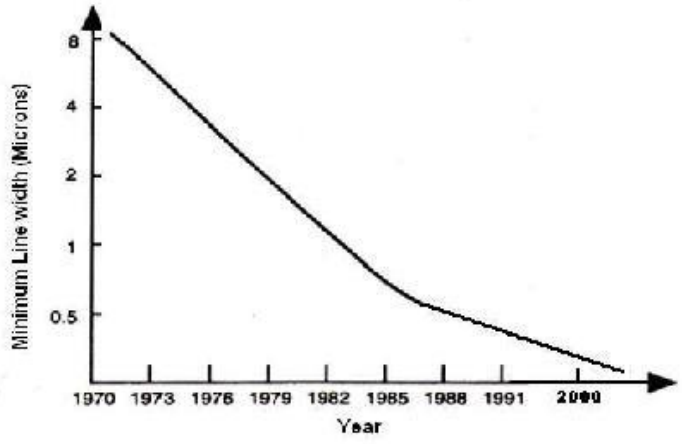
Table 1.1: Evolution of IC's

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 an 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.

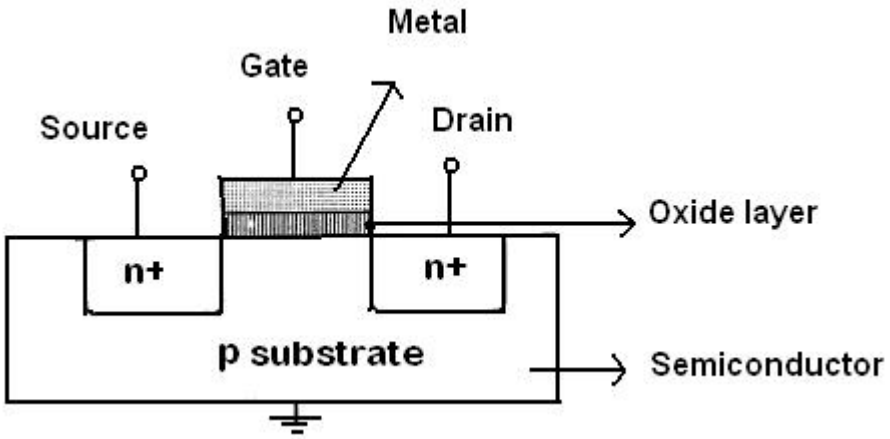


Figure 1.1. Basic MOS structure

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 VD = Vs = Vgs = 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.

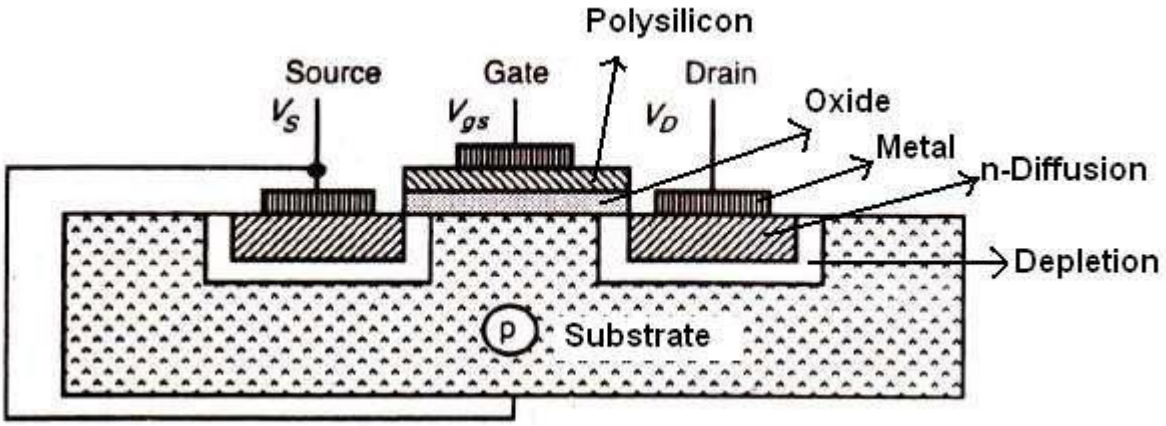


Figure 1.2. Enhancement mode transistor

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 (Vt) must be established between gate and source. Fig1.3(a) Shows the existing situation where a channel is established but no current flowing between source and drain (Vds = 0).

Let us now consider the conditions when current flows in the channel by applying a voltage Vds between drain and source. The IR drop = Vds along the channel. This develops a voltage between gate and channel varying with distance along the channel with the voltage being a maximum of Vgs at the source end. Since the effective gate voltage is Vg= Vgs - Vt, (no current flows when Vgs < Vt) there will be voltage available to invert the channel at the drain end so long as Vgs - Vt ~ Vds· The limiting condition comes when Vds= Vgs - Vt.

For all voltages Vds<Vgs - Vt, the device is in the non-saturated region of operation which is the condition shown in Fig1.3(b) Below.

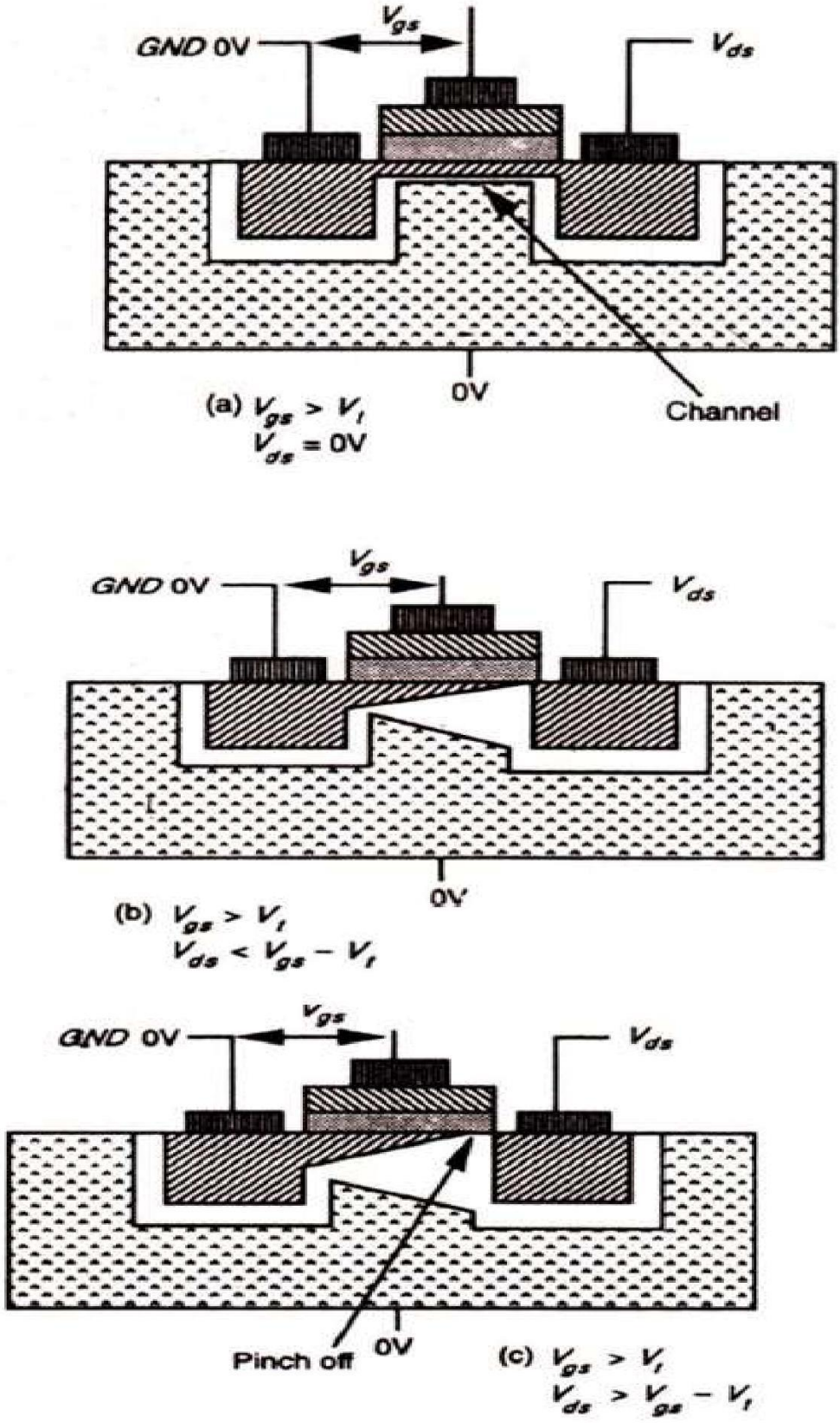


Figure 1.3. Enhancement mode transistor

Let us now consider the situation when Vds is increased to a level greater than Vgs - Vt. In this case, an IR drop equal to Vgs – Vt 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 Vds above Vds = Vgs - Vt. In all cases, the channel will cease to exist and no current will flow when Vgs < Vt. Typically, for enhancement mode devices, Vt = 1 volt for VDD = 5 V or, in general terms, Vt = 0.2 VDD.

DEPLETION MODE TSANSISTOR 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. -Vt for an N channel transistor, or +Vt for an P-channel transistor). The actual concentration of ions in the substrate-to-channel region is used to adjust the threshold voltage (Vt) 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 Vgs = 0, and to cause the channel to cease a negative voltage Vtd must be applied between gate and source.

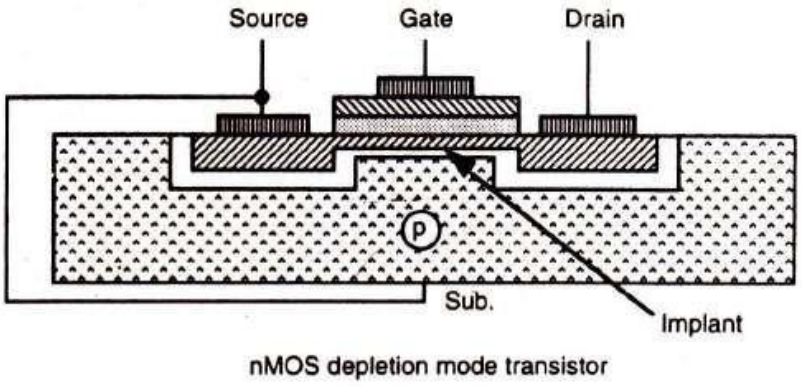


Figure 1.4. Depletion mode transistor

Vtd is typically < - 0.8 VDD, depending on the implant and substrate bias, but, threshold voltage differences apart, the action is similar to that of the enhancement mode transistor.

DRAIN-TO-SOURCE CURRENT IDS VERSUS VOLTAGE VDS 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 Vds applied between drain and source. Since the charge induced is dependent on the gate to source voltage Vgs then Ids is dependent on both Vgs and Vds.

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

$$Ids = -Isd = \frac{\text{Charge induced in channel (Qc)}}{\text{Electron transit time}(\tau)}$$

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

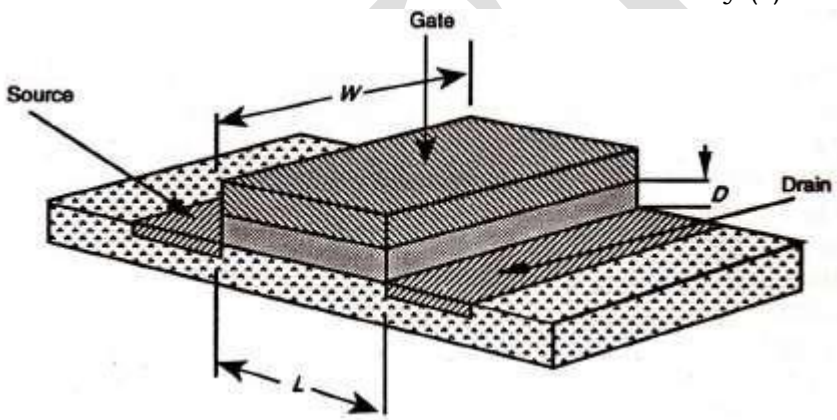


Figure 1.5. nMOS transistor structure

But velocity v= μEds

Where μ =electron or hole mobility and also ,**Eds = Vds/L**

so, v = μ.Vds/L and **τds = L^2 / μ.Vds**

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

$\mu_n \doteq 650\ cm^2/V\ sec$ (surface)
$\mu_p \doteq 240\ cm^2/V\ sec$ (surface)

The Non-saturated Region :

Let us consider the Id vs Vd relationships in the non-saturated region .The charge induced in the channel due to the voltage difference between the gate and the channel, Vgs (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 Vds/2.

