

## Unit 1

### Basic MOS Technology

Transistor was first invented by William.B.Shockley, Walter Brattain and John Bardeen of Bell Laboratories. In 1961, first IC was introduced.

Levels of Integration:-

- i) SSI:- (10-100) transistors => Example: Logic gates
- ii) MSI:- (100-1000) => Example: counters
- iii) LSI:- (1000-20000) => Example:8-bit chip
- iv) VLSI:- (20000-1000000) => Example:16 & 32 bit up
- v) ULSI:- (1000000-10000000) => Example: Special processors, virtual reality machines, smart sensors

Moore's Law:-

“The number of transistors embedded on the chip doubles after every one and a half years.” The number of transistors is taken on the y-axis and the years in taken on the x-axis. The diagram also shows the speed in MHz. the graph given in figure also shows the variation of speed of the chip in MHz.

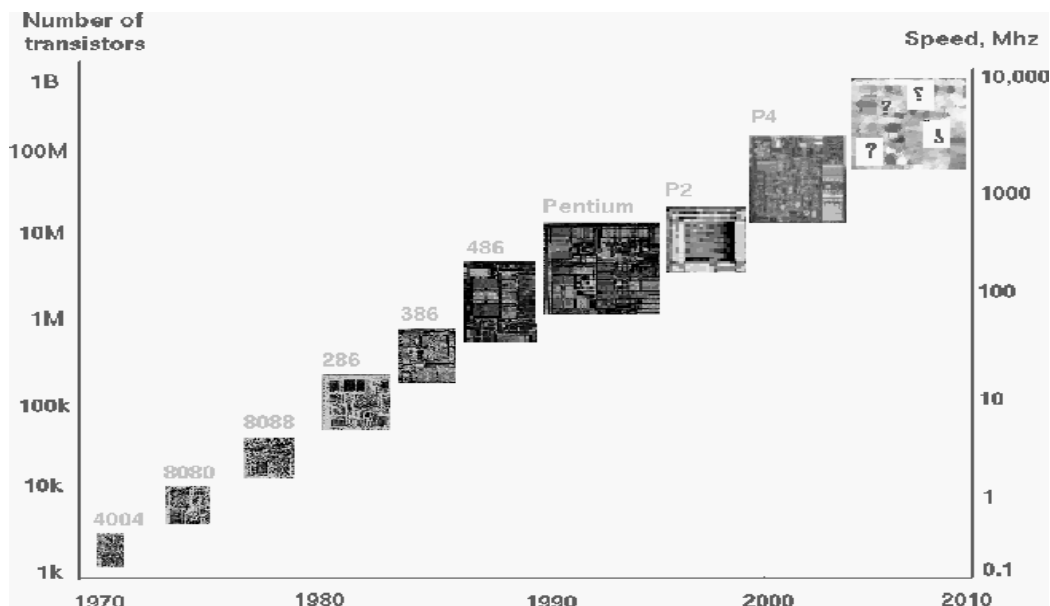


Figure 1. Moore's law

The graph in figure2 compares the various technologies available in ICs.

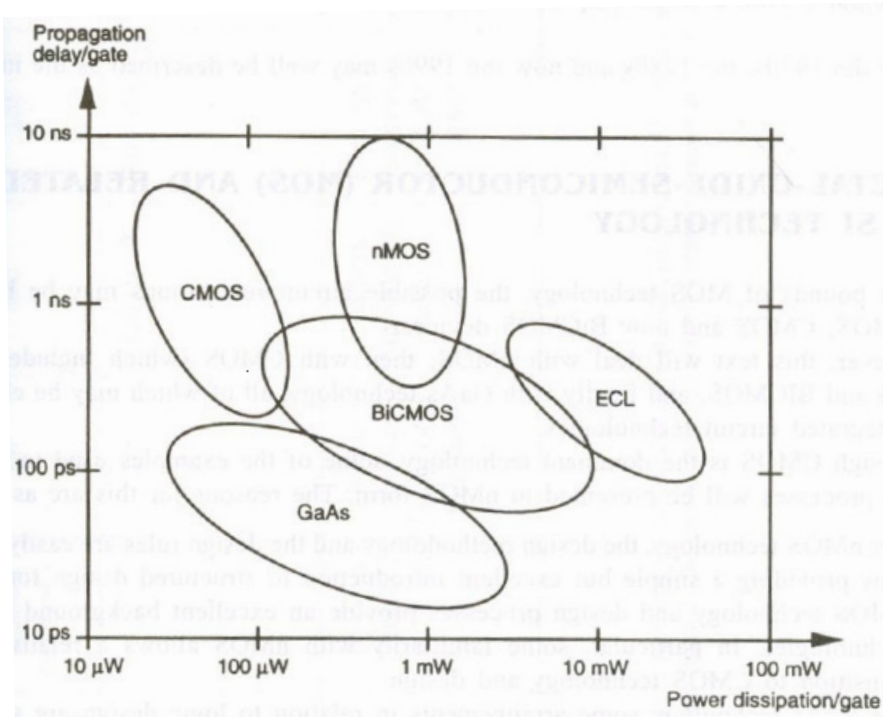


Figure 2. Comparison of available technologies

From the graph we can conclude that GaAs technology is better but still it is not used because of growing difficulties of GaAs crystal. CMOS looks to be a better option compared to nMOS since it consumes a lesser power. BiCMOS technology is also used in places where high driving capability is required and from the graph it confirms that, BiCMOS consumes more power compared to CMOS.

