

Introduction to Ic Technology. ①

Integrated circuits: It is the combination of logic gates fabricated on a small chip of semiconductor.

(or)

A low cost, complete electronic circuit consisting of active and passive components on a small chip of semiconductor.

In Ic's fabrications mostly preferred semiconductor is silicon.

The main features of Si over Ge

- It is ability to conduct electricity in controlled manner.
- It can operate upto 150°C compared Ge (100°C).
- It is suitable for fabrication and grows a stable oxide (SiO_2)
- Undoped resistivity of Si is $23 \times 10^4 \Omega\text{cm}$ and Ge $47 \Omega\text{cm}$.

→ Ge is difficult to dope as compared to Si

→ Si is 2nd most abundant element on earth.

→ Growth of (SiO₂) layer does not requires any special arrangement.

→ Ge has high junction leakage current or reverse current as compared to 'Si'

classification of ICs:-

a) On the basis of fabrication.

b) Scale or level of Integration

c) Application. → Digital IC's.

classification of IC's on Scale, Integration:

Generation.	Products.
1 st ⇒ SSI → 10	→ Logic gates / F/F's
2 nd ⇒ MSI → 100-1000	→ Counters, Mux, Address Shifters
3 rd ⇒ LSI → 1000-20,000	→ 8-bit - up ROM, RAM
4 th ⇒ VLSI → 20,000-10 lakh	→ 16 and 32 bit MP, sophisticated Peripherals GHM DRAM.

5th VLSI : ULTRA LARGE SCALE INTEGRATION

Special type processes machines, Smart Sensors.

6th GSI : Giant scale Integration

→ To be implemented version.

Advantages of IC:

- Higher packaging densities
- High ckt speeds.
- Miniaturization of ckts.

VLSI Technology.

- It is a very large scale Integration.
- Process of integration of millions of transistors on a piece of Si material.
- VLSI is one of the basic building block of today's higher-end or advance technology.
- VLSI is finding its applications in a variety of areas from simple electronic consumer products of very complex circuits used in space electronics.
- Now a days the trend in digital #lw

Product is to integrate as much circuitry as possible on to a single-chip.

Applications:-

VLSI circuits are used almost everywhere

Such as.

- automobiles
- Digital cameras
- cell phones
- Different computer systems and other electronic products.

VLSI design:-

→ VLSI has offered new opportunity to design the circuits which never possible before.

Ex: PLD's (Programmable Logic Devices)

EDAT Tools (Electronic design automation tools).

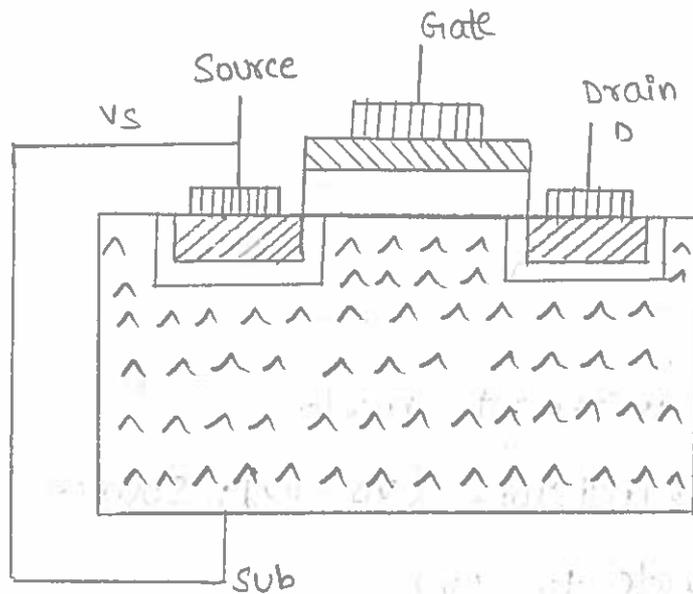
VLSI design styles:-

Two design styles

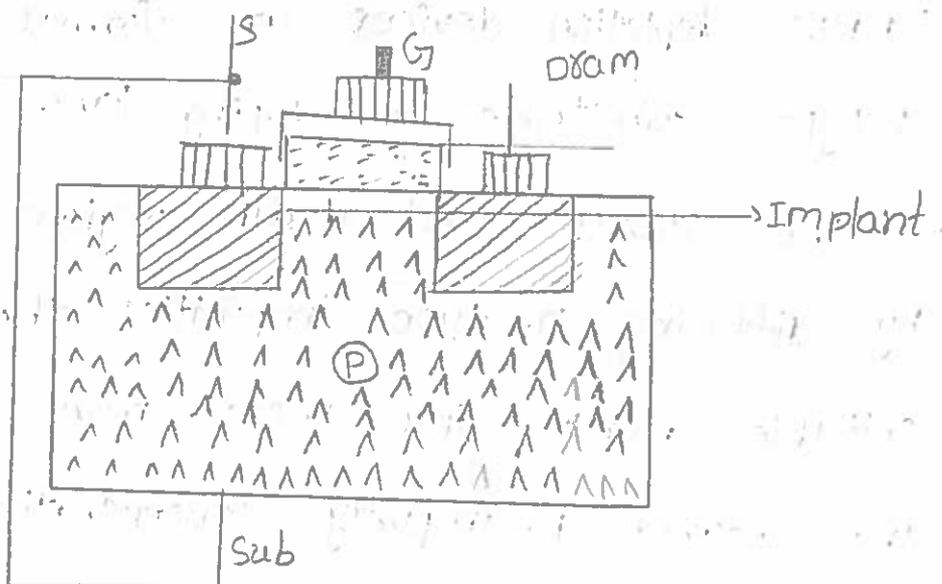
- a) semi custom
- b) Full custom.

Basic MOS transistor:

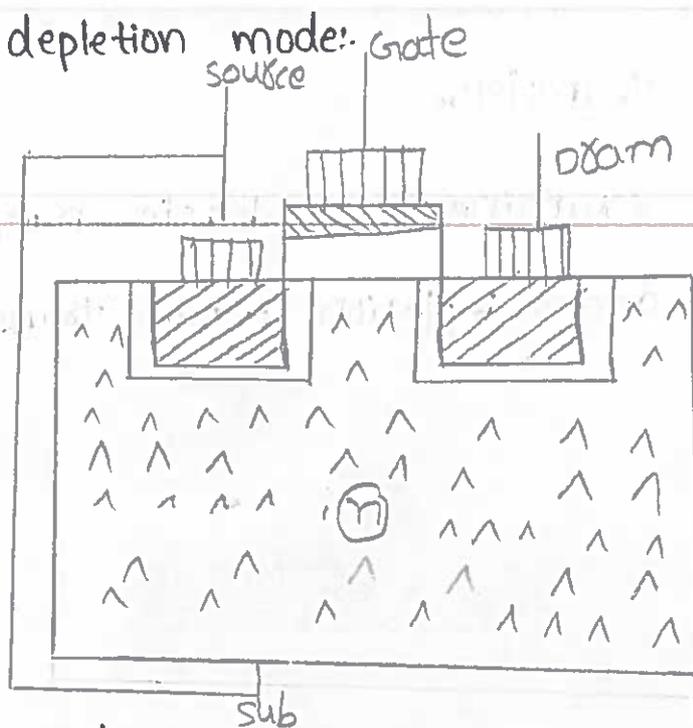
Let us examine the basic nmos enhancement and depletion mode transistors.



nmos enhancement mode:-



nMOS depletion mode:



p-MOS enhancement mode:

MOS transistors ($V_D = 0V$); source gate and substrate to $0V$)

N-MOS.

- nMOS depletion devices are formed in a p-type substrate of 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-regions.

④
→ The source and drain are isolated from one another by two diodes.

→ connections to the S and D 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 S and D.

and this is commonly achieved by two ways.

1) Enhancement mode

2) Depletion mode.