Also, the effective gate voltage **Vg = Vgs – Vt** where Vt is the threshold voltage needed to invert the charge under the gate and establish the channel.

Where Eg = average electric field gate to channel

εins = relative permittivity of insulation between gate and channel (=**4** For silicon oxide)

εo = permittivity of free space (=**$8.85X10^{-14}$ F/Cm**).

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

$$Q_c = \frac{WL\varepsilon_{ins}\varepsilon_0}{D}\left((V_{gs} - V_t) - \frac{V_{ds}}{2}\right)$$

So, by combining the two equations ***Qc*** and τ_{sd} ,we get

$$I_{ds} = \frac{\varepsilon_{ins}\varepsilon_0\mu}{D}\frac{W}{L}\left((V_{gs} - V_t) - \frac{V_{ds}}{2}\right)V_{ds}$$

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 Vds < Vgs – Vt and

$$K = \frac{\varepsilon_{ins}\varepsilon_0\mu}{D}$$

Generally, a constant β is defined as

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

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

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

$$C_g = \frac{\varepsilon_{ins}\varepsilon_0 WL}{D} \text{ (parallel plate)}$$

- The capacitance formed by gate and channel has a parallel plate geometry.

$$C_g = \frac{\varepsilon_{ins}\varepsilon_0 WL}{D}$$

Then in terms of C_g.

$$K = \frac{\mu \cdot C_g}{WL}$$

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

- Gate capacitance per unit area C_0 or C_{ox} is defined as :

$$C_0 = \frac{C_g}{WL}$$

Therefore I_{ds} can be written as -

$$I_{ds} = C_0 \mu \frac{W}{L}\left[(V_{gs} - V_t)V_{ds} - \frac{V_{ds}^2}{2}\right]$$

The Saturated Region

Saturation begins when Vds = Vgs - V, 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 Vds 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

$$I_{ds} = C_0 \mu \frac{W}{2L}(V_{gs} - V_t)^2$$

The expressions derived above for Ids hold for both enhancement and depletion mode devices. Here the threshold voltage for the nMOS depletion mode device (denoted as Vtd) is negative.

- Typical characteristics nMOS transistors are shown in Fig. pMOS transistor characteristics are similar with suitable reversal of polarity.
- Following expression summarizes current in three regions.

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{Cut - off} \\ \beta\left(-V_{gs} - V_t - \frac{V_{ds}}{2}\right)V_{ds} & V_{ds} < V_{dsat} \quad \text{Linear} \\ \frac{\beta}{2}(V_{gs} - V_t)^2 & V_{ds} > V_{dsat} \quad \text{Saturation} \end{cases}$$

MOS Transistor Threshold Voltage Vt:

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 Vt may be expressed as: $V_t = \phi_{ms} \frac{Q_B - Q_{SS}}{C_0} + 2\phi_{fN}$

where Q_B = the charge per unit area in the depletion layer below the oxide

Qss = charge density at Si: SiO2 interface

Co =Capacitance per unit area.

Φns = work function difference between gate and Si

$Q_B = \sqrt{2\varepsilon_0 \varepsilon_{Si} qN(2\phi_{fN} + V_{SB})}$ coulomb/m^2

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

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

ΦfN = Fermi level potential between inverted surface and bulk Si

For polynomial gate and silicon substrate, the value of Φns is negative but negligible and the magnitude and sign of Vt are thus determined by balancing the other terms in the equation. To evaluate the Vt the other terms are determined as below.

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 (Vsb) 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.

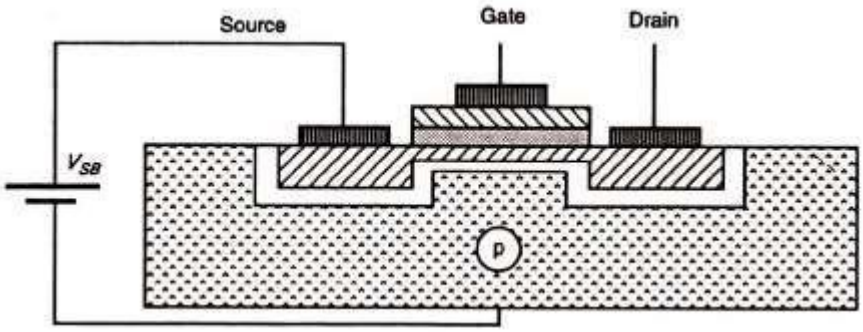


Figure 1.6 Body effect - nMOS device

Increasing Vsb causes the channel to be depleted of charge carriers and thus the threshold voltage is raised. Change in Vt is given by $\Delta Vt = \gamma.(Vsb)^{1/2}$ 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 $V_t = V_t(0) + \left(\frac{D}{\varepsilon_{ins}\varepsilon_0}\right)\sqrt{2\varepsilon_0\varepsilon_{Si}QN.(V_{SB})^{1/2}}$

Where Vt(0) is the threshold voltage for Vsd = 0

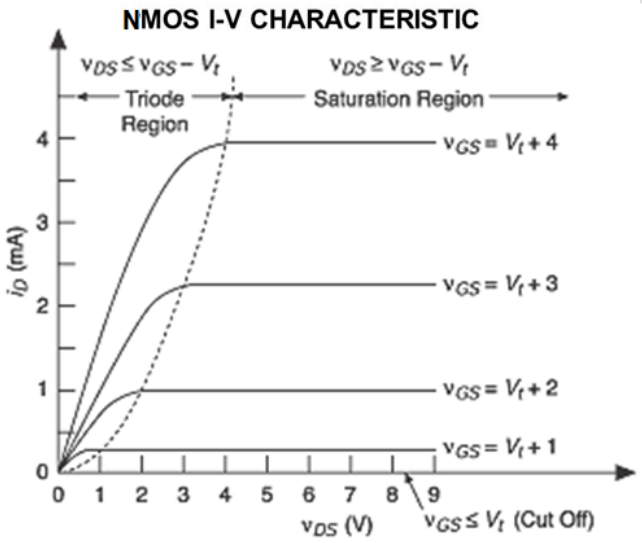
For n-MOS depletion mode transistors, the body voltage values at different VDD voltages are given below.

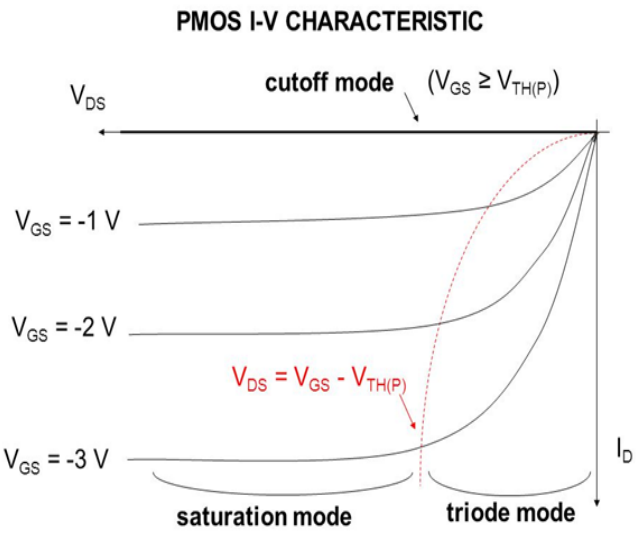
VSB = 0 V ; Vsd = -0.7VDD (= - 3.5 V for VDD =+5V)

VSB = 5 V ; Vsd = -0.6VDD (= - 3.0 V for VDD =+5V)

MOSFET I-V Characteristics:
Summary of Analytical Equations

nMOS		
Mode	I_D	**Voltage Range**
Cut-off	0	$V_{GS} < V_T$
Linear	$(\mu_n C_{ox}/2)(W/L)[2(V_{GS}-V_T)V_{DS}-V_{DS}^2]$	$V_{GS} \geq V_T,\ V_{DS} < V_{GS}-V_T$
Saturation	$(\mu_n C_{ox}/2)(W/L)(V_{GS}-V_T)^2(1+\lambda V_{DS})$	$V_{GS} \geq V_T,\ V_{DS} \geq V_{GS}-V_T$
pMOS		
Cut-off	0	$V_{GS} > V_T$
Linear	$(\mu_n C_{ox}/2)(W/L)[2(V_{GS}-V_T)V_{DS}-V_{DS}^2]$	$V_{GS} \leq V_T,\ V_{DS} > V_{GS}-V_T$
Saturation	$(\mu_n C_{ox}/2)(W/L)(V_{GS}-V_T)^2(1+\lambda V_{DS})$	$V_{GS} \leq V_T,\ V_{DS} \leq V_{GS}-V_T$





TRANS CONDUCTANCE (g_m):

$g_{ds}=\partial I_{DS}/\partial V_{DS}\sim\beta(Vgs-Vt)=\lambda$

- Transconductance shows relationship between output current I_{ds} and input voltage V_{gs} and is given as -

$$g_m = \left.\frac{\delta I_{ds}}{\delta V_{gs}}\right|_{V_{ds}=\text{constant}}$$

- By definition of current, the charge in channel Q_c is such that,

$$\tau = \frac{Q_c}{I_{ds}}$$

where,

τ is transit time for crossing the channel.

- Now change in current is,

$$\delta I_{ds} = \frac{\delta Q_c}{\tau_{ds}}$$

but $$\tau_{ds} = \frac{L^2}{\mu V_{ds}}$$

Therefore,

$$\delta I_{ds} = \frac{\delta Q_c V_{ds}\mu}{L^2}$$

Since $Q = CV$, change in charge

$$\delta Q_c = C_g \delta V_{gs}$$

From which

$$\delta I_{ds} = \frac{C_g \delta V_{gs} \mu V_{ds}}{L^2}$$

- Transconductance is given by -

$$g_m = \frac{\delta I_{ds}}{\delta V_{gs}} = \frac{C_g \mu V_{ds}}{L^2}$$

Under saturation region, $V_{ds} = V_{gs} - V_t$

So that,

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

From equation (2.15),

$$C_{g-ch} = \frac{\varepsilon_{ins}\varepsilon_0 WL}{D}$$

$$\therefore \quad g_m = \frac{\mu\varepsilon_{ins}\varepsilon_0}{D}\frac{W}{L}(V_{gs} - V_t)$$

In terms of β

$$g_m = \beta (V_{gs} - V_t)$$

- Above expressions of g_m reveals that g_m of a MOS device can be increased by increasing its width. However, this will also increase the input capacitance and area occupied.
- The resistance of FET in linear region near the origin is known as on state resistance R_{on} and is given by,

$$R_{on} = [g_m]^{-1} = \frac{1}{\beta (V_{gs} - V_t)} \qquad \text{where, } \beta = K \cdot \frac{W}{L}$$

Output Conductance gds:

The output conductance g_{ds} can be expressed as - g_{ds} =

$$g_{ds} = \partial I_{ds} / \partial V_{ds} \quad \text{at } v_{gs} \text{ is constant.}$$

$$= \lambda \cdot I_{ds} = \left(\frac{1}{L}\right)^2$$

For the MOS device, strong dependence on the channel length demonstrated as -

$$\lambda \propto \frac{1}{L} \quad \text{and} \quad I_{ds} \propto \frac{1}{L}$$

MOS Transistor Figure of Merit (ω_0)

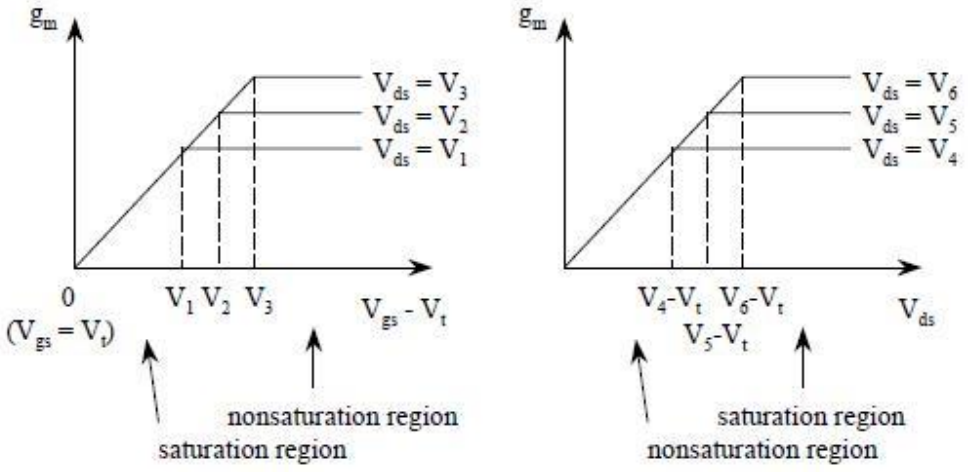
- Figure of merit is a measure for frequency response and switching performance of a MOS transistor. Figure of merit is defined as :

$$\omega_0 = \frac{g_m}{C_g}$$

By using equation

$$\omega_0 = \frac{\mu}{L^2}(V_{gs} - V_t) \qquad \omega_0 = \frac{1}{\tau_{sd}}$$