### Levels of Integration:-

- i) Small Scale Integration:- (10-100) transistors => Example: Logic gates
- ii) Medium Scale Integration:- (100-1000) => Example: counters
- iii) Large Scale Integration:- (1000-20000) => Example: 8-bit chip
- iv) Very Large Scale Integration:- (20000-1000000) => Example: 16 & 32 bit up
- v) Ultra Large Scale Integration:- (1000000-10000000) => Example: Special processors, virtual reality machines, smart sensors

### Basic MOS Transistors:

#### Why the name MOS?

We should first understand the fact that why the name Metal Oxide Semiconductor transistor, because the structure consists of a layer of Metal (gate), a layer of oxide ( $\text{SiO}_2$ ) and a layer of semiconductor. Figure 3 below clearly tell why the name MOS.

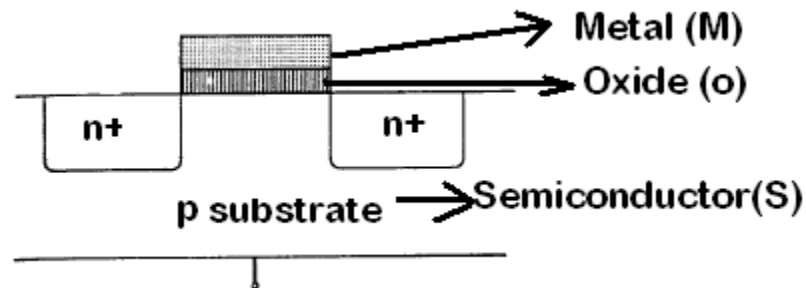


Figure 3. cross section of a MOS structure

We have two types of FETs. They are Enhancement mode and depletion mode transistor. Also we have PMOS and NMOS transistors.

In **Enhancement mode transistor** channel is going to form after giving a proper positive gate voltage. We have NMOS and PMOS enhancement transistors.

In **Depletion mode transistor** channel will be present by the implant. It can be removed by giving a proper negative gate voltage. We have NMOS and PMOS depletion mode transistors.

#### N-MOS enhancement mode transistor:-

This transistor is normally off. This can be made ON by giving a positive gate voltage. By giving a +ve gate voltage a channel of electrons is formed between source drain.

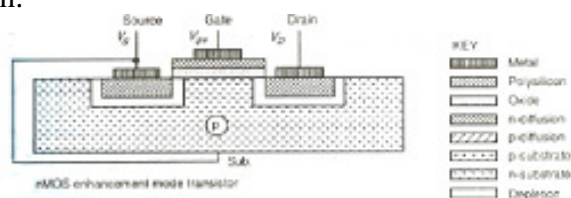


Fig 5. Nmos Enhancement transistor

#### P-Mos enhancement mode transistors:-

This is normally on. A Channel of Holes can be performed by giving a -ve gate voltage. In P-Mos current is carried by holes and in N-Mos its by electrons. Since the mobility is of holes less than that of electrons P-Mos is slower.

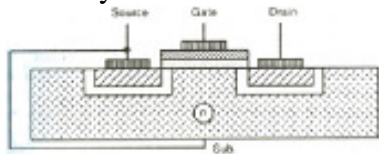
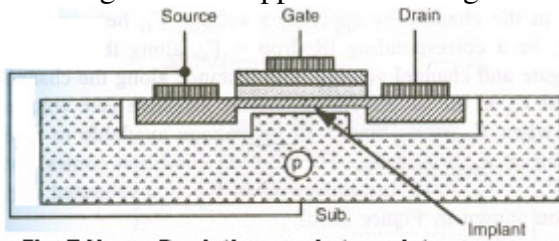


Fig 6. Pmos Enhancement transistor

**N-MOS depletion mode transistor:-**

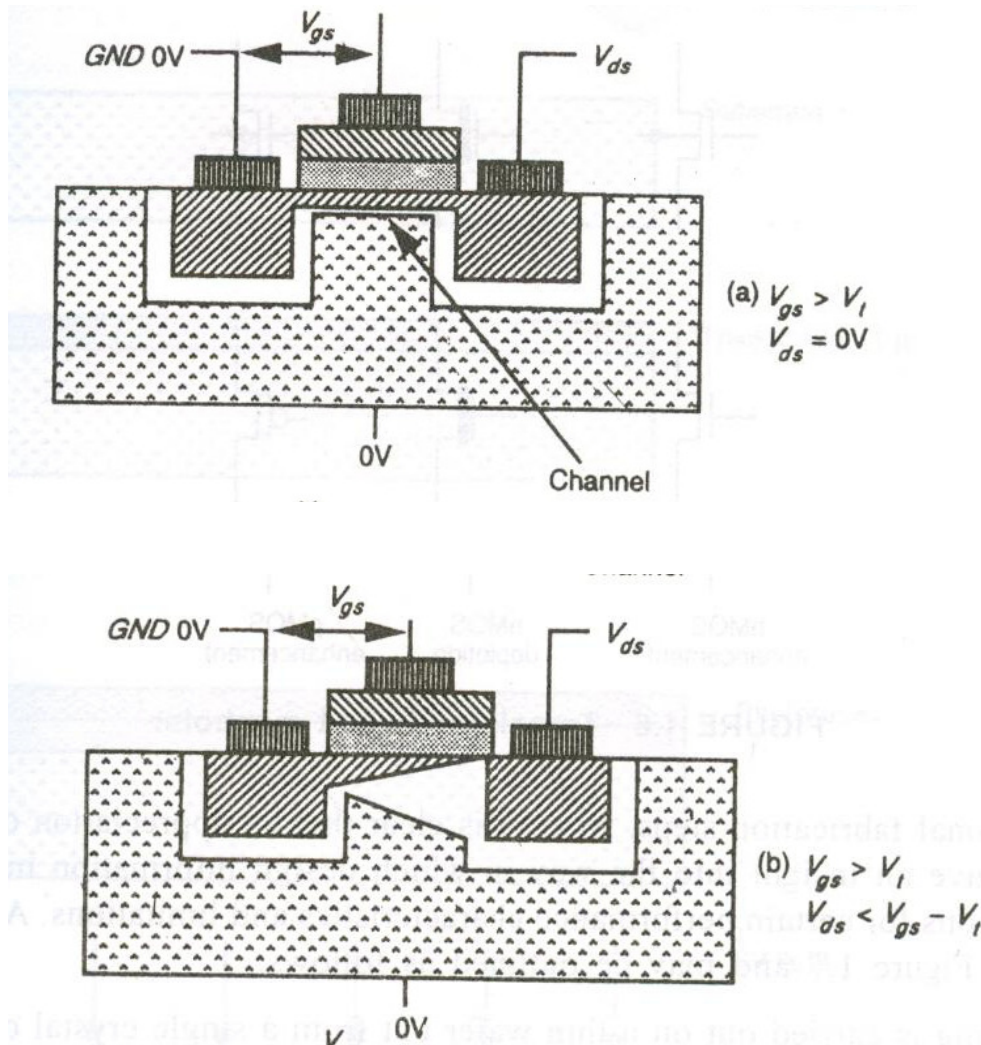
This transistor is normally ON, even with  $V_{gs}=0$ . The channel will be implanted while fabricating, hence it is normally ON. To cause the channel to cease to exist, a -ve voltage must be applied between gate and source.