Enhancement mode:

→ In which the channel is not established and the device in a non-conducting condition, (if $V_D = V_S = V_{GS} = 0$).

→ If this gate is connected to a suitable +ve voltage with respect to gate source, then \bar{e} field established b/w 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 b/w source and drain.

Depletion mode:

In which, under the condition $V_{GS} = 0$ by implanting suitable impurities in the regions b/w source and drain during manufacture and prior to depositing the insulation and gate. Hence channel established between S and D.

→ And the channel is closed by applying a suitable $-ve$ voltage to the gate.

Enhancement mode operation:

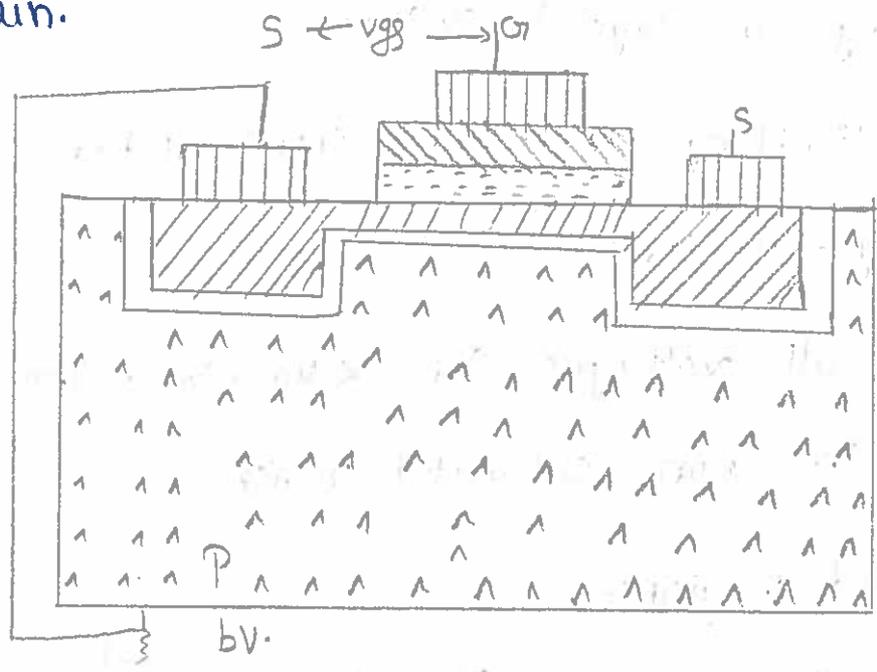
→ To establish the channel between source and drain, a minimum voltage is required.

→ V_T is defined as the V_{GS} voltage at which a MOS device begins to conduct.

In order to establish the channel, a min voltage level of V_T must be established b/w gate to source or (source to substrate).

Initially $V_{GS} = V_T$ and $V_{DS} = 0$.

\therefore No current flows between source and Drain.



$V_{GS} > V_T$
 $V_{DS} = 0V$

→ Now applying V_{DS} between S and D, There is resistive drop along the channel, begins to change shape of the channel.

(IR drop = V_{DS}).

→ This results in the voltage b/w gate and channel (V_G and V_G') varying with distance.

At the source end of the channel:

V_{GS} will be maximum, since there is full V_G is effective is available to invert the channel.

However at the drain end, of the channel, only $(V_g - V_d)$ is effect.

So long as $V_{gs} - V_t > V_{ds}$.

The limiting condition comes when

$$V_{gs} - V_t = V_{ds}$$

For all voltages $V_{ds} < V_{gs} - V_t$. The device is in the non-saturated region.

Saturated region:

If $V_{ds} > V_{gs} - V_t$. (V_{ds} is induced to a level greater $V_{gs} - V_t$).

In this an IR drop $= V_{gs} - V_t$ takes place over $<$ the whole length of the channel.

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'.

Diffusion current completes path from source to drain in this case, causing

the channel to exhibit a high resistance and behave as a constant current source.

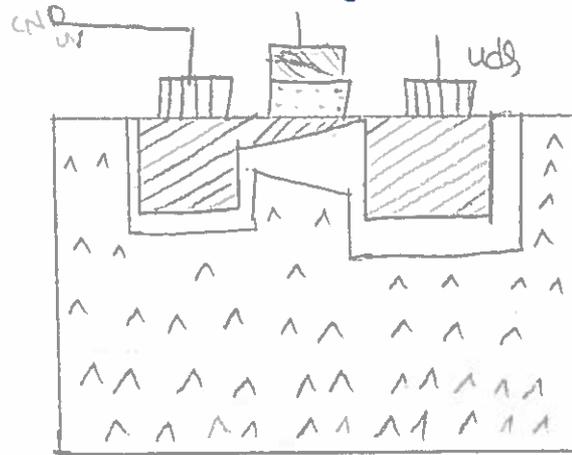
This region is, known as saturation, is characterised by

$$V_{ds} = V_{gs} - V_t$$

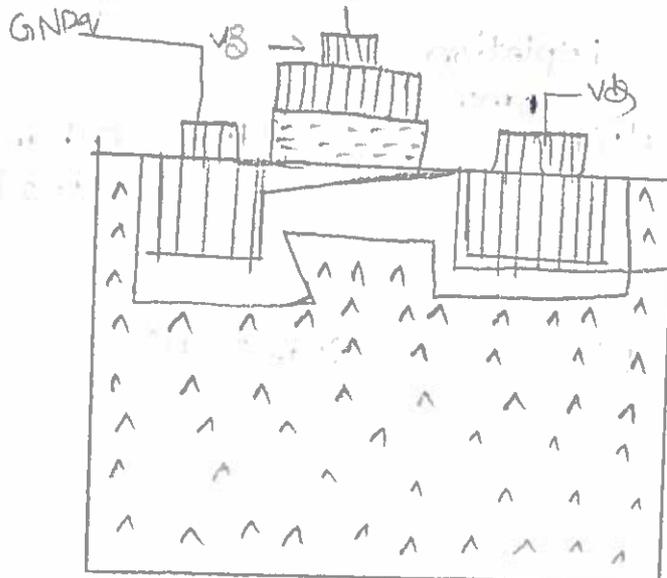
In all cases, the channel will cease to exist and no current flow when $V_{gs} < V_t$.

For enhancement mode devices $V_t = 1V$, $V_{DD} = 5V$.

Or general terms $V_t = 0.2 V_{DD}$.



$V_{gs} > V_t$
 $V_{ds} < V_{gs} - V_t$
Non-saturated



Pinched off
 $V_{gs} > V_t$
 $V_{ds} > V_{gs} - V_t$

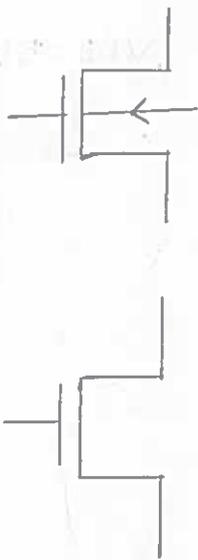
Saturated region

Depletion mode transistor action:

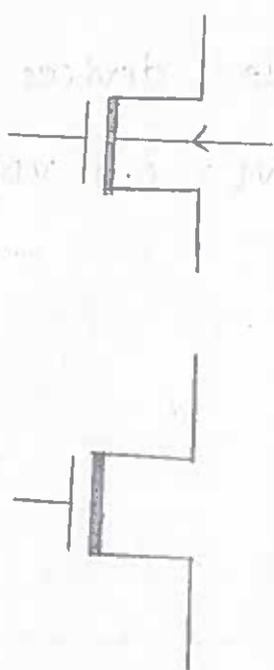
For depletion mode devices the channel is established, because of the implant, even when $V_{GS} = 0$, and to cause the channel to cease to exist a negative voltage V_{td} must be applied b/w G & S.

V_{td} is typically $< -0.8 V_{DD}$.