For devices in non-saturation region

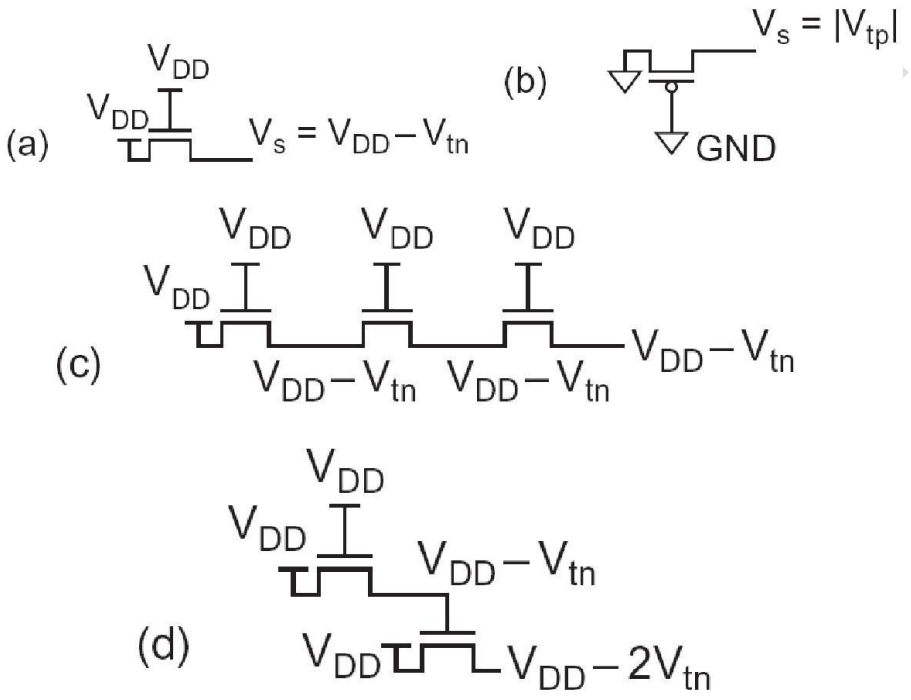


PASS TRANSISTOR:

The nMOS transistors pass '0's well but '1' s poorly. Figure(a) shows an nMOS transistor with the gate and drain tied to VDD. Imagine that the source is initially at Vs = 0.

Vgs > Vtn, so the transistor is ON and current flows. If the voltage on the source rises to Vs = VDD - Vtn, Vgs falls to Vtn and the transistor cuts itself OFF. Therefore, nMOS transistors attempting to pass a '1' never pull thesource above VDD - Vtn. This loss is sometimes called a **threshold drop**. Moreover, when the source of the nMOS transistor rises, Vsb becomes nonzero and this nonzero source to body potential introduces the body effect that increases the threshold voltage. Similarly, pMOS transistors pass '1`s well but '0's poorly. If the pMOS source drops belo w |Vtp|, the transistor cuts off. Hence, pMOS transistors only pull down to within a threshold above GND, as shown in Figure(b).

As the source can rise to within a threshold voltage of the gate, the output of several transistors in series is no more degraded than that of a single transistor (Figure (c)). However, if a degraded output drives the gate of another transistor, the second transistor can produce an even further degraded output (Figure d).



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 of connected from the drain to the positive supply rail VDD. 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. 1.7.

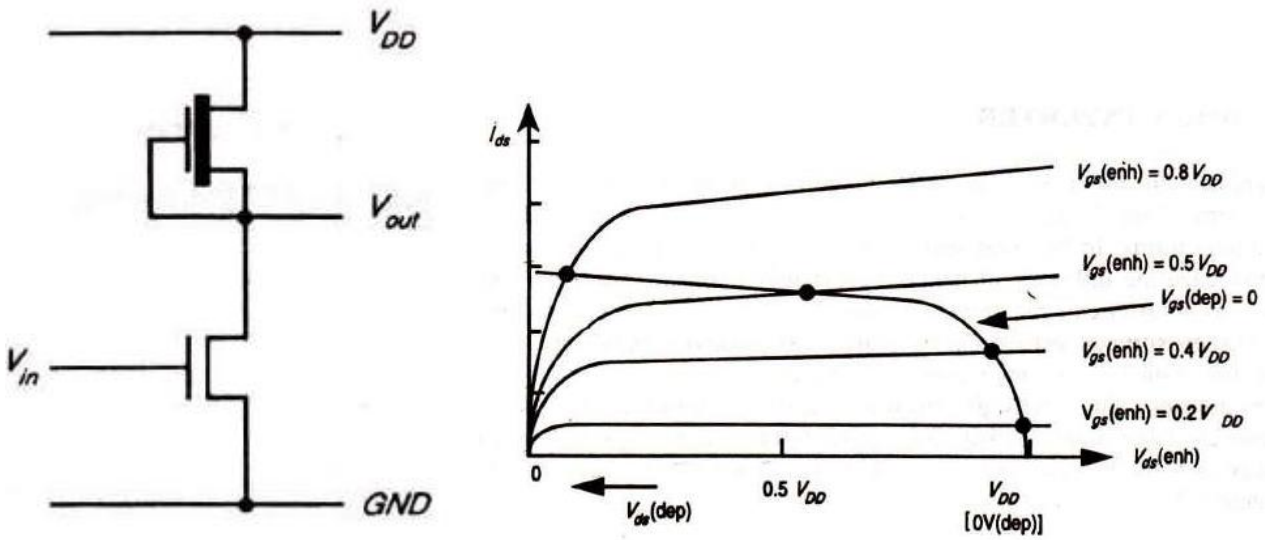


Figure 1.7 depletion mode transistor as load

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 Ids for both transistors must be equal.

nMOS Inverter transfer characteristic.

The transfer characteristic is drawn by taking Vds on x-axis and Ids on Y-axis for both enhancement and depletion mode transistors. So, to obtain the inverter transfer characteristic for

Vgs = 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 Vin(=Vgs p.d. transistor) exceeds the Pulldown threshold voltage current begins to flow. The output voltage Vout thus decreases and the subsequent increases in Vin will cause the Pull-down transistor to come out of saturation and become resistive.

CMOS Inverter Analysis:

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.

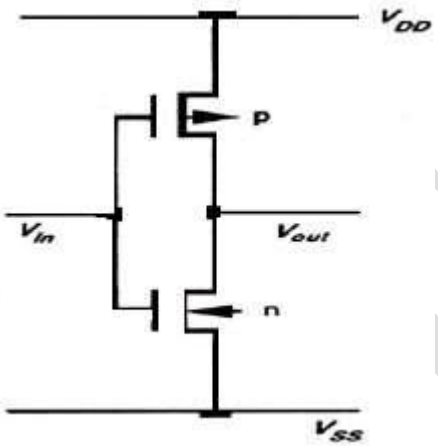
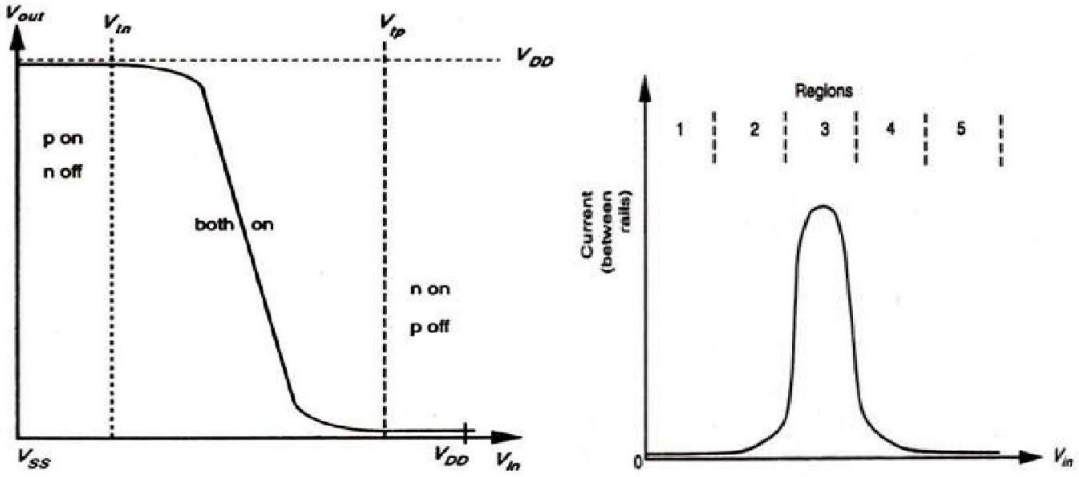


Figure 1.8 CMOS Inverter

In the inverter circuit Figure 1.8, 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 Vi= 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 VDD through the p-transistor.

Hence the output voltage is logic 1. In **region 5**, Vin = 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 VDD to VSS. 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

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

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

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

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 VDD and Vss 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 *(Zp.u}Zp.d.)*for an nMOS Inverter driven by another nMOS Inverter:

Let us consider the arrangement shown in Fig.1.9 (a). in which an inverter is driven from the output of another similar inverter. Consider the depletion mode transistor for which Vgs = 0 under all conditions, and also assume that in order to cascade inverters without degradation the condition

$$V_{in} = V_{out} = V_{inv}$$

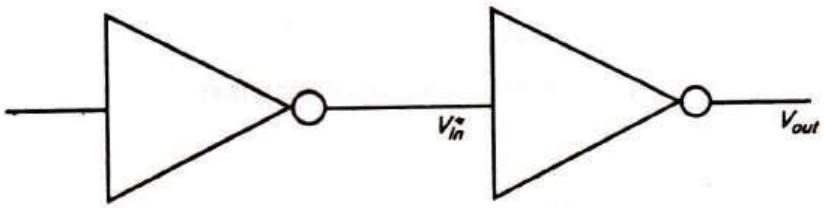


Fig 1.9.(a). Inverter driven by another inverter.

For equal margins around the inverter threshold, we set Vinv = 0.5VDD · 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

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

where Wp.d , Lp.d , Wp.u. and Lp.u are the widths and lengths of the pull-down and pull-up transistors respectively.

So,we can write that

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

Equating (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 Vt ,Vinv and Vtd 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

Here $\sqrt{Z_{p.u.}/Z_{p.d.}} = 2$ $\qquad 0.5 = 0.2 + \dfrac{0.6}{\sqrt{Z_{p.u.}/Z_{p.d.}}}$

So, we get

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

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 1.9(b) below (These transistors are called pass transistors).

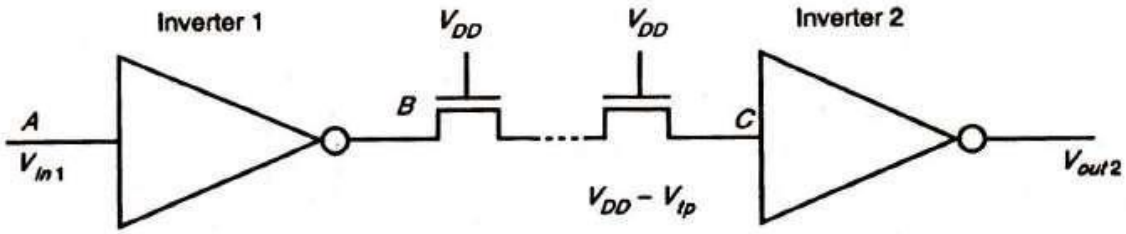


Fig 1.9.(b). Inverter driven by another inverter.

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 VDD. but the voltage into inverter2 at point C is now reduced from VDD by the threshold voltage of the series pass transistor. With all pass transistor gates connected to VDD there is a loss of Vtp, 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

Vin2 = VDD- Vtp

where Vtp = threshold voltage for a pass transistor.