**Fig. 7 Nmos Depletion mode transistor**

NOTE: Mobility of electrons is 2.5 to 3 times faster than holes. Hence P-MOS devices will have more resistance compared to NMOS.

**Enhancement mode Transistor action:-**



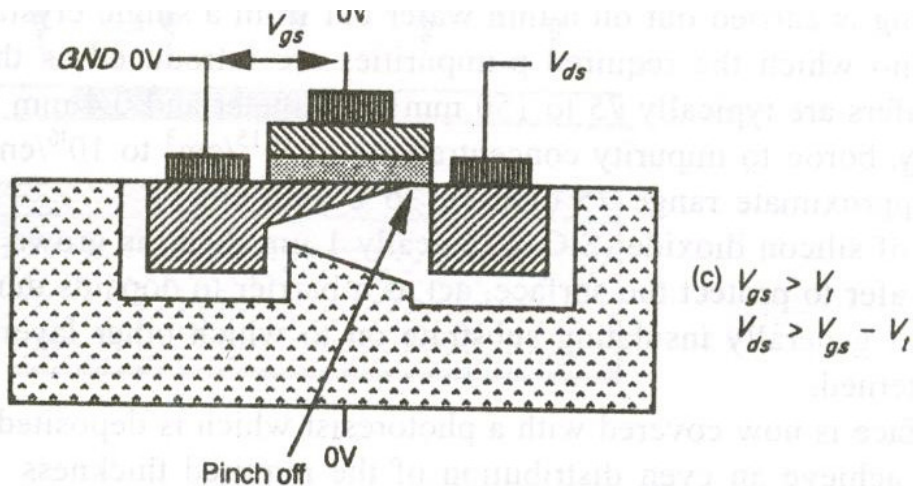


Figure 8(a)(b)(c) Enhancement mode transistor with different  $V_{ds}$  values

To establish the channel between the source and the drain a minimum voltage ( $V_t$ ) must be applied between gate and source. This minimum voltage is called as “Threshold Voltage”. The complete working of enhancement mode transistor can be explained with the help of diagram a, b and c.

a)  $V_{gs} > V_t$   
 $V_{ds} = 0$

Since  $V_{gs} > V_t$  and  $V_{ds} = 0$  the channel is formed but no current flows between drain and source.

b)  $V_{gs} > V_t$   
 $V_{ds} < V_{gs} - V_t$

This region is called the non-saturation Region or linear region where the drain current increases linearly with  $V_{ds}$ . When  $V_{ds}$  is increased the drain side becomes more reverse biased(hence more depletion region towards the drain end) and the channel starts to pinch. This is called as the pinch off point.

c)  $V_{gs} > V_t$   
 $V_{ds} > V_{gs} - V_t$

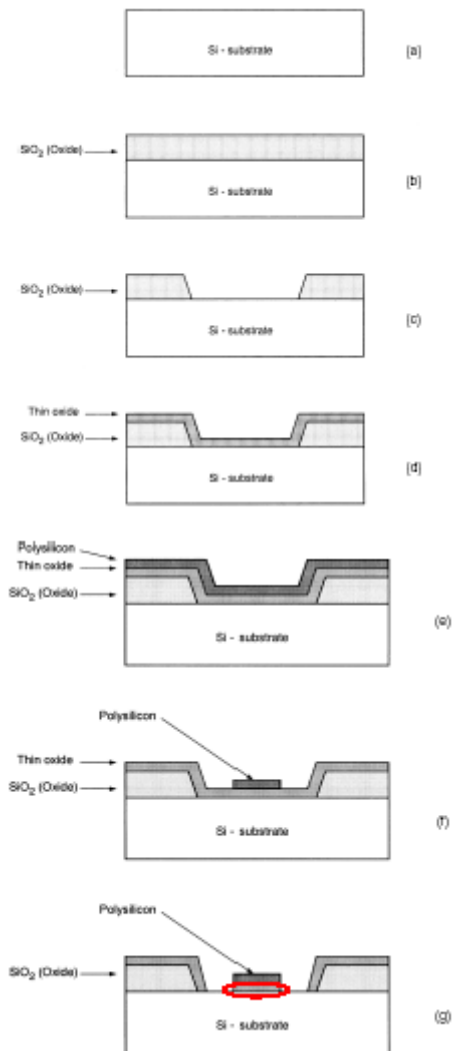
This region is called Saturation Region where the drain current remains almost constant. As the drain voltage is increased further beyond ( $V_{gs}-V_t$ ) the pinch off point starts to move from the drain end to the source end. Even if the  $V_{ds}$  is increased more and more, the increased voltage gets dropped in the depletion region leading to a constant current.