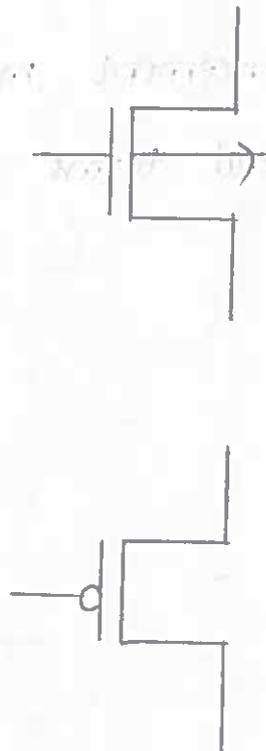
Commonly used symbols:



nmos
enhancement.



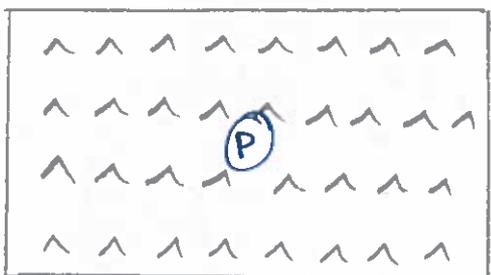
Depletion
nmos.



Pmos enhancement
mode.

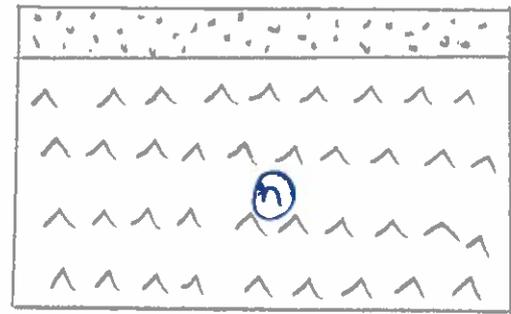
NMOS fabrication:

4)



substrate.

2)



Thick oxide (1 μm)

3)

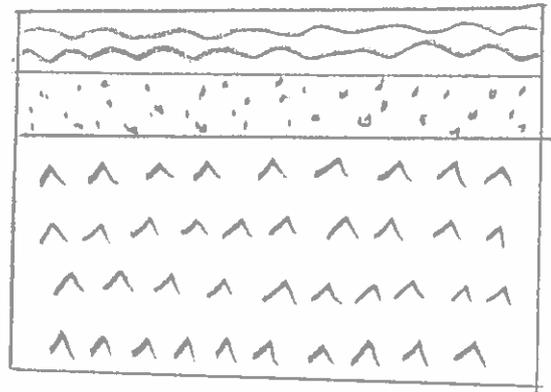
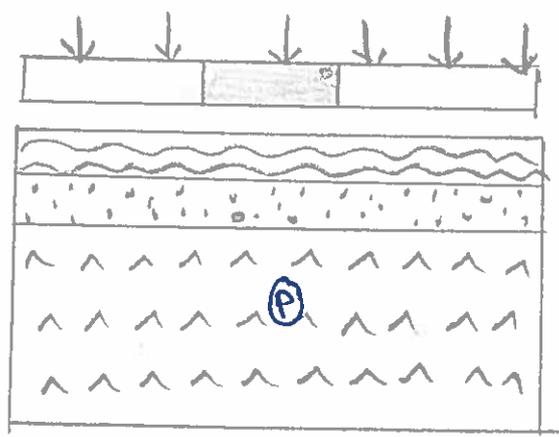


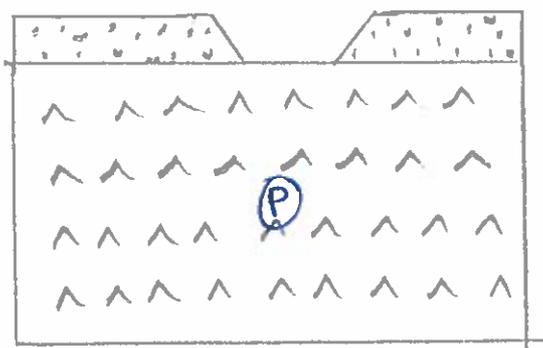
Photo resist

4)



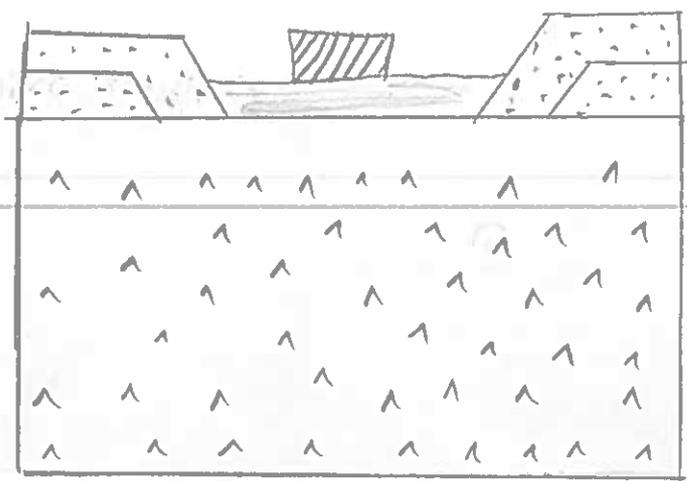
UV light mask.

5)



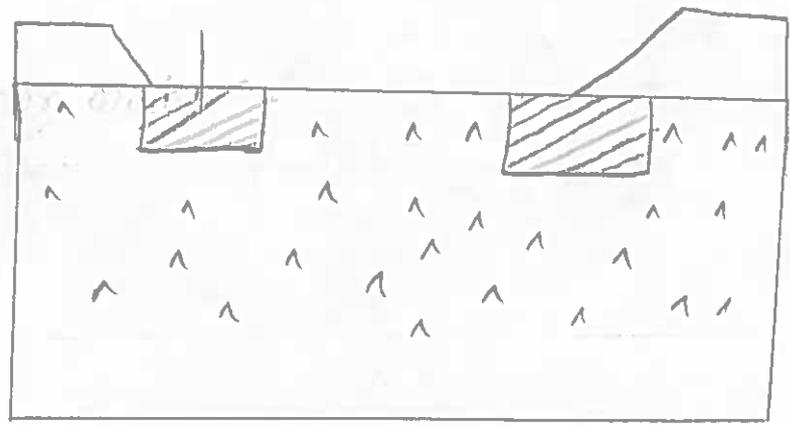
window in oxide

6



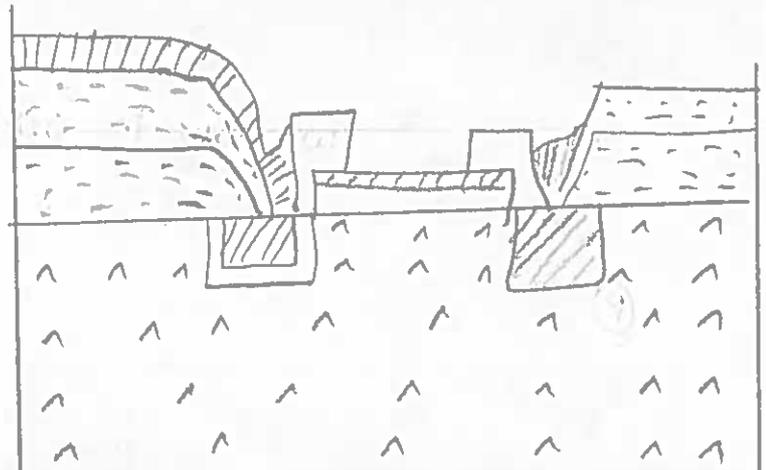
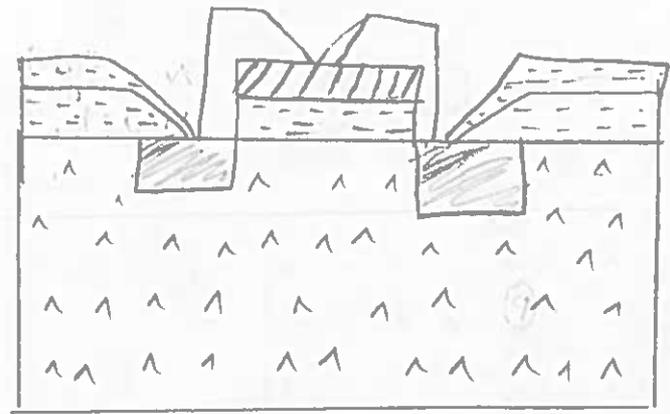
Patterned prepreg
(1-2 cm)
on thin oxide
(800-1000A)