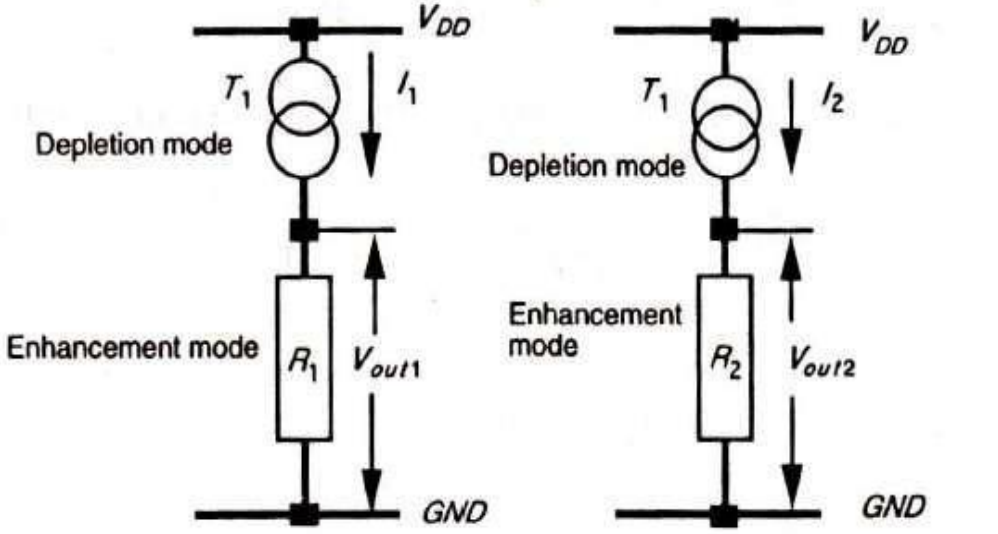
Let us consider the inverter 1 shown in Fig. (a) with input = VDD· If the input is at VDD, then the Pull-down transistor T2 is conducting but with a low voltage across it; therefore, it is in its resistive region represented by R1 in Fig. (a) below. Meanwhile, the pull up transistor T1 is in saturation and is represented as a current source.

For the pull down transistor

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

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

Since Vds is small, Vds/2 can be neglected in the above expression.



(a) Inverter 1 with input = V_{DD} (b) Inverter 2 with input = $V_{DD} - V_{tp}$

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 Vgs = 0

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

The product **I1R1 = Vout1**

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

So,

Let us now consider the inverter 2 Fig.b .when input = VDD- Vtp.

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

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

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 Vout1 = Vout2.

That is $I_1R_1=I_2R_2$

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

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

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 Zp.u/ Zpd. ratio of ≥ 4/1.

(ii).An inverter driven through one or more pass transistors should have a Zp.u./Zp.d ratio of ≥8/1

ALTERMTIVE 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 RL : 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.

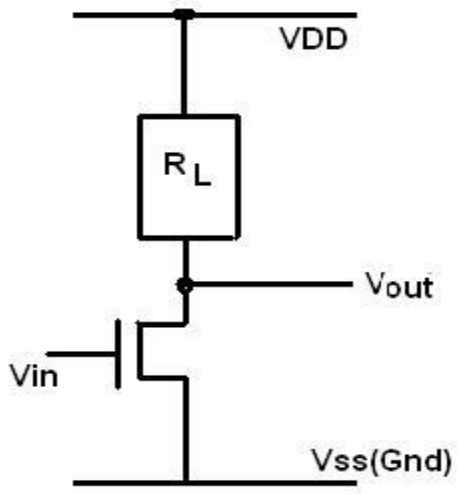


Fig 1.10(a). Inverter Load resistance RL.

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 Vin = logical 1.

(b) Switching of output from 1 to 0 begins when Vin exceeds Vt, of pull-down device.

(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.

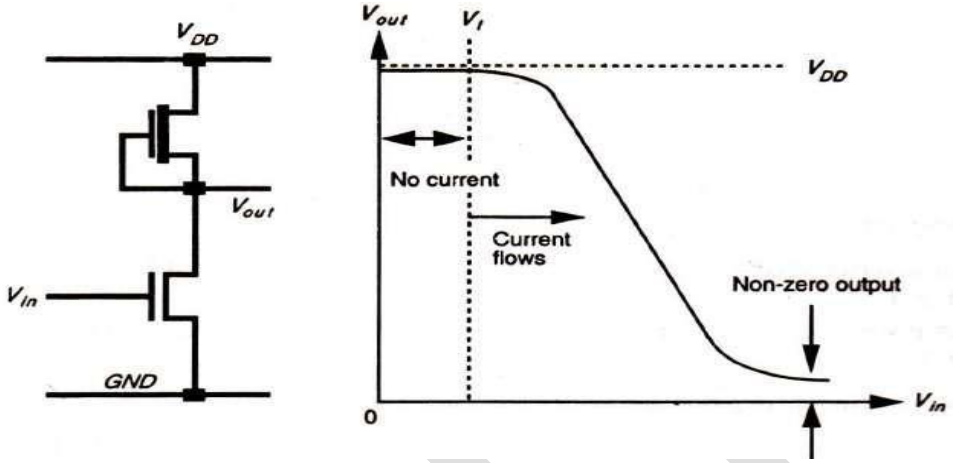


Fig 1.10(b)nMOS depletion mode transistor pull-up and transfer characteristic

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 Vin =logical 1 (VGG is returned to VDD).

(b) Vout can never reach VDD (logical I) if VGG = VDD as is normally the case.

(c) VGG may be derived from a switching source, for example, one phase of a clock, so that dissipation can be greatly reduced.

(d) If VGG is higher than VDD then an extra supply rail is required.

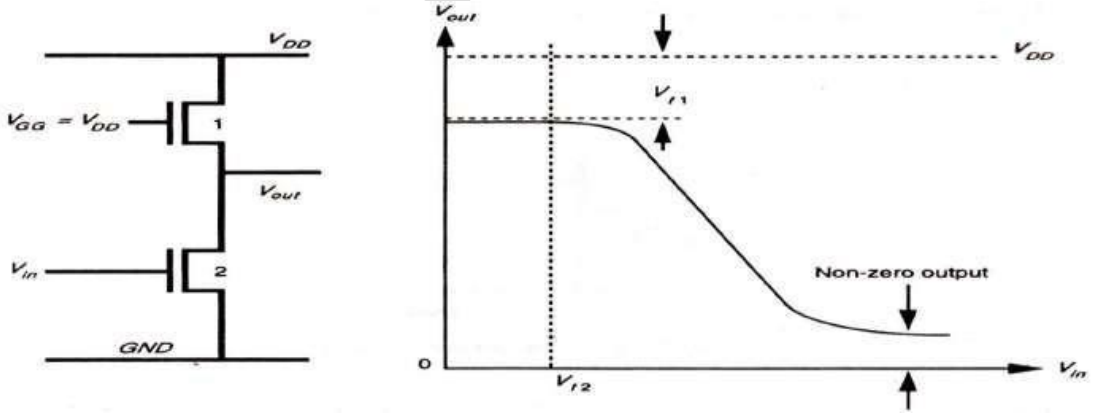


Fig 1.10(c) nMOS enhancement mode pull-up and transfer characteristic

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

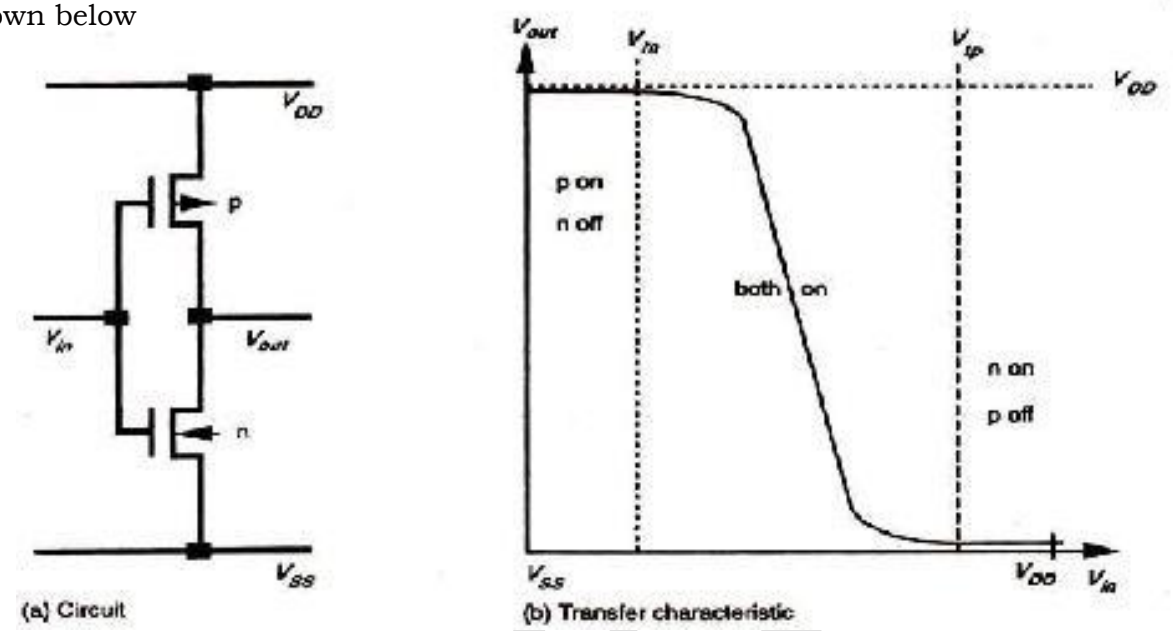


Fig 1.10(d) Complementary transistor pull-up and transfer characteristic

The salient features of this arrangement are

(a) No current flows either for logical 0 or for logical 1 input.

(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.

BiCMOS Gates:

➢ A known deficiency of MOS technology is its limited load driving capabilities (due to limited current sourcing and sinking abilities of pMOS and nMOS transistors.

➢ BiCMOS is a modified CMOS technology that includes bipolar junction transistors as circuit elements. In digital design, BiCMOS stages are used to drive high- capacitance lines more efficiently than MOSFET only circuits. BiCMOS Circuits employ CMOS logic circuits that are connected a bipolar output driver stage, as shown in fig.

The CMOS network provides logic operations and bipolar transistors are used to drive the output. Only one BJT is active at a time. BJT Q1 provides the high output voltage while Q1 discharges the output capacitance and gives the low output value.

- ➢ Bipolar transistors have
  - ❖ higher gain
  - ❖ better noise characteristics
  - ❖ better high frequency characteristics

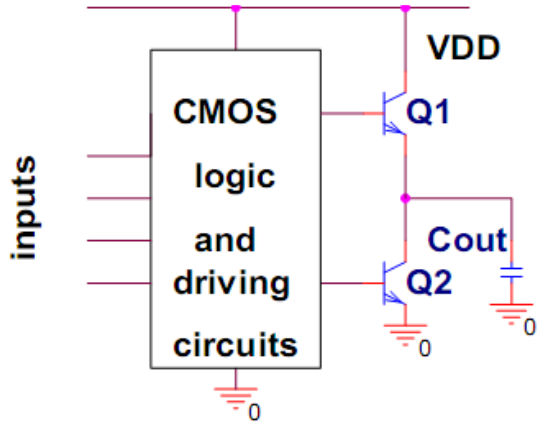


Fig 1.11 General form of a BiCMOS circuit

BiCMOS Inverter:

- The inverter circuit consists of two bipolar transistors T_1 and T_2, one nMOS transistor T_3 and one pMOS transistor T_4. Both the MOS devices are enhancement mode devices. The functioning of the circuit is as follows :

a) With V_{in} at logic 0 i.e. 0 volts (GND), T_3 is off which keeps T_1 non-conducting. However T_4 is on and supplies base current to T_2 which conducts and acts as a current source to charge the load C_L toward + 5 volts (V_{DD}). The output V_{out} goes to + 5 V less the base to emitter drop V_{BE} of T_2.

b) When V_{in} = logic 1 i.e. + 5 V (V_{DD}), T_4 is off so that T_2 will be non-conducting. But T_3 is on and supplies current to base of T_1 which conducts and acts as a current sink to the load C_L which discharges through it to 0 volts (GND). The V_{out} falls to 0 volts plus the saturation voltage V_{CEsat} between collector and emitter of T_1.

c) Charging and discharging of the load C_L is very fast because transistors T_1 and T_2 present low impedances when turned on into saturation.

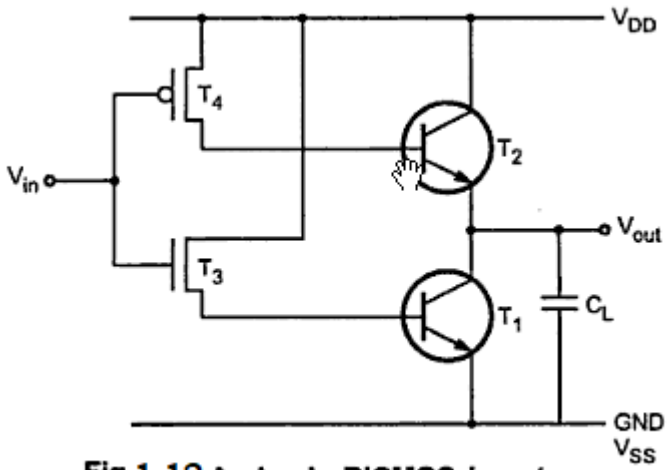


Fig.1.12 A simple BiCMOS inverter

- The output logic levels will approximate the rail voltages since V_{CEsat} is quite small and V_{BE} equals 0.7 volts approximately. The inverter offers a low output impedance and a high input impedance. It occupies a relatively small area but still has a high current drive capability. The inverter circuit has high noise margins.