The typical threshold voltage for an enhancement mode transistor is given by  $V_t = 0.2 * V_{dd}$ .

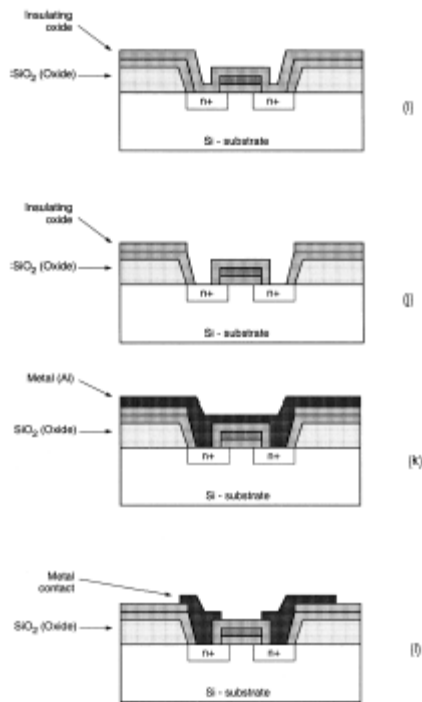
## Depletion mode Transistor action:-

We can explain the working of depletion mode transistor in the same manner, as that of the enhancement mode transistor only difference is, channel is established due to the implant even when  $V_{gs} = 0$  and the channel can be cut off by applying a  $-ve$  voltage between the gate and source. Threshold voltage of depletion mode transistor is around  $0.8 * V_{dd}$ .

## NMOS Fabrication:







**Figure 9 NMOS Fabrication process steps**

The process starts with the oxidation of the silicon substrate (Fig. 9(a)), in which a relatively thick silicon dioxide layer, also called field oxide, is created on the surface (Fig. 9(b)). Then, the field oxide is selectively etched to expose the silicon surface on which the MOS transistor will be created (Fig. 9(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. 9(d)). On top of the thin oxide, a layer of polysilicon (polycrystalline silicon) is deposited (Fig. 9(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. 9(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. 9(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 9(h) shows that the doping penetrates the exposed areas on the silicon surface, ultimately creating 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. Once the source and drain regions are completed, the entire surface is again covered with an insulating layer of silicon dioxide (Fig. 9 (i)). The insulating oxide layer is then patterned in order to provide contact windows for the drain and source junctions (Fig. 9 (j)). The surface is covered with evaporated aluminum which will form the interconnects (Fig. 9 (k)). Finally, the metal layer is patterned and etched, completing the interconnection of the MOS transistors on the surface (Fig. 9 (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 fabrication:** When we need to fabricate both nMOS and pMOS transistors on the same substrate we need to follow different processes. The three different processes are, P-well process, N-well process and Twin tub process.

### P-WELL PROCESS:

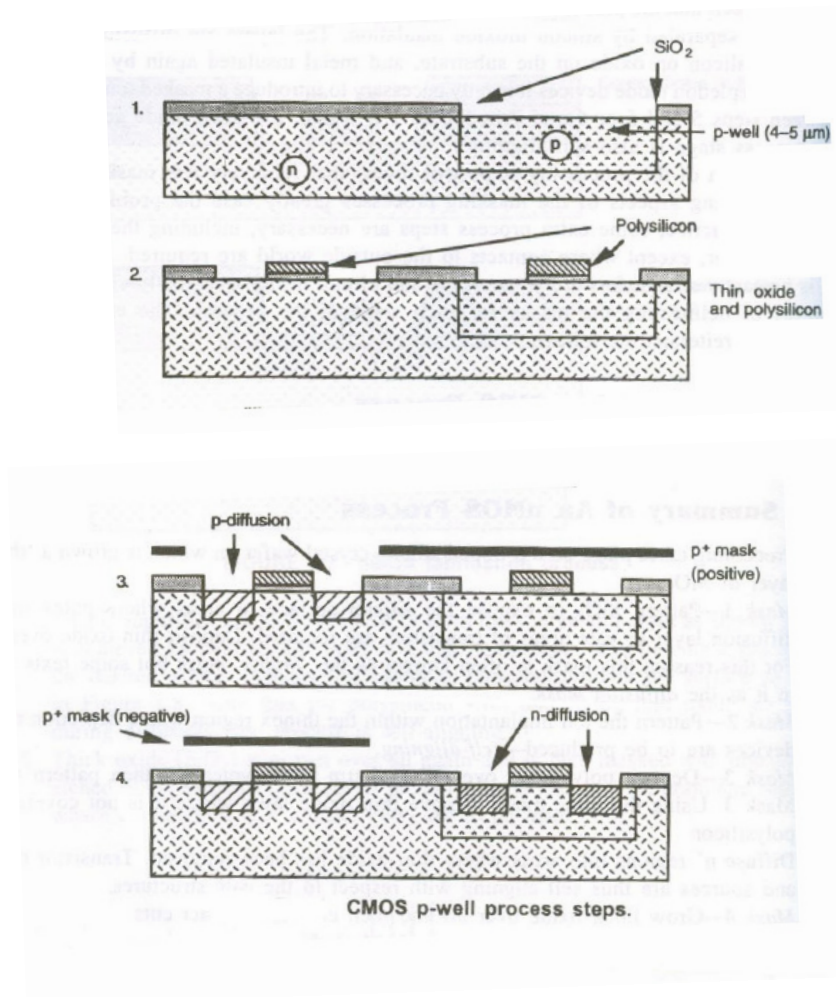


Figure 10 CMOS Fabrication (P-WELL) process steps



The p-well process starts with a n type substrate. The n type substrate can be used to implement the pMOS transistor, but to implement the nMOS transistor we need to provide a p-well, hence we have provided the place for both n and pMOS transistor on the same n-type substrate.