7



In + diffusion (1 umdep)

8



Patterned metalized prepreg (Al-1um)

An nmos process is illustrated basically 9 steps, they are

- 1) processing is carried out on a thin wafer cut from a single crystal of Si of high purity into which the required p-impurities are introduced as the crystal is grown. Such wafers are typically 75 to 150 mm in diameter and 0.4 mm thick and are doped with, say, boron to impurity concentration of $10^{15}/\text{cm}^3$ to $10^{16}/\text{cm}^3$, giving resistivity in the approximate range $25 \Omega \text{ cm}$ to $2 \Omega \text{ cm}$.
- 2) A layer of (SiO_2), typically $1 \mu\text{m}$ thick, is grown all over the surface of the wafer to protect the surface, act as barrier to dopants during processing, and provide a generally insulating substrate onto which other layers may be deposited and patterned.
- 3) The surface is now covered with a photo-resist which is deposited onto the wafer

and spun to achieve an even distribution of the required thickness.

4) The photoresist layer is then exposed to ultraviolet light through a mask which defines those regions into which diffusion is to take place together with transistor channels.

→ Those areas exposed to UV radiation are polymerised (hardened), but that the areas required for diffusion are shielded by the mask and remain unaffected.

5) These areas are subsequently readily etched away together with the underlying SiO_2 so that the wafer surface is exposed in the window defined by the mask.

6) The remaining photoresist is removed and a thin layer of SiO_2 (0.1 μm typical) is grown over the entire chip surface. It is exposed in the window defined by the mask.

and the polysilicon is deposited on top of this to form gate structure.

The polysilicon layer consists of heavily doped polysilicon by ~~chemical~~ chemical vapour deposition (CVD).

7) Further photoresist coating and masking allows the polysilicon to be patterned and then the thin oxide is removed to expose areas into which n-type impurities are to be diffused. expose areas into which n-type impurities are to be diffused to form the source and drain shown. Diffusion is achieved by heating the wafer to a high temperature and passing a gas containing the desired n-type impurity (phosphorus) over the surface. Note that the polysilicon with underlying thin oxide and thick oxide acts as masks during diffusion. the process is self aligning.

8) Thick oxide (SiO_2) is grown over all again and is then masked with photoresist and etched to expose selected areas of the polysilicon gate, and the drain and source areas where connections (i.e. contact cuts) are to be made.

9) The whole chip then has metal (Al) deposited over its surface to a thickness typically of $1 \mu\text{m}$. This metal layer is then masked and etched to form the required interconnection pattern.

Summary of nmos process:-

Processing takes place on a p-doped silicon crystal wafer on which is grown a thick layer of SiO_2 .

Mask-1: Pattern SiO_2 to expose the Si surface in areas where paths in the diffusion layer or source, drain or gate areas of transistors are required. Deposit thin oxide over all. For this reason, the mask is often known as the 'thin oxide mask' (or) diffusion mask.

Mask-2: Pattern ion implantation within the thin oxide region (3n depletion mode)- Self aligning.

Mask-3: Deposit polysilicon over all (1.5 μm thick) Diffuse n⁺ regions into areas where thin oxide has been removed. Transistor drains and sources are thus self-aligning with respect to the gate structures.

Mask 4:- Grow thick oxide over all and then etch for contact cuts.

Mask 5: Deposit metal and pattern with mask 5.

Mask:6- would be required for the over glassing process step.

CMOS Fabrication.

There are a number of approaches to CMOS-fabrication, including 1) p-well, 2) n-well, 3) twin tub and 4) SOI (Silicon-on-insulator) process.

→ p-well process is widely used in practice and the n-well process is also popular.

P-well process:

A brief overview of the fabrication steps may be obtained with reference to nmos, note that the basic processing steps are same nature when compared to nmos.

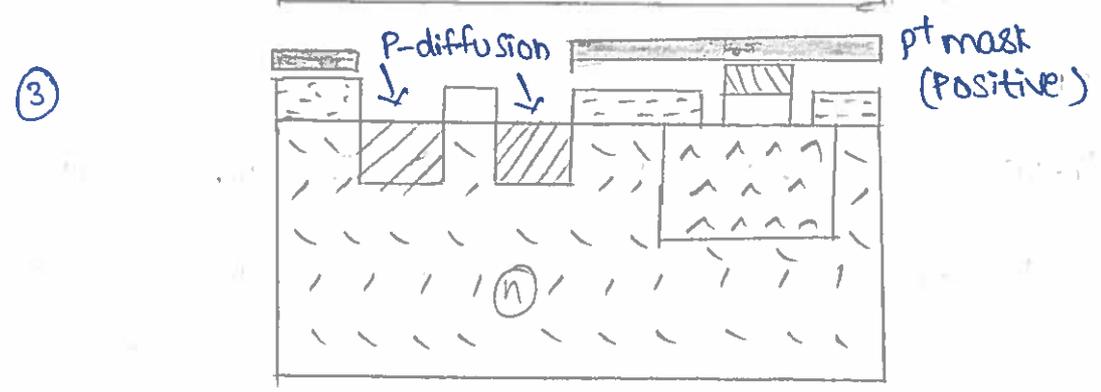
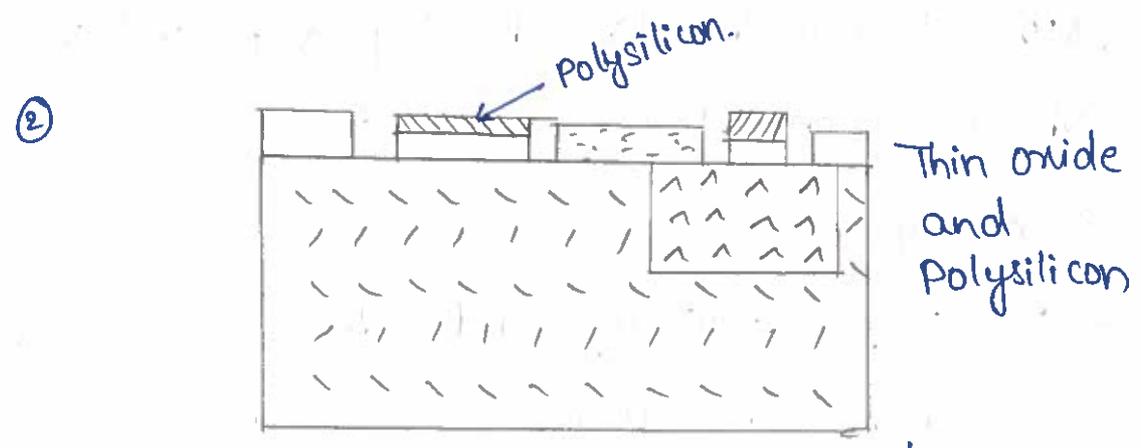
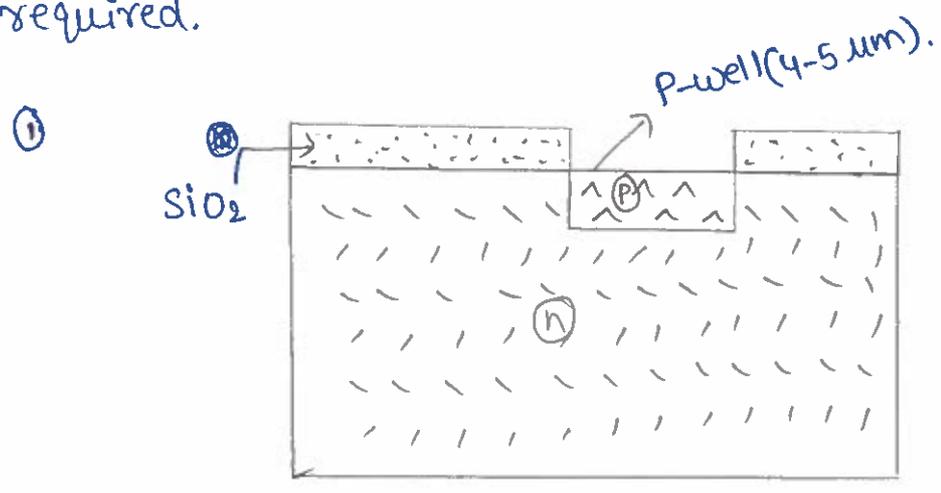
The structure of p-well, consists of an n-type substrate in which p-devices may be formed by suitable masking and diffusion and ~~indicate~~ in order to accommodate n-type devices, a deep p-well is diffused into the n-type substrate.