- However there is a constant D.C. path between the rails through T_3 and T_1 which allows a significant static current flow whenever V_{in} = logic 1. This is not a desirable arrangement. Also, there is another problem, that there is no discharge path for current from the base of either npn transistor when it is being turned off. This adversely affects the speed of action of the circuit.

- The problem of the D.C. path through T_1 and T_3 is eliminated in an improved inverter circuit shown in Fig.**1.13**. However the output voltage swing gets reduced because the output cannot go below the base to emitter voltage V_{BE} of transistor T_1.

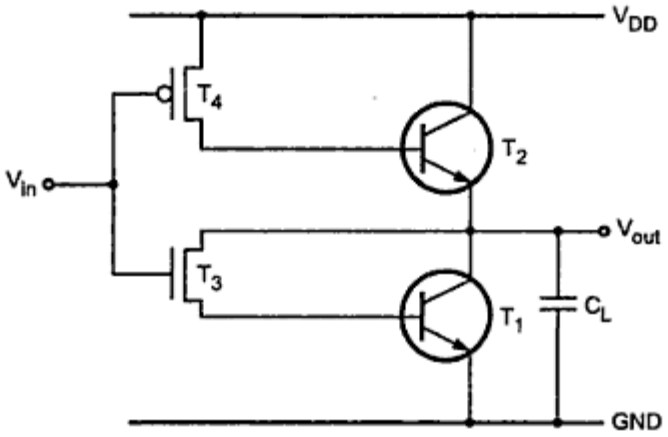


Fig.1.13 An alternative BiCMOS inverter with no static current flow

- A further improvement in inverter circuit can be achieved using resistors as shown in Fig.**1.14**. Here, the resistors provide an improved swing of output voltage when either bipolar transistor is off.

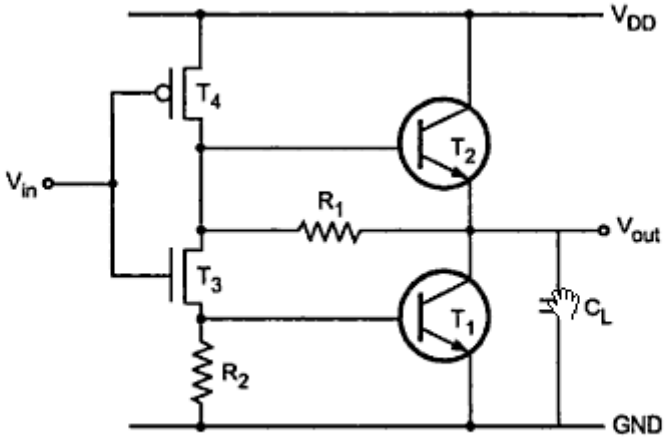


Fig1.14An improved BiCMOS inverter with better output logic levels

- They also provide discharge paths for the base currents during turn-off. However, fabricating resistors of suitable values in not always convenient and may occupy larger space. Hence other arrangements, like the one shown in Fig. are used.

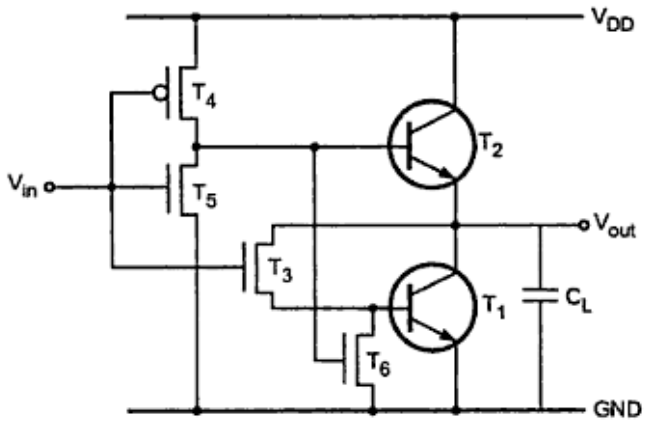


Fig1.15 An improved BiCMOS inverter using MOS transistors for base current discharge

- In the circuit shown in Fig. 1.15, arrangement is made to turn on transistors T_5 and T_6 when T_2 and T_1 respectively are being turned off. That is when T_2 is to be tuned off, T_5 gets turned on and provides discharge path for base current of T_2. Thus we observe that BiCMOS inverters are more suitable where high load current sinking and sourcing is required.

BICMOS NAND circuit: This is the two input BICMOS NAND gate.It consists of

1) 2 PMOS PA and PB
2) 4 NMOS, NA1, NB1, NA3, NB3
3) Two BJT's QP and Q0.

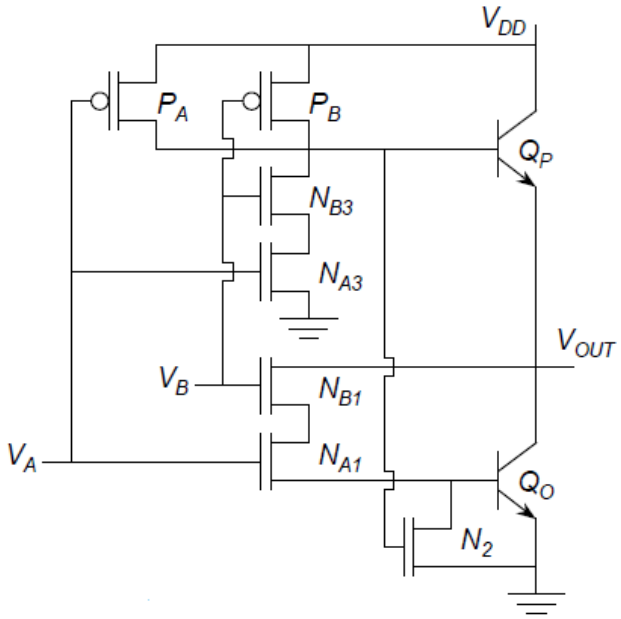


Fig 1.16 BICMOS NAND circuit:

Case 1: Both A and B are low: If both the inputs VA and VB(Refer the circuit) are low(0). PA and PB will be ON and the base of QA will be high. THus the top BJT(QA) will be ON which pulls the output UP. If there is a capacitive load. The output current will be almost 101 times the base current i.e it will be 101IB(Assuming beta of transistor is

Note: PA, PB, NA1 and NB1 are used for logical purposes (Just like in CMOS)

What is the purpose of N2, NB3 and NA3?

For high switching speeds of the BJT's, we need to remove the base charge from the transistor. In order to remove the charge from the base of the transistor we need a mechanism, that can be achieved by using these 3 NMOS.

Why to remove the Base charge anyway?

If we do not remove the base charge, the transistor will be in ON state and takes a lot of time to go to the OFF state. If we do not remove the base charge, the whole purpose of BICMOS is lost.

N2 MOS is getting input from the base of QA transistor. Since the base of QA is HIGH (say 5v), it will on N2 MOS. Since N2 is ON, it will PULL THE BASE CHARGE OUT OF Q0 transistor. So, if both the inputs are low, QA is ON and Q0 is OFF. OUTPUT IS HIGH and OUTPUT CURRENT IS LARGE. HIGH SWITCHING SPEED.

Case 2 : A and B are high,NB3 and NA3 NMOS make sure that QA is in OFF state. PA and PB are OFF. NB1, NA1 will be ON. The output is discharged via NB1, NA1 and Q0.(High Speed). N2 will be effectively out of circuit in this case. So, Q0 is ON and QP is OFF. You can correlate with the other 2 inputs.

Latch-up in CMOS circuits :

A byproduct of the Bulk CMOS structure is a pair of parasitic bipolar transistors. The collector of each BJT is connected to the base of the other transistor in a positive feedback structure.

A phenomenon called latch up can occur when

(1) both BJT's conduct, creating a low resistance path between Vdd and GND and

(2) the product of the gains of the two transistors in the feedback loop, b1 x b2, is greater than one. The result of latch up is at the minimum a circuit malfunction, and in the worst case, the destruction of the device.

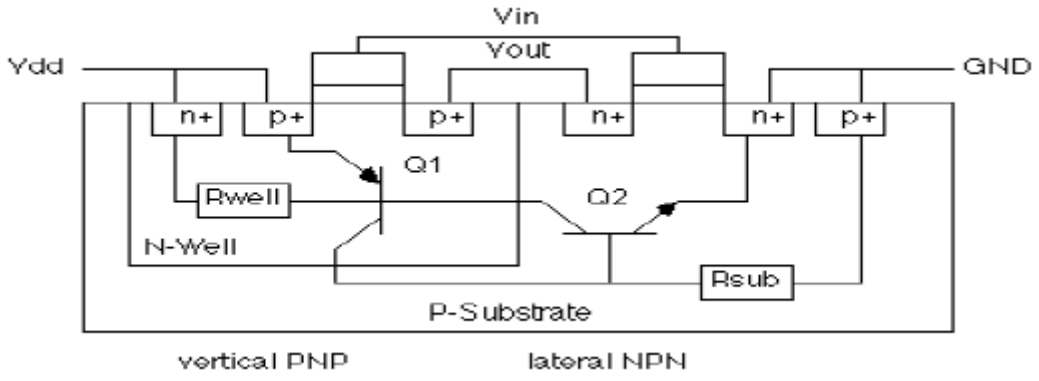


Fig 1.17(a) Cross section of parasitic transistors in Bulk CMOS

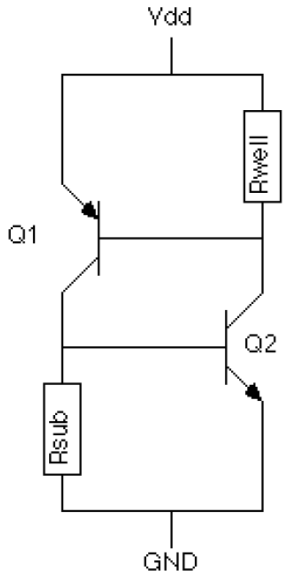


Fig 1.17(b) General form of a BiCMOS circuit

Latchup may begin when Vout drops below GND due to a noise spike or an improper circuit hookup (Vout is the base of the lateral NPN Q2). If sufficient current flows through Rsub to turn on Q2 (I Rsub > 0.7 V), this will draw current through Rwell. If the voltage drop across Rwell is high enough, Q1 will also turn on, and a self-sustaining low resistance path between the power rails is formed. If the gains are such that b1 x b2 > 1, latchup may occur. Once latchup has begun, the only way to stop it is to reduce the current below a critical level, usually by removing power from the circuit.

The most likely place for latch up to occur is in pad drivers, where large voltage transients and large currents are present.

Preventing latch up :

Fab/Design Approaches

1. Reduce the gain product b1 x b1

move n-well and n+ source/drain farther apart increases width of the base of Q2 and reduces gain beta2 > also reduces circuit density

buried n+ layer in well reduces gain of Q1

2. Reduce the well and substrate resistances, producing lower voltage drops

higher substrate doping level reduces Rsub

reduce Rwell by making low resistance contact to GND

guard rings around p- and/or n-well, with frequent contacts to the rings, reduces the parasitic resistances.