Mask sequence.

Mask 1:

Mask 1 defines the areas in which the deep p-well diffusion takes place.

Mask 2:

It defines the thin oxide region (where the thick oxide is to be removed or stripped and thin oxide grown)

Mask 3:

It's used to pattern the polysilicon layer which is deposited after thin oxide.

Mask 4:

A p+ mask (anded with mask 2) to define areas where p-diffusion is to take place.

Mask 5:

We are using the -ve form of mask 4 (p+ mask) It defines where n-diffusion is to take place.

Mask 6:

Contact cuts are defined using this mask.

Mask 7:

The metal layer pattern is defined by this mask.

Mask 8:

An overall passivation (overglass) is now applied and it also defines openings for accessing pads.

The cross section below shows the CMOS pwell inverter.

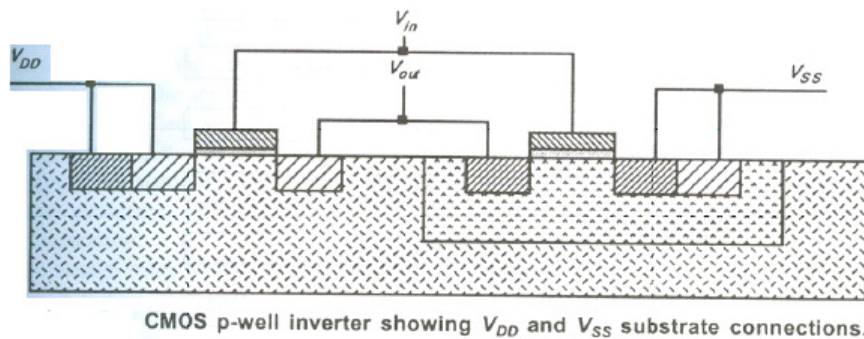
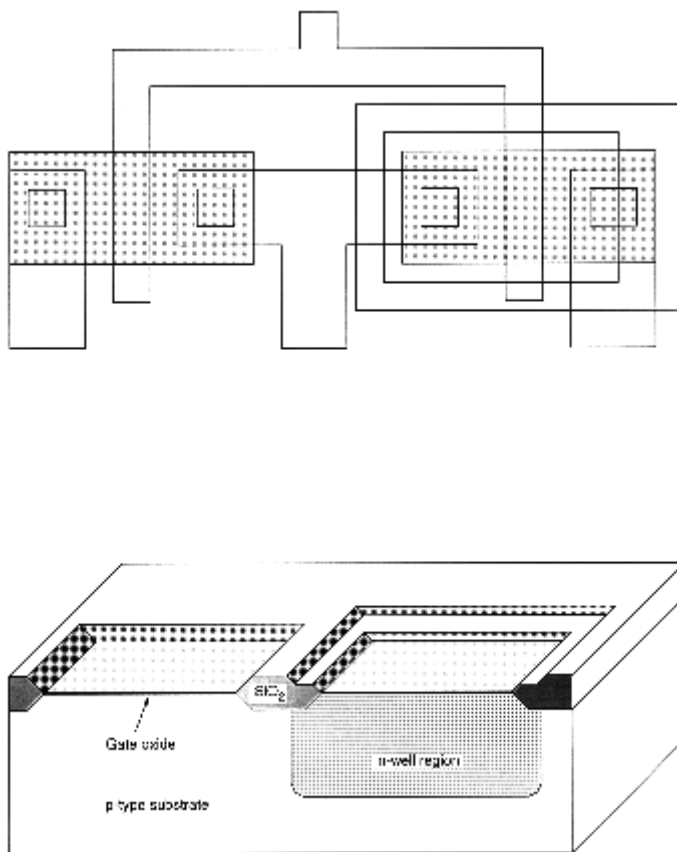


Figure 11 CMOS inverter (P-WELL)

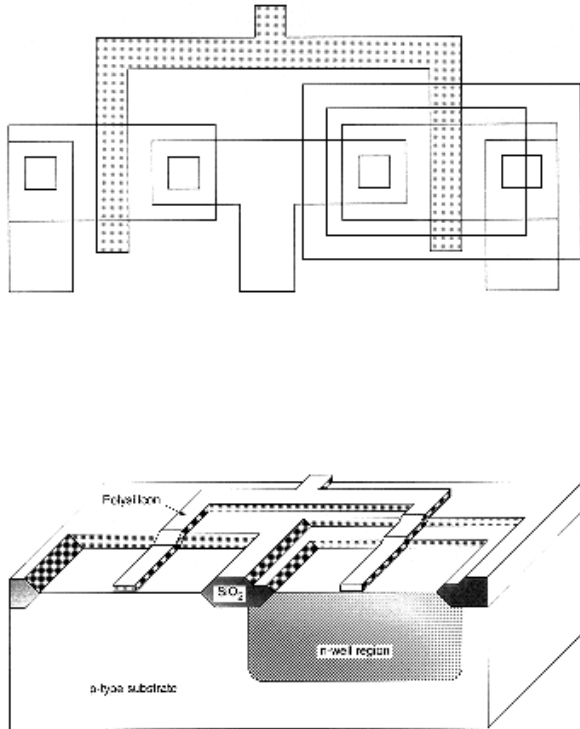
## N-WELL PROCESS:

In the following figures, some of the important process steps involved in the fabrication of a CMOS inverter will be shown by a top view of the lithographic masks and a cross-sectional view of the relevant areas.