This diffusion must 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-transistor. to achieve low threshold voltages (0.6 to 1.0V), we need either deep well diffusion or high well ~~diffusion~~ resistivity.

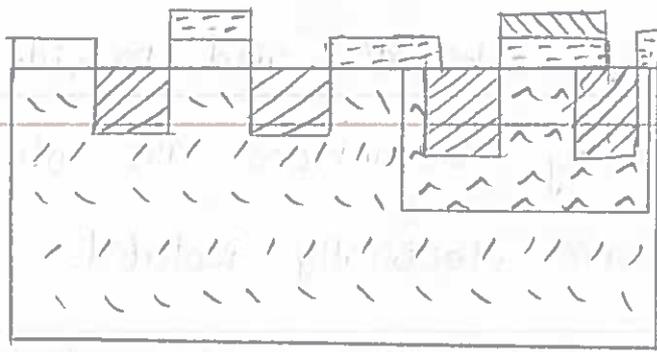
However, deep wells require larger spacing b/w the n and p-type transistors & wires.

p-well acts as substrates for the n-devices with in the n-substrate and provides that voltage Polarity restrictions are observed, the two areas are electrically isolated.

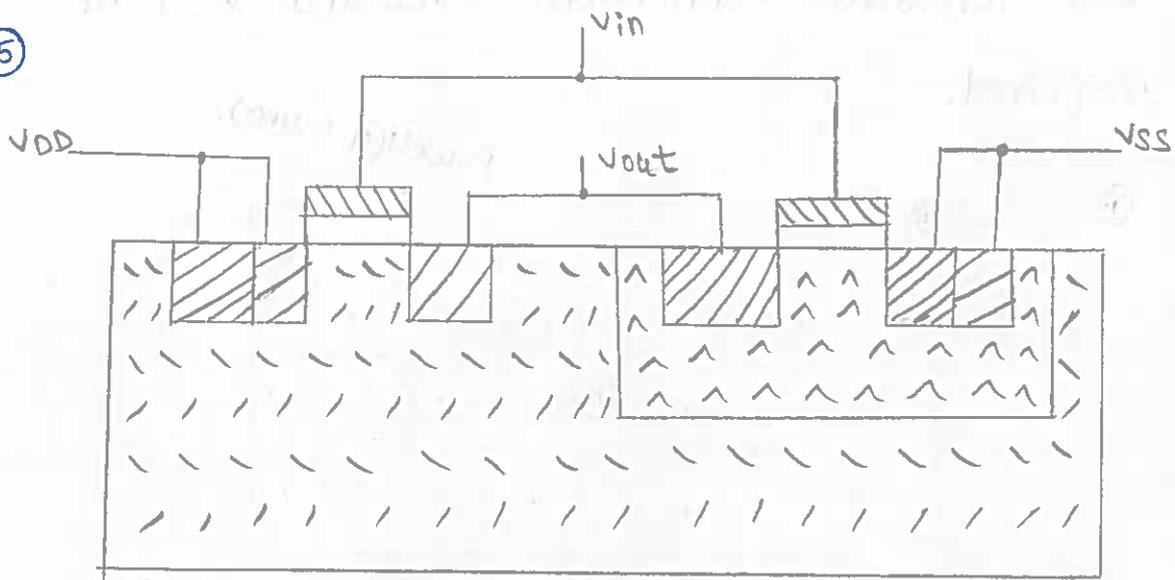
However, there are now in effect two substrates two substrate connections (V_{DD} and V_{SS}) are required.



④



⑤



CMOS P-well inverter showing V_{DD} and V_{SS} Substrate connections.

Summary

→ Masking, patterning and diffusion-are same as in nmos fabrication another processing steps are.

Mask 1: Defines the areas in which the P-well diffusions are to take place.

Mask: 2 Defines thinox regions, namely those areas where the thick oxide is to be stripped and thin oxide grown to accommodate p and n-transistors and diffusion wires.

Mask: 3 used to pattern the polysilicon layer which is deposited after the thin oxide

Mask 4: A p⁺ mask is now used to define all areas where p-diffusion is to take place

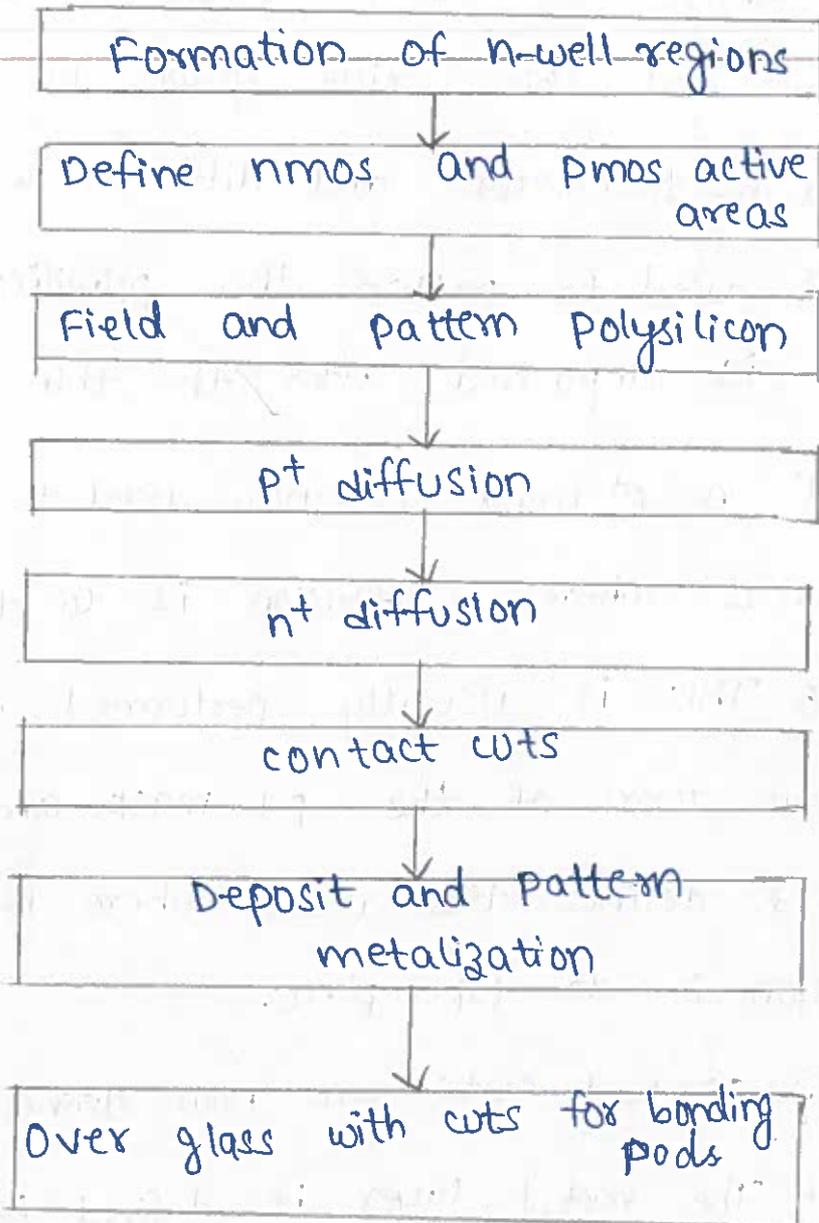
Mask 5: This is usually performed using the negative form of the p⁺ mask and, with mask 2, defines those areas where n-type diffusion is to take place.