MOS TECHNOLOGY FABRICATION:

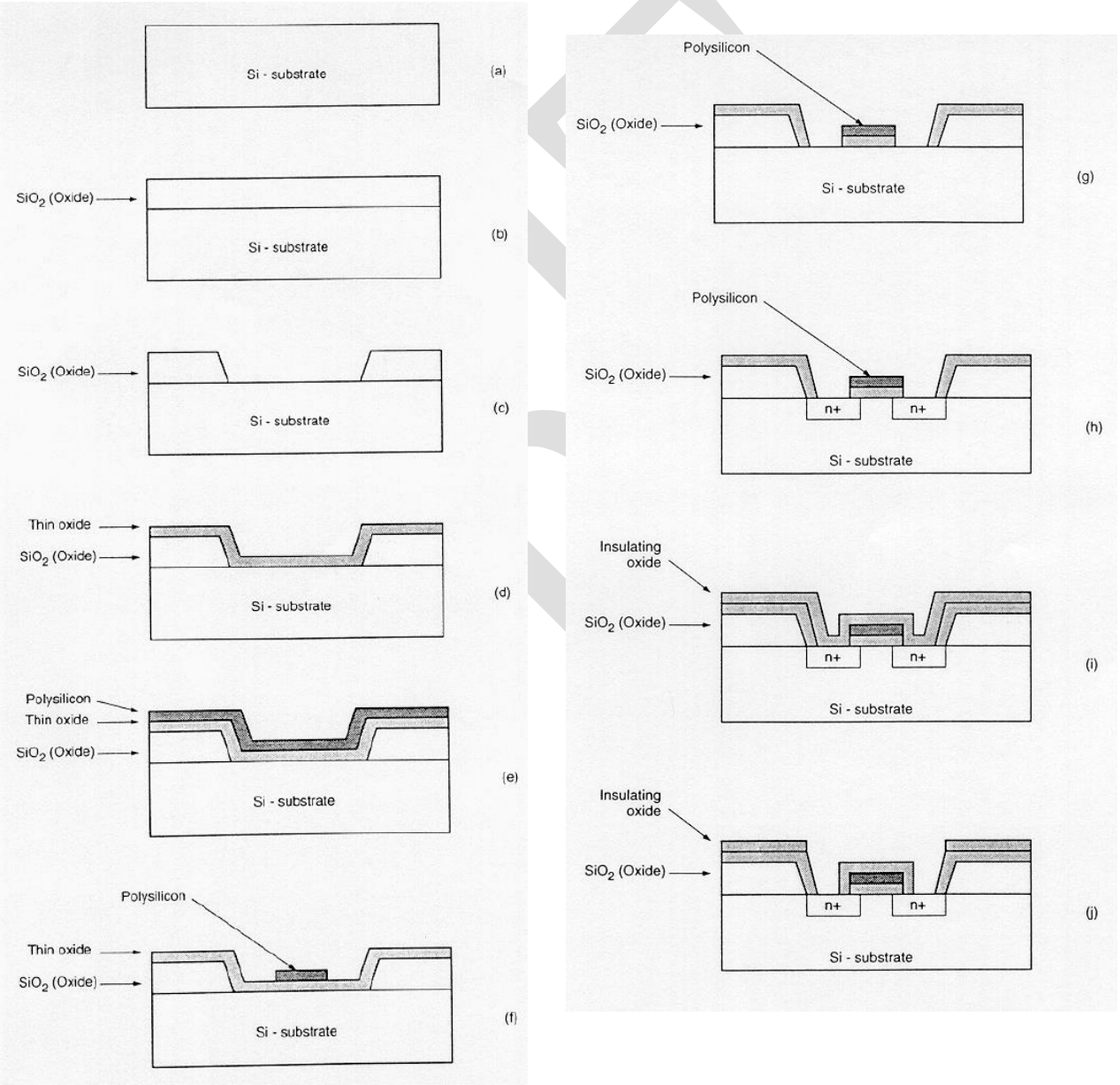
- nMOS Fabrication
- CMOS Fabrication
    - p-well process
    - n-well process
    - twin-tub process
- All the devices on the wafer are made at the same time
- After the circuitry has been placed on the chip
    - the chip is overglassed (with a passivation layer) to protect it
    - only those areas which connect to the outside world will be left uncovered (the pads)
- The wafer finally passes to a test station
    - test probes send test signal patterns to the chip and monitor the output of the chip
- The *yield* of a process is the percentage of die which pass this testing
- The wafer is then scribed and separated up into the individual chips. These are then packaged
- Chips are 'binned' according to their performance

nMOS Fabrication:

The process starts with the oxidation of the silicon substrate (Fig. 1.18(a)), in which a relatively thick silicon dioxide layer, also called field oxide, is created on the surface (Fig. 1.18(b)). Then, the field oxide is selectively etched to expose the silicon surface on which the MOS transistor will be created (Fig. 1.18(c)). Following this step, the surface is covered with a thin, high-quality oxide layer, which will eventually form the gate oxide of the MOS transistor (Fig. 1.18(d)).

On top of the thin oxide, a layer of polysilicon (polycrystalline silicon) is deposited (Fig. 1.18(e)). Polysilicon is used both as gate electrode material for MOS transistors and also as an interconnect medium in silicon integrated circuits. Undoped polysilicon has relatively high resistivity. The resistivity of polysilicon can be reduced, however, by doping it with impurity atoms. After deposition, the polysilicon layer is patterned and etched to form the interconnects and the MOS transistor gates (Fig. 1.18(f)). The thin gate oxide not covered by polysilicon is also etched away, which exposes the bare silicon surface on which the source and drain junctions are to be formed (Fig. 1.18(g)).

The entire silicon surface is then doped with a high concentration of impurities, either through diffusion or ion implantation (in this case with donor atoms to produce n-type doping). Figure 4(h) shows that the doping penetrates the exposed areas on the silicon surface, ultimately two n-type regions (source and drain junctions) in the p-type substrate. The impurity doping also penetrates the polysilicon on the surface, reducing its resistivity. Note that the polysilicon gate, which is patterned before doping actually defines the precise location of the channel region and, hence, the location of the source and the drain regions. Since this procedure allows very precise positioning of the two regions relative to the gate, it is also called the self-aligned process.



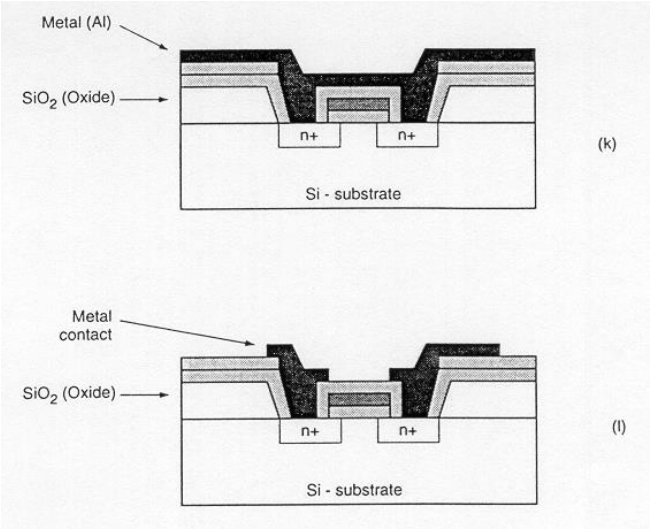


Figure-1.18: Process flow for the fabrication of an n-type MOSFET on p-type silicon.

Once the source and drain regions are completed, the entire surface is again covered with an insulating layer of silicon dioxide (Fig. 1.18(i)). The insulating oxide layer is then patterned in order to provide contact windows for the drain and source junctions (Fig. 1.18(j)). 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 (Fig. 1.18(l)). Usually, a second (and third) layer of metallic interconnect can also be added on top of this structure by creating another insulating oxide layer, cutting contact (via) holes, depositing, and patterning the metal.

CMOS TECHNOLOGY:

Fabrication

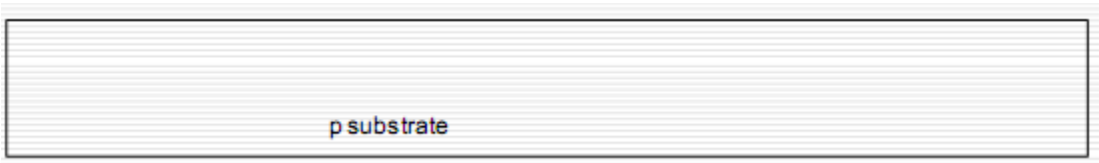
Fabrication involves the implementation of semiconductor processes to build a MOSFET transistor and compatible passive components as an integrated circuit.

N-Well CMOS Fabrication Major Steps

† Start with blank wafer † Build inverter from the bottom up

† First step will be to form the n-well

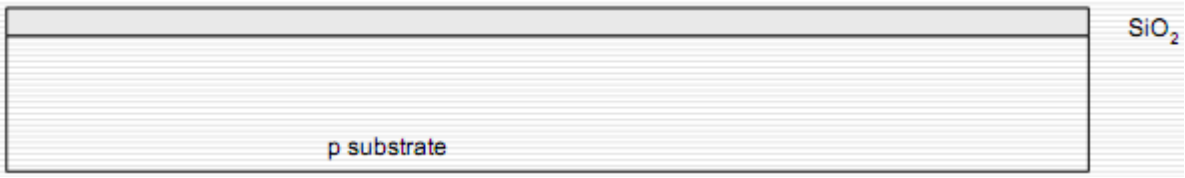
Cover wafer with protective layer of SiO2 (oxide), Remove layer where n-well should be built, Implant or diffuse n dopants into exposed wafer Strip off SiO2



Oxidation:

† Grow SiO2 on top of Si wafer 900 – 1200 C with H2O or O2 in oxidation furnace

† Now shape/pattern it for n-well



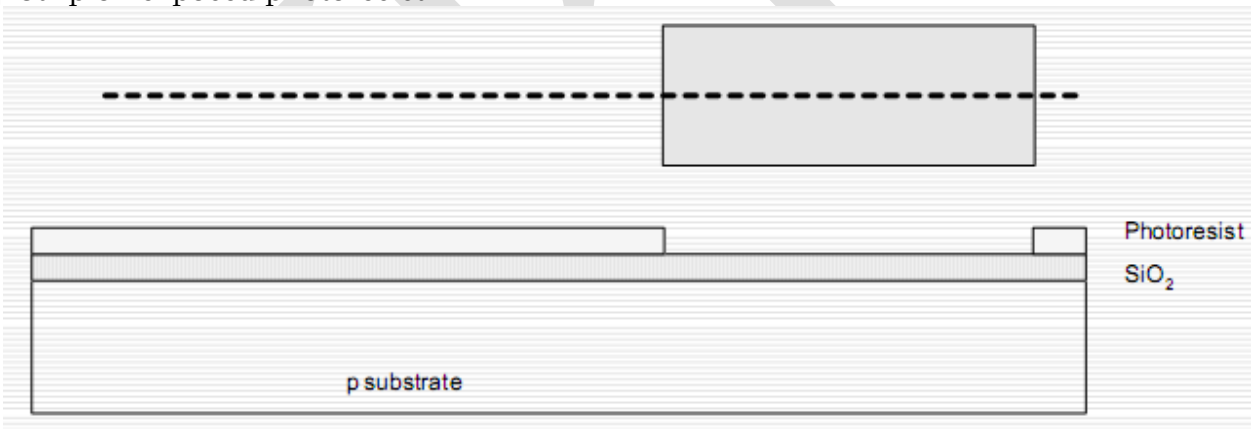
Photoresist:

† Spin on photoresist Photoresist is a light-sensitive organic polymer Softens where exposed to light

Lithography:

† Expose photoresist through n-well mask

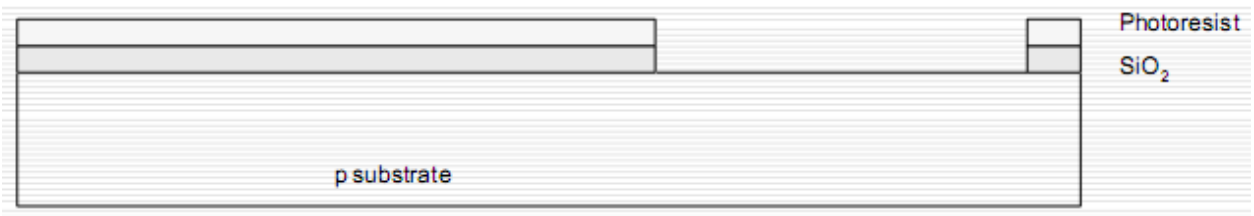
† Strip off exposed photoresist



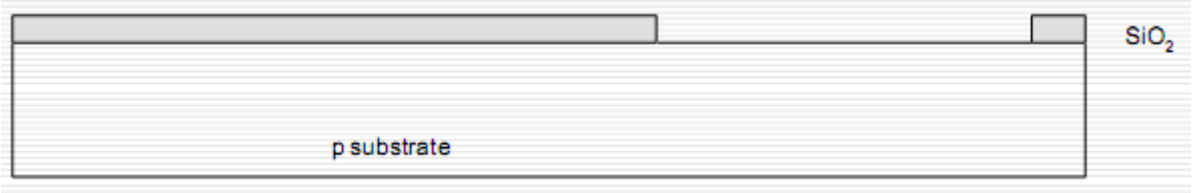
Etch:

† Etch oxide with hydrofluoric acid (HF)