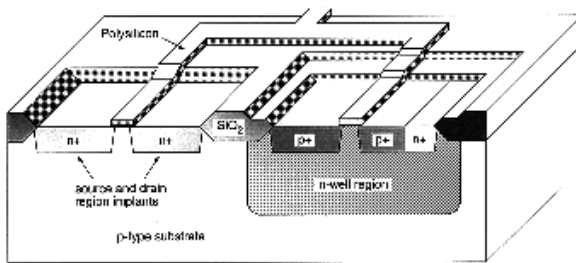
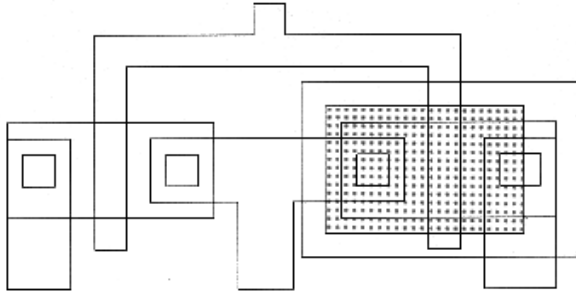
The n-well CMOS process starts with a moderately doped (with impurity concentration typically less than  $10^{15}$  cm<sup>-3</sup>) p-type silicon substrate. Then, an initial oxide layer is grown on the entire surface. The first lithographic mask defines the n-well region. Donor atoms, usually phosphorus, are implanted through this window in the oxide. Once the n-well is created, the active areas of the nMOS and pMOS transistors can be defined. Figures 12.1 through 12.6 illustrate the significant milestones that occur during the fabrication process of a CMOS inverter.



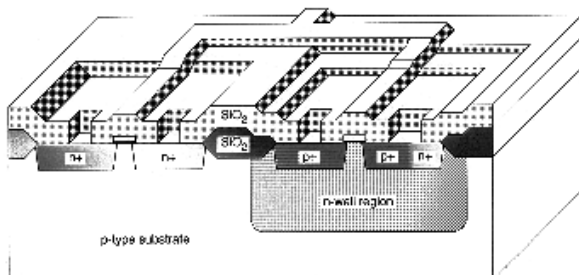
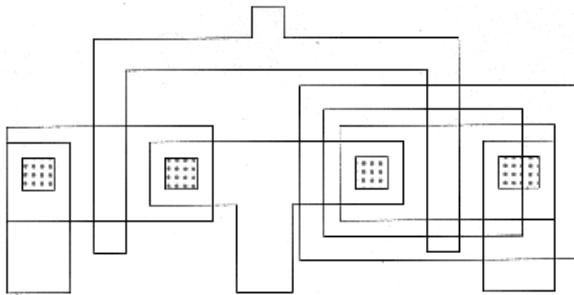
**Figure-12.1:** Following the creation of the n-well region, a thick field oxide is grown in the areas surrounding the transistor active regions, and a thin gate oxide is grown on top of the active regions. The thickness and the quality of the gate oxide are two of the most critical fabrication parameters, since they strongly affect the operational characteristics of the MOS transistor, as well as its long-term reliability.



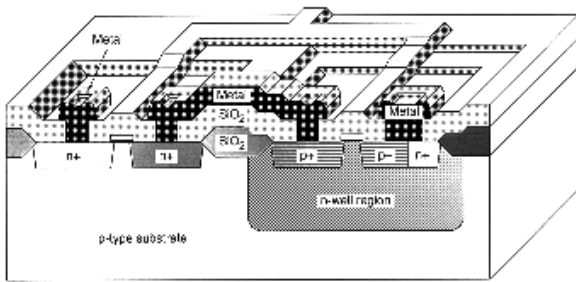
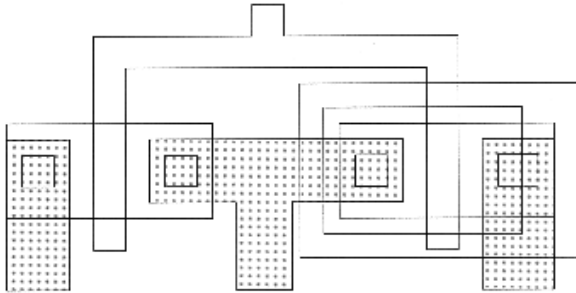
**Figure-12.2:** The polysilicon layer is deposited using chemical vapor deposition (CVD) and patterned by dry (plasma) etching. The created polysilicon lines will function as the gate electrodes of the nMOS and the pMOS transistors and their interconnects. Also, the polysilicon gates act as self-aligned masks for the source and drain implantations that follow this step.



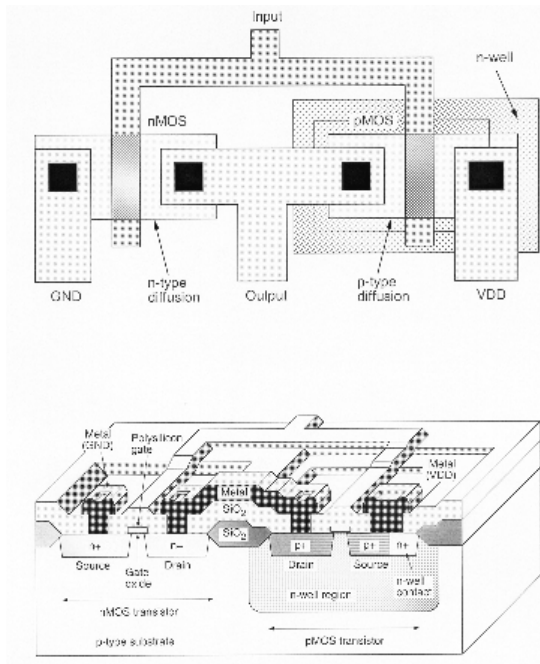
**Figure-12.3:** Using a set of two masks, the n+ and p+ regions are implanted into the substrate and into the n- well, respectively. Also, the ohmic contacts to the substrate and to the n-well are implanted in this process step.



**Figure-12.4:** An insulating silicon dioxide layer is deposited over the entire wafer using CVD. Then, the contacts are defined and etched away to expose the silicon or polysilicon contact windows. These contact windows are necessary to complete the circuit interconnections using the metal layer, which is patterned in the next step.