Mask 6: contact cuts are now defined

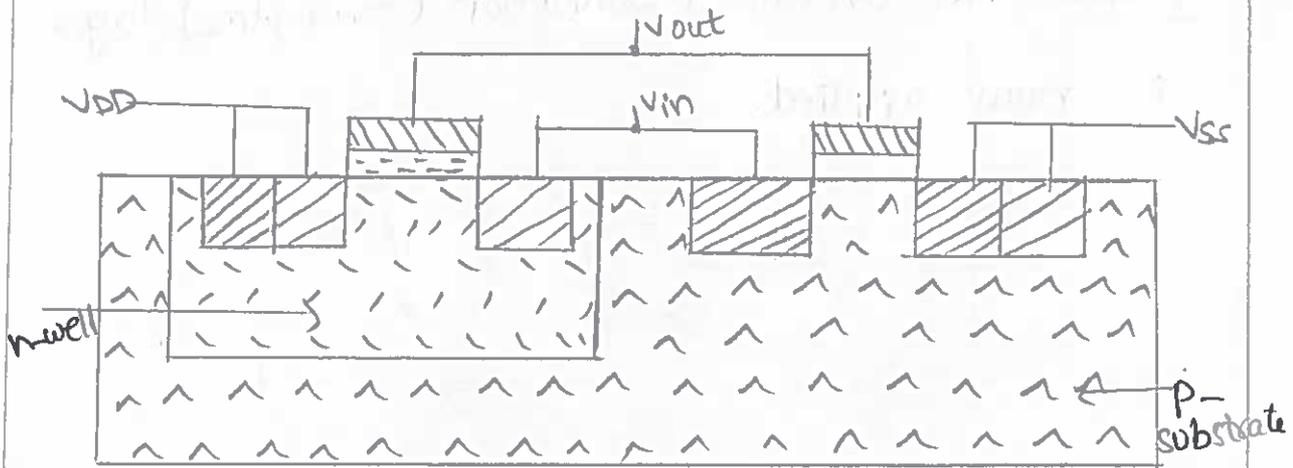
Mask 7: The metal layer is defined by this mask.

Mask 8: An overall passivation (cover glass) layer is now applied.

The n-well process:-



Main steps in a typical n-well process.

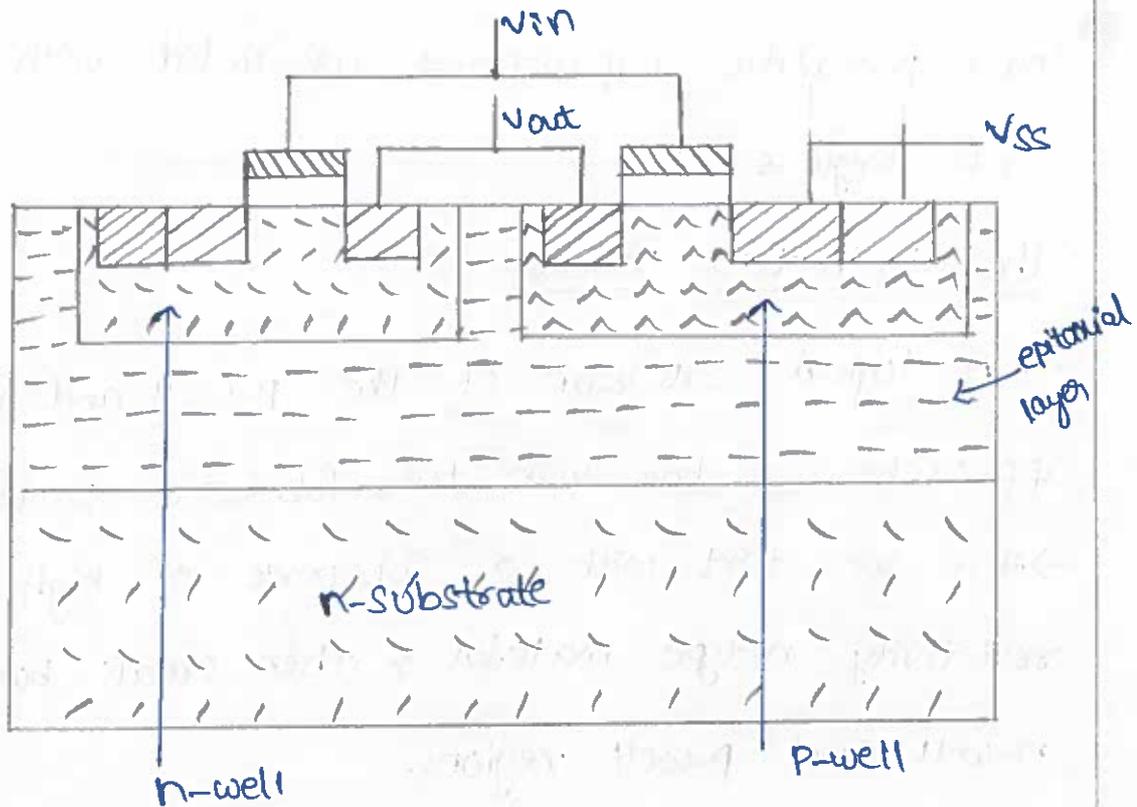


→ N-well CMOS circuits are also superior to P-well because of the lower substrate bias effects on transistor threshold voltages and lower parasitic capacitances associated with S & D regions.

The Twin-tub process:

- A logical extension of the p-well and n-well approaches is the twin tub fabrication process
- Here we start with a substrate of high resistivity n-type material & then create both n-well and p-well regions.
- Through this process it is possible to preserve the performance of n-transistors without compromising the p-transistors.
- Doping control is more readily achieved and some relaxation in manufacturing tolerances results. This is particularly important as far as latch-up is concerned.
- Latch up problems can be considerably reduced by using a low-resistivity epitaxial P-type substrate as the starting material,

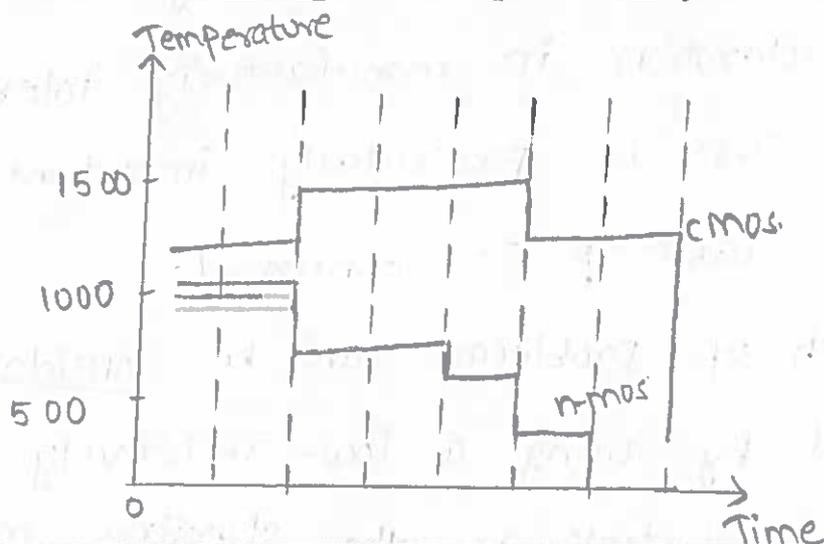
which subsequently acts as a very low resistance ground-plane to collect substrate currents.