† Only attacks oxide where resist has been Exposed



† Strip off remaining photoresist Use mixture of acids called piranha etch

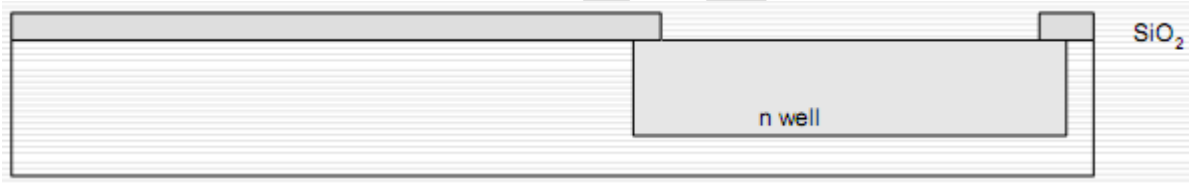


n-well:

† n-well is formed with diffusion or ion implantation

† Diffusion: Place wafer in furnace with arsenic gas Heat until as atoms diffuse into exposed Si

† Ion Implantation: Blast wafer with beam of as ions blocked by SiO, only enter exposed Si

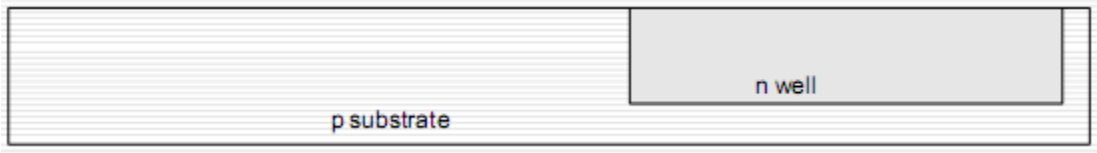


Strip Oxide:

† Strip off the remaining oxide using HF

† Back to bare wafer with n-well

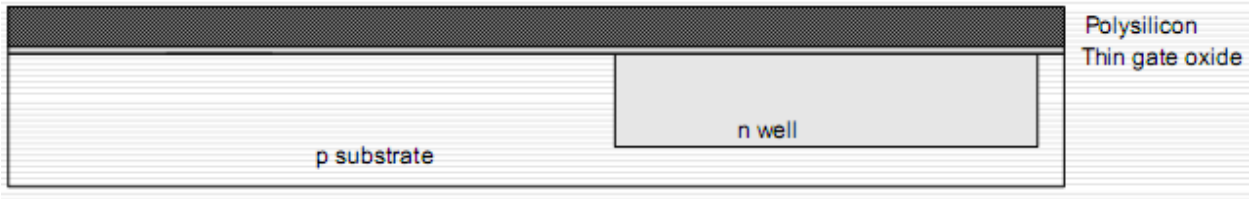
† Subsequent steps involve similar series of steps



Polysilicon:

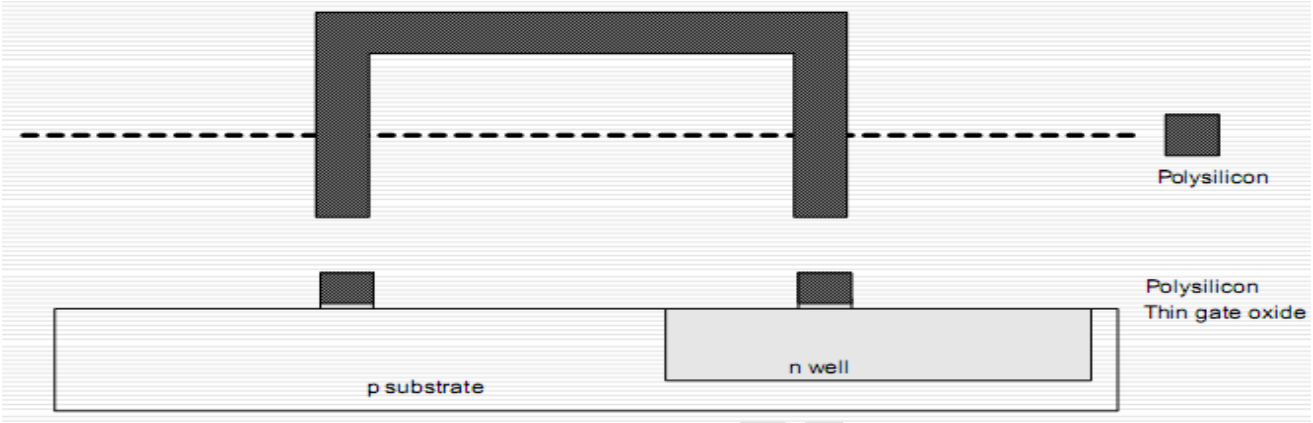
† Deposit very thin layer of gate oxide < 20 Å> (6-7 atomic layers)

† Chemical Vapor Deposition (CVD) of silicon layer Place wafer in furnace with Silane gas (SiH4) Forms many small crystals called polysilicon Heavily doped to be good conductor



Polysilicon Patterning:

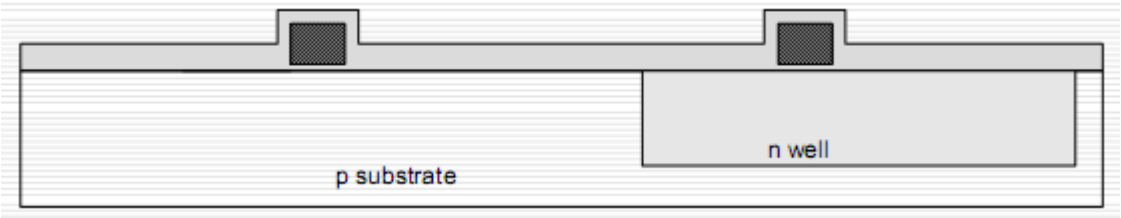
† Use same lithography process to pattern polysilicon



Self-Aligned Process:

† Use oxide and masking to expose where n+ dopants should be diffused or implanted

† N-diffusion forms nMOS source, drain, and n-well contact

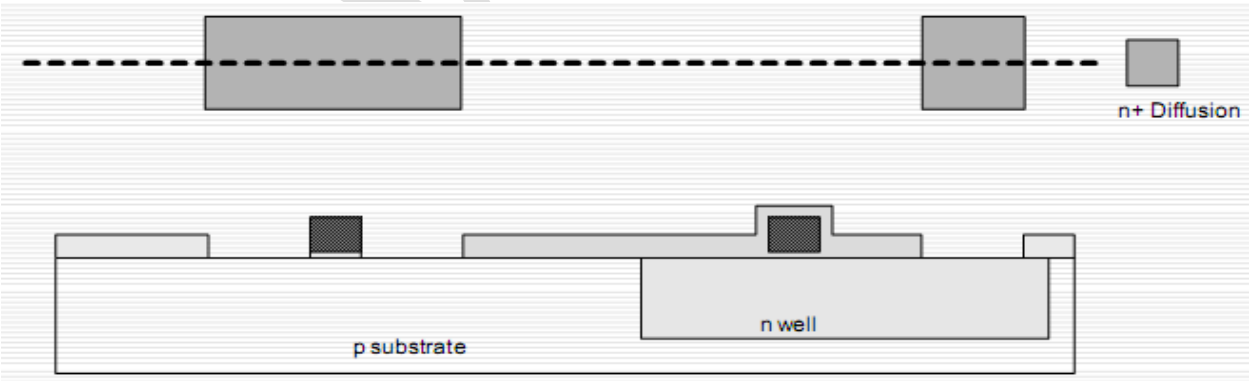


N-diffusion:

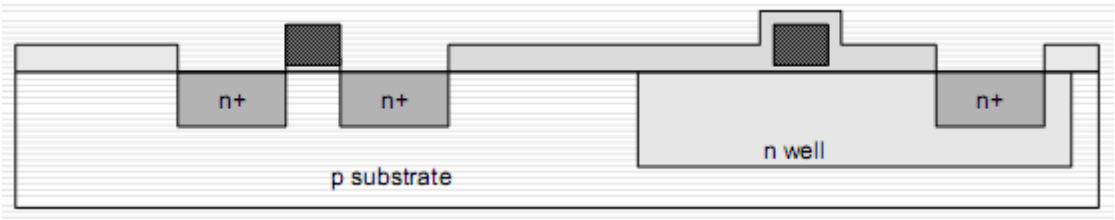
† Pattern oxide and form n+ regions

† Self-aligned process where gate blocks diffusion

† Polysilicon is better than metal for self-aligned gates because it doesn't melt during later processing

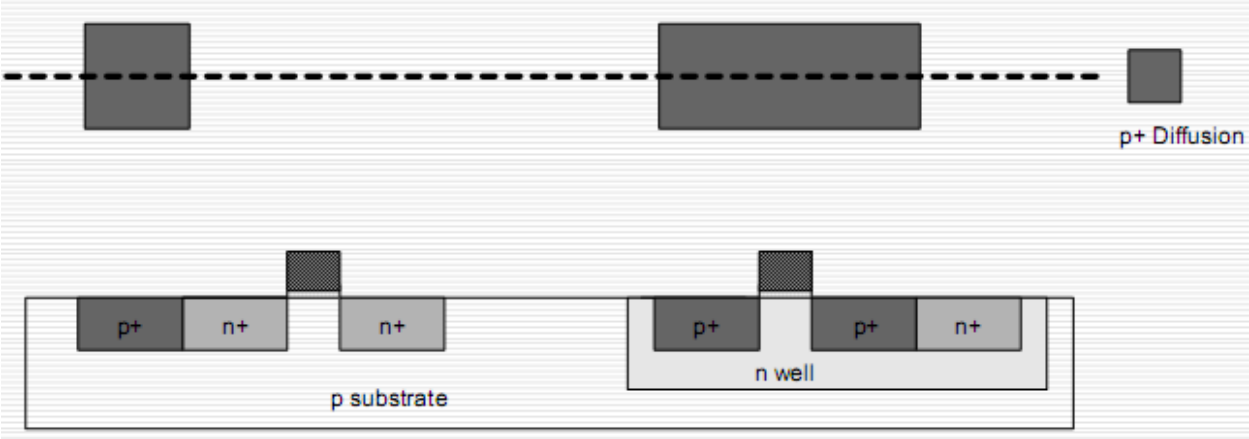


Historically dopants were diffused Usually ion implantation today But regions are still called diffusion



Strip off oxide to complete patterning step

P-Diffusion:† Similar set of steps form p+ diffusion regions for pMOS source and drain and substrate contact

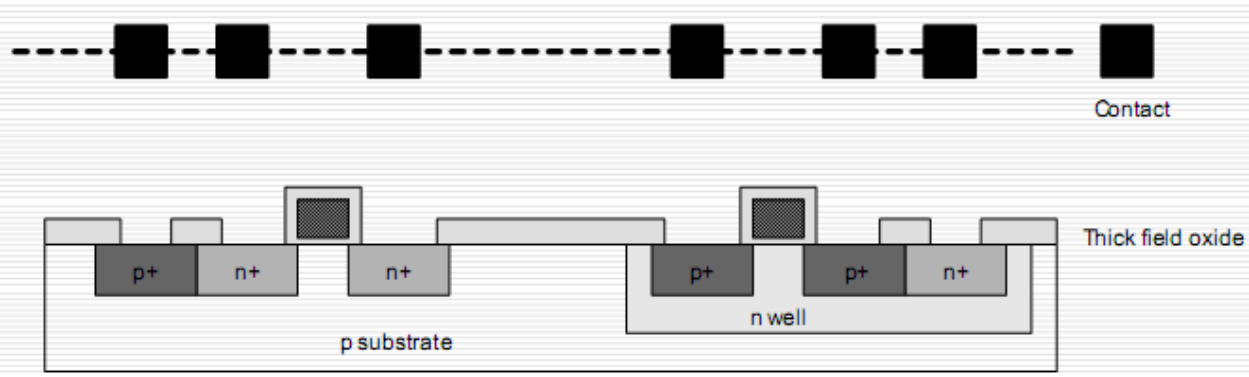


Contacts:

† Now we need to wire together the devices

† Cover chip with thick field oxide

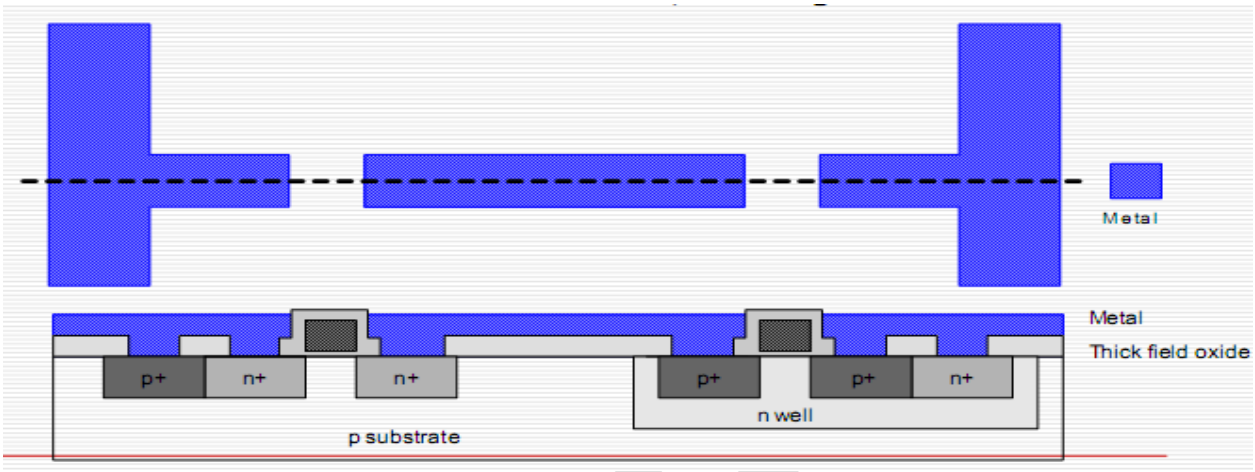
† Etch oxide where contact cuts are needed



Metallization:

† Sputter on aluminum over whole wafer

† Pattern to remove excess metal, leaving wires



Passive layer formation.