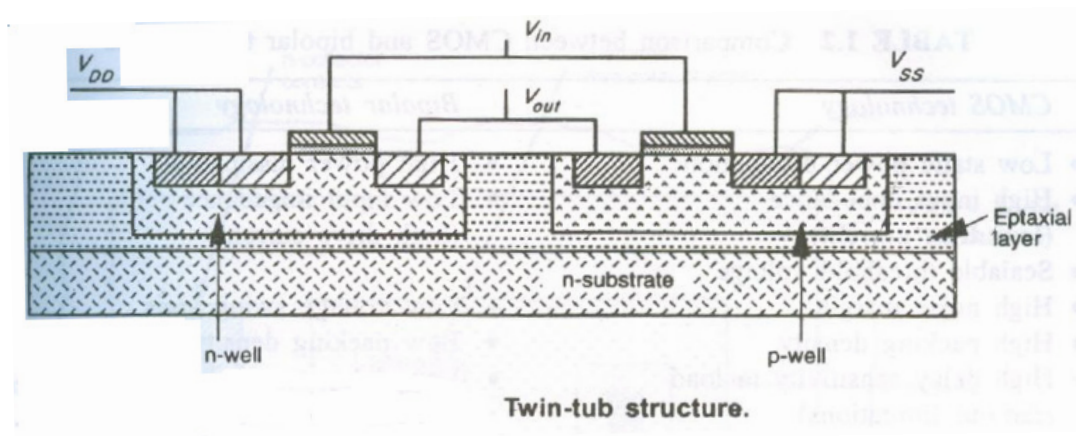
**Figure-12.5:** Metal (aluminum) is deposited over the entire chip surface using metal evaporation, and the metal lines are patterned through etching. Since the wafer surface is non-planar, the quality and the integrity of the metal lines created in this step are very critical and are ultimately essential for circuit reliability.



**Figure-12.6:** The composite layout and the resulting cross-sectional view of the chip, showing one nMOS and one pMOS transistor (built-in n-well), the polysilicon and metal interconnections. The final step is to deposit the passivation layer (for protection) over the chip, except for wire-bonding pad areas.

### Twin-tub process:

Here we will be using both p-well and n-well approach. The starting point is a n-type material and then we create both n-well and p-well region. To create the both well we first go for the epitaxial process and then we will create both wells on the same substrate.



**Figure 13 CMOS twin-tub inverter**

**NOTE: Twin tub process is one of the solutions for latch-up problem.**



### **Bi-CMOS technology: - (Bipolar CMOS)**

The driving capability of MOS transistors is less because of limited current sourcing and sinking capabilities of the transistors. To drive large capacitive loads we can think of Bi-Cmos technology.

This technology combines Bipolar and CMOS transistors in a single integrated circuit, by retaining benefits of bipolar and CMOS, BiCMOS is able to achieve VLSI circuits with speed-power-density performance previously unattainable with either technology individually.

### **Characteristics of CMOS Technology**

- Lower static power dissipation
- Higher noise margins
- Higher packing density – lower manufacturing cost per device
- High yield with large integrated complex functions
- High input impedance (low drive current)
- Scalable threshold voltage
- High delay sensitivity to load (fan-out limitations)
- Low output drive current (issue when driving large capacitive loads)
- Low transconductance, where transconductance,  $g_m \propto V_{in}$
- Bi-directional capability (drain & source are interchangeable)
- A near ideal switching device

### **Characteristics of Bipolar Technology**

- Higher switching speed
- Higher current drive per unit area, higher gain
- Generally better noise performance and better high frequency characteristics
- Better analogue capability
- Improved I/O speed (particularly significant with the growing importance of package limitations in high speed systems).
- high power dissipation
- lower input impedance (high drive current)
- low voltage swing logic
- low packing density
- low delay sensitivity to load
- high  $g_m$  ( $g_m \propto V_{in}$ )
- high unity gain band width (ft) at low currents
- essentially unidirectional

From the two previous paragraphs we can get a comparison between bipolar and CMOS technology.

The diagram given below shows the cross section of the BiCMOS process which uses an npn transistor.

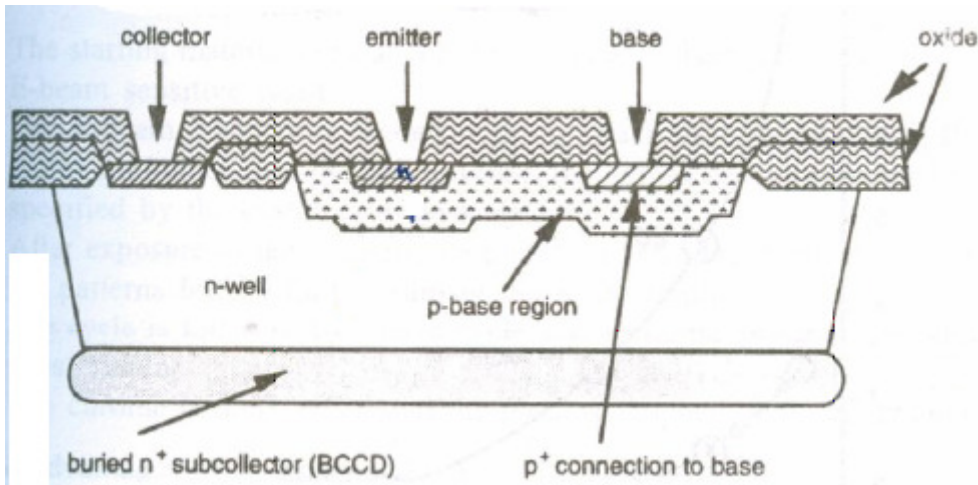


Figure 14 Cross section of BiCMOS process

The figure below shows the layout view of the BiCMOS process.