→ n-well process performing PMOS process

→ Modern process lines are achieved by combining both P & n well process.

Thermal aspects of processing:-



→ The process involved in masking nmos & cmos devices having differing high temperature sequence.

CMOS p-well process has a high temperature p-well diffusion process ($1100-1250^{\circ}\text{C}$), the nmos process having no such requirement because simplicity and easy of fabrication.

Berkeley n-well process. (Fabrication steps).

Mask 1: Defines the phosphorous doped n-wells



Grow gate oxide, then cover with silicon nitride.



Mask 2: Define thin oxide areas above the p-substrate, leaving all n-regions covered



Nitride is selectively etched from the regions where thick oxide is desired



Boron implant is introduced to act as p-type channel stop



Field oxidation



Mask 3: Nitride layer is selectively etched above n-well.



Phosphorous implant is introduced to form n-type channel stop in the n-well



The remaining nitride layer is etched.



Implant for threshold adjustment



Heavily doped (n-type) polysilicon is deposited over wafer



Mask 4: Formation of polysilicon gates of n-channel devices



Formation of n^+ regions in thinner areas not covered by polysilicon using arsenic implant



Mask 5:

Formation of p^+ regions through boron



Thick oxide overall



Mask 6: Define contact cuts



Mask 7: Deposit 'Al' and pattern



Mask 8: Passivate and make cuts for bonding Pads.

Bicmos technology

- The drawback of CMOS technology is limited load driving capabilities of MOS transistor
- This is due to the limited current sourcing and current sinking abilities associated with both P & n-transistors.
- So Bipolar transistors also provide higher gain and have better noise and high frequency characteristics than MOS transistors.
- By using Bicmos gates, speed up the VLSI circuits.

Applications:

- ALU, ROM

→ Register file

Now considering the comparisons between the characteristics of CMOS and Bipolar ckt.

CMOS technology

- Low static power dissipation
- High input impedance (low drive current)
- Scalable V_t
- High noise margin
- High packing density
- High delay sensitivity to load
- Low output drive current
- Bidirectional capability (Drain and source are interchangeable).

Bipolar technology.

- High power dissipation
- Low input impedance (High drive current).
- Low voltage swing logic
- Low packing density
- Low delay sensitivity to load
- High g_m ($g_m \propto e^{V_{in}}$)
- High f_t at low currents
- Essentially unidirectional.

To achieve n-well Bicomos technology, combining n-well cmos process with two additional layers

→ 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, buried n^+ subcollector (BCSD) is added to reduce the n-well (collector) resistance and thus improve the quality of the bipolar transistor.

Bicomos fabrication process (n-well)

→ Fabrication of Bicomos is similar to the cmos but extra two additional processing steps are

- (i) p^+ base region
 - (ii) n^+ collector
 - (iii) Buried collector
- } masks.

n-well BiCMOS fabrication process steps:

single poly. single metal CMOS	Additional steps for bipolar devices
<ul style="list-style-type: none">→ Form n-well→ Define active areas→ channel stop→ Threshold V_t adjustment→ Define poly gates→ Form n^+, p^+ active areas→ Define contacts→ Define metal areas	<ul style="list-style-type: none">→ Form buried n^+ layer (BCCO)→ Form deep n^+ collector→ Form p^+ base for bipolar

Applications:

CMOS → For logic gates implementation

Bi CMOS → For I/O & driver circuits (Buffer).

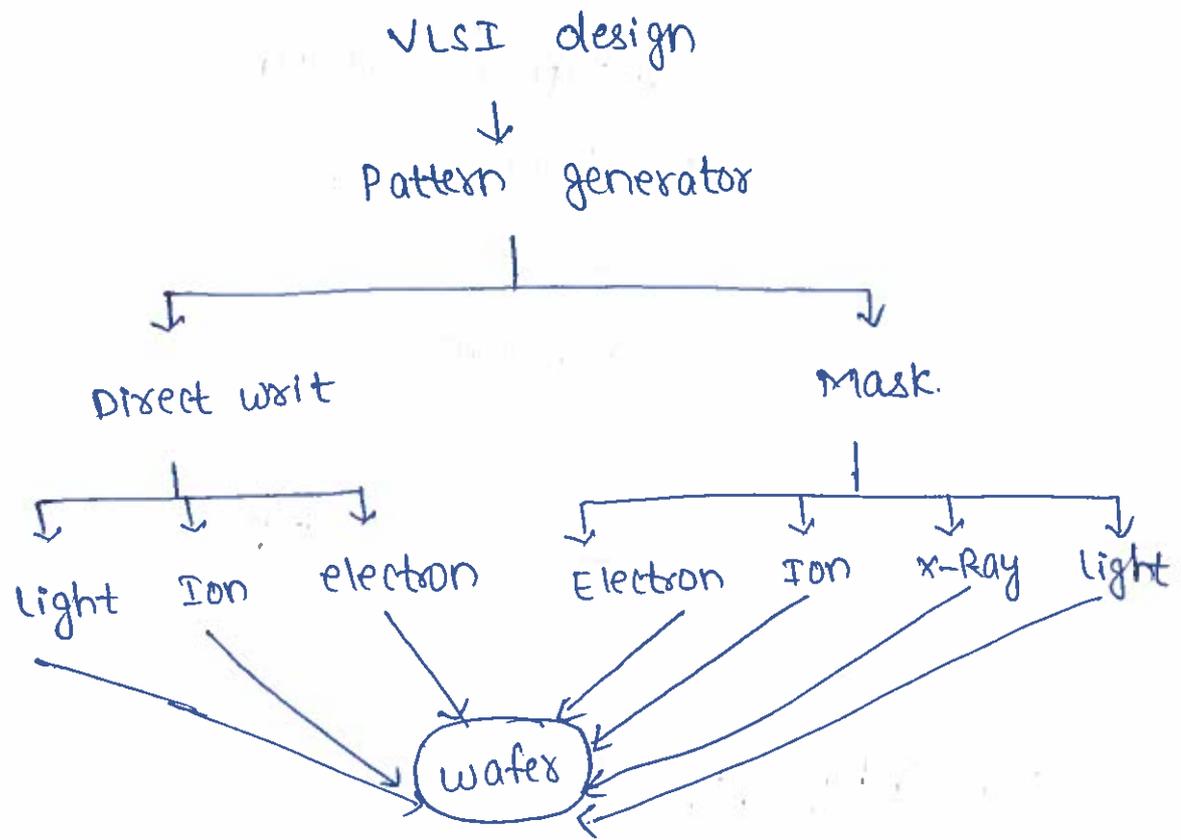
Lithography: (Printing method).

→ It is a process of transferring patterns of geometric shapes in a mask to a layer of radiation sensitive (resist) material for covering surface of semiconductor wafer.

→ various lithographic process employed in integrating VLSI patterns on substrate.

Lithograph:

A printed image produced by lithography an image produced by etching the image onto a flat surface, then copying the etched surface by applying ink to it and pressing another material against it.