P-well process is similar but starts with a p-well implant rather than an n-well implant.

Twin-Tub (Twin-Well) CMOS Process:

This technology provides the basis for separate optimization of the nMOS and pMOS transistors, thus making it possible for threshold voltage, body effect and the channel trans conductance of both types of transistors to be tuned independently. Generally, the starting material is a n+ or p+ substrate, with a lightly doped epitaxial layer on top. This epitaxial layer provides the actual substrate on which the n-well and the p-well are formed. Since two independent doping steps are performed for the creation of the well regions, the dopant concentrations can be carefully optimized to produce the desired device characteristics. The Twin-Tub process is shown below.

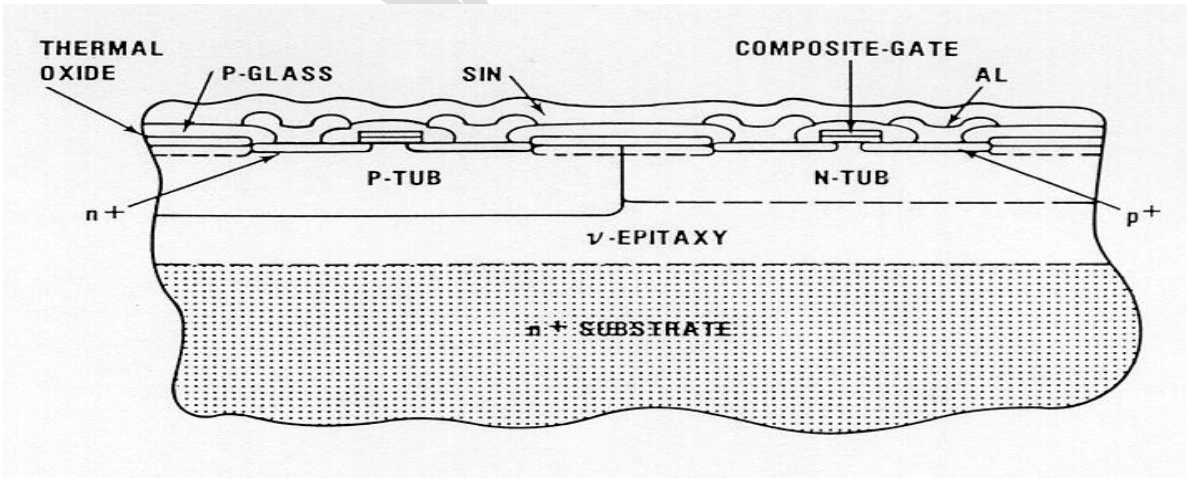


Figure-1.19: Process flow for the fabrication of Twin-Well MOSFET.

In the conventional p & n-well CMOS process, the doping density of the well region is typically about one order of magnitude higher than the substrate, which, among other effects, results in unbalanced drain parasitics. The twin-tub process avoids this problem.

Silicon-on-Insulator (SOI) CMOS Process:

Rather than using silicon as the substrate material, technologists have sought to use an insulating substrate to improve process characteristics such as speed and latch-up susceptibility. The SOI CMOS technology allows the creation of independent, completely isolated nMOS and pMOS transistors virtually side-by-side on an insulating substrate. The main advantages of this technology are the higher integration density (because of the absence of well regions), complete avoidance of the latch-up problem, and lower parasitic capacitances compared to the conventional p & n-well or twin-tub CMOS processes. A cross-section of nMOS and pMOS devices using SOI process is shown below.

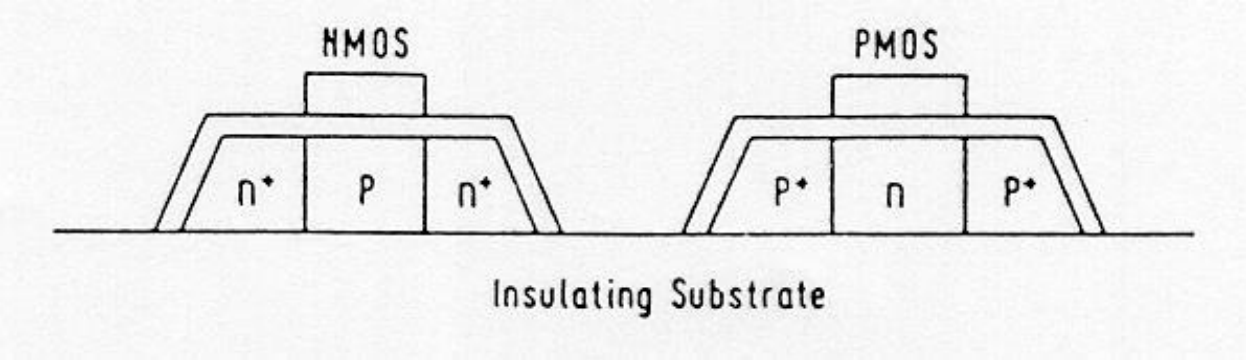


Figure-1.20: Process flow for Silicon-on-Insulator.

The SOI CMOS process is considerably more costly than the standard p & n-well CMOS process. Yet the improvements of device performance and the absence of latch-up problems can justify its use, especially for deep-sub-micron devices.

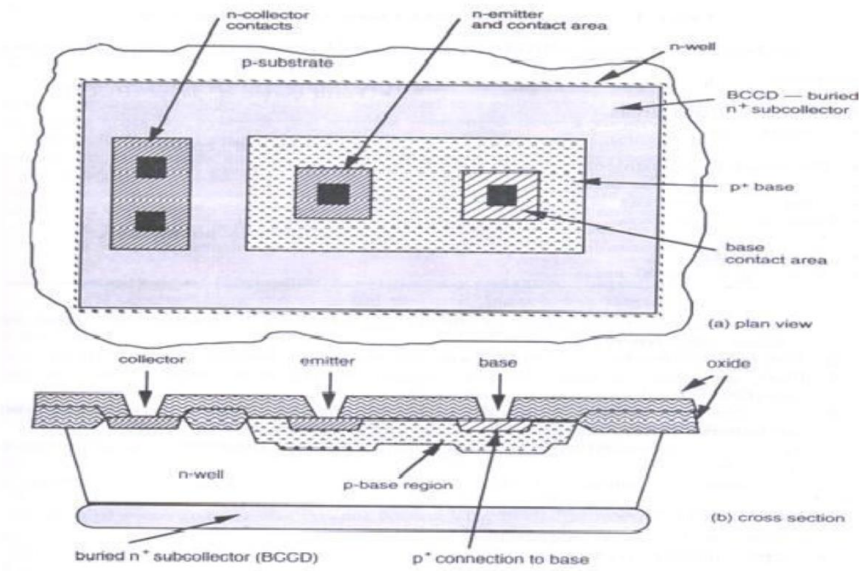
BI-CMOS TECHNOLOGY:

A BiCMOS circuit consists of both bipolar junction transistors and MOS transistors on a single substrate. The deficiency of MOS technology is the limited load driving capabilities of MOS transistors. This is due to the limited current sourcing and current sinking abilities associated with both p- and n-transistors and although it is possible, to design so-called super buffers using MOS transistors alone, such arrangements do not always compare well with the capabilities of bipolar transistors.

Bipolar transistors also provide higher gain and have better noise and high frequency characteristics than MOS 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. BiCMOS is able to achieve VLSI circuits with speed-power density performance previously not possible with either technology individually.

Theoretically there should be little difficulty in extending CMOS fabrication processes to include bipolar as well as MOS transistors. In fact, a problem of p-well and n-well CMOS processing is that parasitic bipolar transistors are inadvertently formed as part of the outcome of fabrication. The production of npn bipolar transistors with good performance characteristics can be achieved, for example, by extending the standard n-well CMOS processing to include further masks to add two additional layers- the n+ subcollector and p+ base layers. The npn transistor is formed in an n-well and the additional p+ base region is located in the well to form the p-base region of the transistor. The second additional layer, the buried n+ subcollector (BCCD), is added to reduce the n-well (collector) resistance and thus improve the quality of the bipolar transistor.

The simplified general arrangement of such a bipolar npn transistor is shown in below figure.

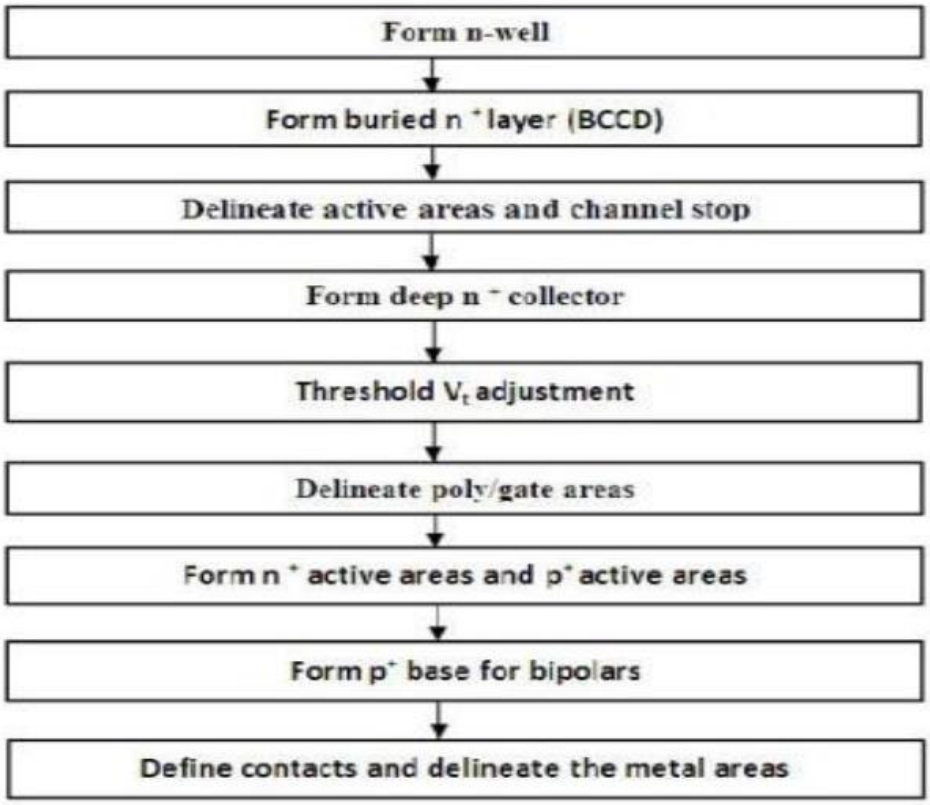


Arrangement of BiCMOS npn transistor

BiCMOS Fabrication in an N-well Process

The basic process steps used are those already outlined for CMOS but with additional process steps and additional masks defining (i) the p+ base region; (ii) n+ collector area; and (iii) the buried subcollector (BCCD). Below Table sets out the process steps for a single poly, single metal CMOS n-well process, showing the additional process steps for the bipolar devices.

N-Well BiCMOS fabrication Process Steps :



n-well BiCMOS fabrication steps

Comparison between CMOS and Bipolar Technologies

MOS Technology	Bipolar Technology
Low static power dissipation High input impedance (low drive current)	High power dissipation Low input impedance (high drive current)
Scalable threshold voltage High nose margin	----- Low voltage swing logic
High packing density High delay sensitivity to load (fan-out limitations)	Low packing density Low delay sensitivity to load
Low output drive current Low gm	High output drive current High gm
Bidirectional capability (drain and source are interchangeable) A near ideal switching device	High fT at low currents Essentially unidirectional