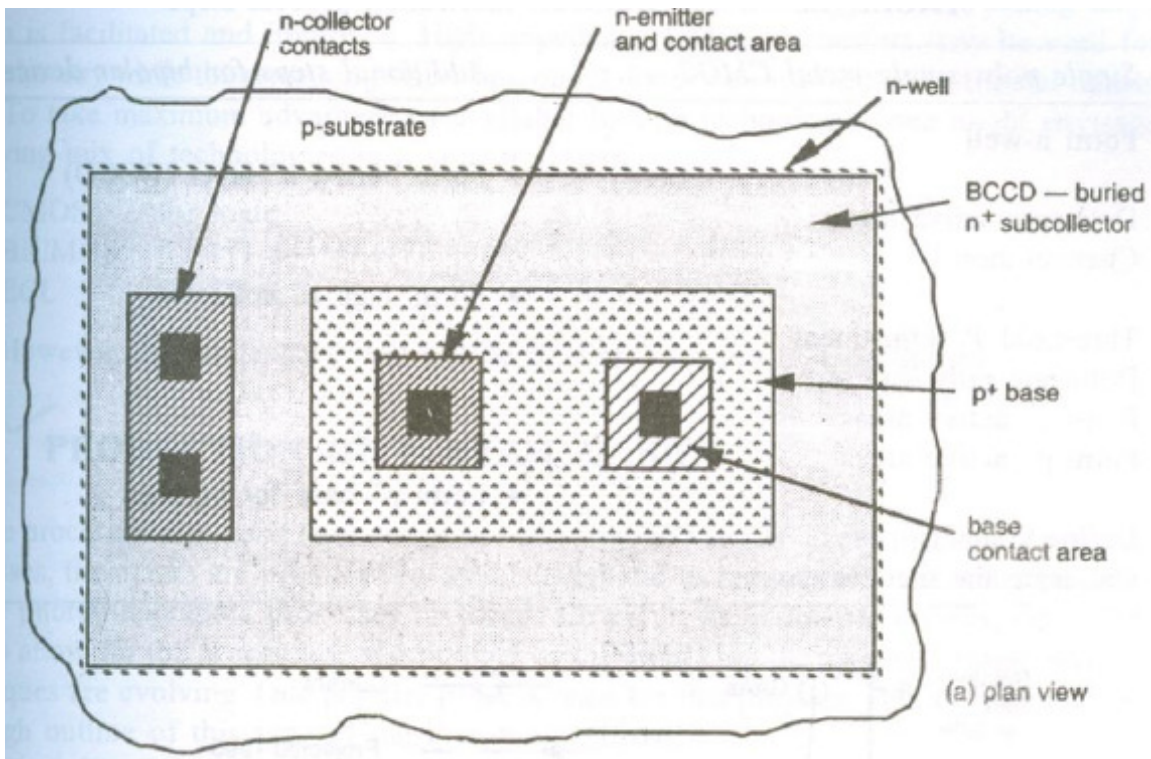


Fig.15. Layout view of BiCMOS process

The graph below shows the relative cost vs. gate delay.

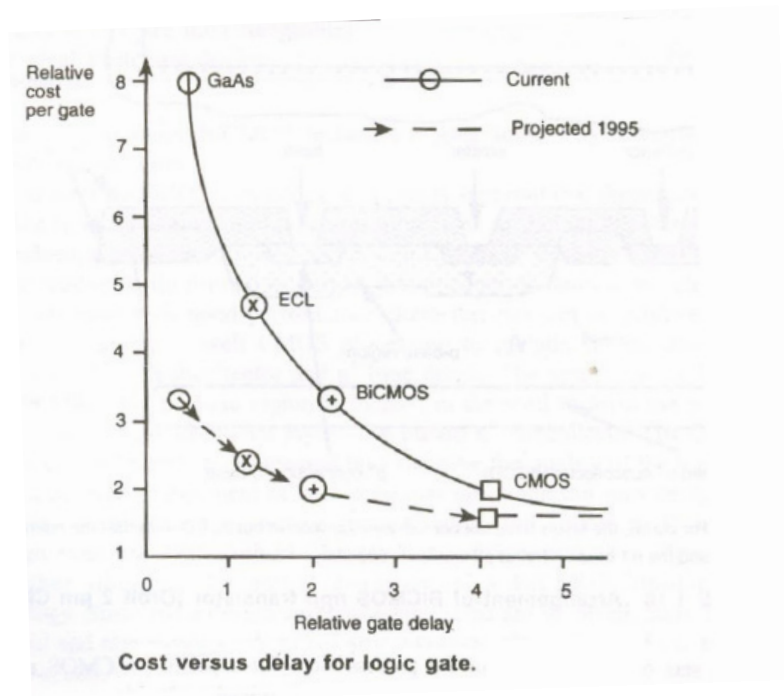


Fig.16. cost versus delay graph

### Production of e-beam masks:

In this topic we will understand how we are preparing the masks using e-beam technology. The following are the steps in production of e-beam masks.

- Starting materials is chromium coated glass plates which are coated with e-beam sensitive resist.
- E-beam machine is loaded with the mask description data.
- Plates are loaded into e-beam machine, where they are exposed with the patterns specified by mask description data.
- After exposure to e-beam, plates are introduced into developer to bring out patterns.
- The cycle is followed by a bake cycle which removes resist residue.
- The chrome is then etched and plate is stripped of the remaining e-beam resist.

We use two types of scanning, Raster scanning and vector scanning to map the pattern on to the mask.

In raster type, e-beam scans all possible locations and a bit map is used to turn the e-beam on and off, depending on whether the particular location being scanned is to be exposed or not.

---

In vector type, beam is directed only to those location which are to be exposed.

**Advantages e-beam masks:**

- Tighter layer to layer registration;
- Small feature sizes