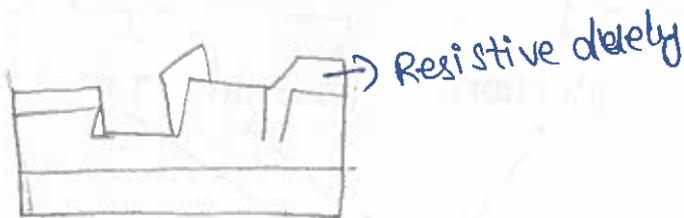
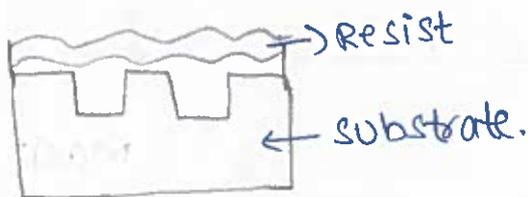
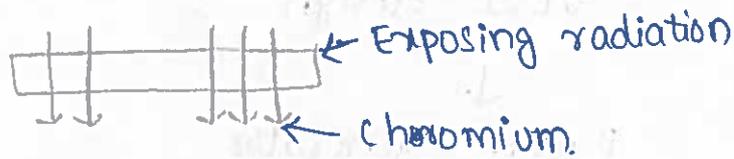
→ In this process, impressing several flat embossed slabs onto a paper. Every slab has

definite colour. after this work, the ckt patterns are transferred on wafers.

Optical lithography:

→ The formation of patterns/images in a photoresist in optical lithography due to visible or uv radiations which utilizes proximity or projection printing

→ The exposing radiation is transmitted through the clear mask transferring patterns on substrate.



Electron lithography:

The electron lithography has much higher resolution compared to optical lithography

→ The electron lithography is mostly used for pattern generates functions such as:

Mask fabrication, direct reaction with some material wafer and direct on wafer.

→ These are two types of electron lithography

- scanning type
- projection type

→ The output format from the CAD system is converted into format of individual systems.

The data is then decomposed into simple elements such as trapezoidal or rectangles by electron exposure machines

Diffusion:

The process of introducing controlled amount of dopants into semi conductors is called "Diffusion."

The conductivity of Si is changed because of diffusion producing selectively n-type and P-type regions requires that diffusion to be carried out at high temperature.

- Diffusion is carried out at $800-1200^{\circ}\text{C}$.
- Diffusion of boron into Si produces p-type Si and diffusion of Arsenic or Phosphorus into Si produces n-type a semi conductor.

Ion Implantation:

The process of introducing high energy charged particles into the substrate called "Ion Implantation". It produces extremely thin, selectively placed doped layers. Ion implantation is preferred when small areas are to be doped at lower temperatures.

Implantation Mechanism:

A beam of implanted ions with desired charge to mass ratio are accelerated ratio. These ions collides with the silicon wafer and loses energy. The loss of energy takes place by 2 mechanisms.

- 1) Nuclear Stopping.
- 2) Electronic Stopping.

→ The ion current during implantation is used for controlling dopant dose. The damage to silicon lattice is fulfilled by annealing treatment.

Metallization:-

Metallization is process of formation of metal ^{films} for interconnections, ohmic contacts and rectifying metal semiconductor contacts.

→ The methods used for making metal films are chemical vapour Deposition (CVD) and physical vapour deposition (PVD).

Requirements of metallization for VLSI application:

- Low Resistivity
- Ease of pattern generation
- Not contaminate devices
- Mechanical stability
- Provide good device contact
- Low stress

Role of Metallization in IC fabrication:-

- Metallization controls speed of ckt by its interconnection resistance of the contact.
- It decides the threshold voltage of Mos.
- The resistance of metallization cause the ohmic contact. The ohmic contact should have least resistance and must be stable electrically and mechanically.
- Metallization provides a connection to outside world.

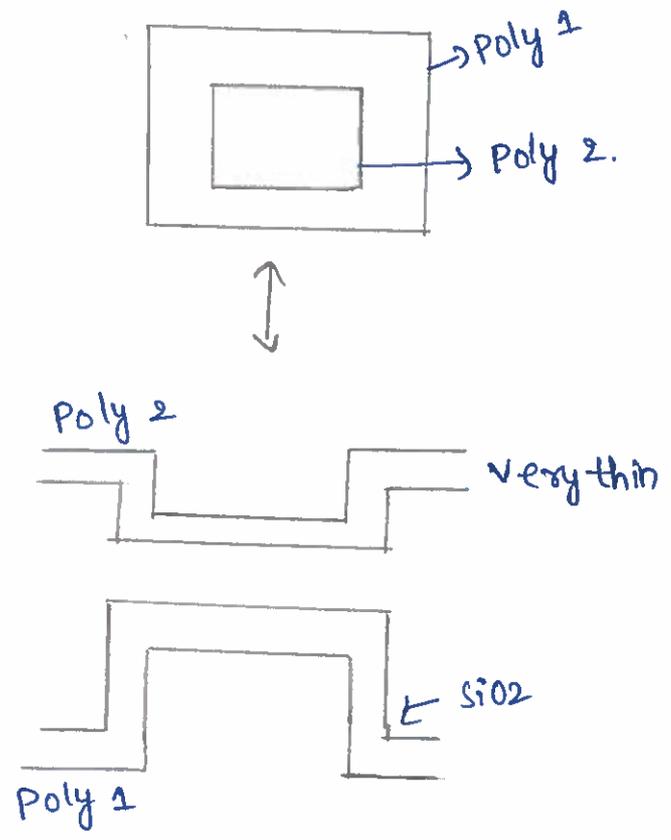
Integrated Resistors and capacitors:-

Resistors:-

- An undoped 'si' highly resistive, this property is used to fabricate resistors.
- In mixed signal CMOS, a resistive metal nichrome is added to high value, quality resistors.

capacitors:

- These are required for switched capacitor analog ckts and dynamic memory cells.
- These capacitors are fabricated by adding one extra layer of polysilicon by different techniques.



Unit 1

Chapter 1

Section 1.1

Page 1

1.1.1

1.1.2

1.1.3

1.1.4

1.1.5

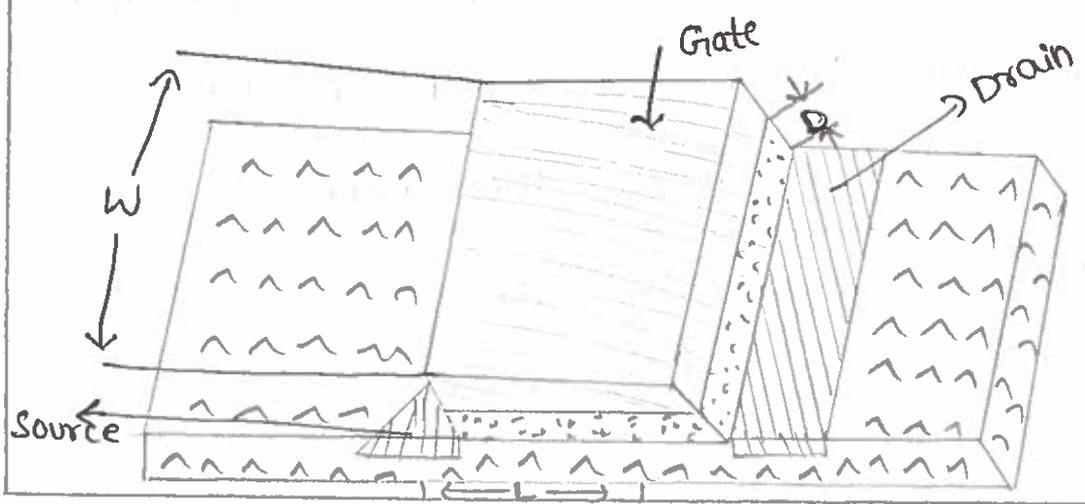
Basic electrical properties of MOS and BiCMOS circuits:-

Drain to source current I_{ds} versus voltage V_{ds} relationships:

The whole concept of the MOS transistor develops from the use of a voltage on the gate to induce a charge in the channel b/w source and drain, which may then be caused to move from source to drain under the influence of an \vec{e} field created by voltage V_{ds} applied b/w drain and source.

Here charge induced is dependent on v_{gs} , then I_{ds} is dependent on both v_{gs} & V_{ds} .

$$I_{ds} = -I_{sd} = \frac{\text{charge induced in channel } (C_{ch})}{\text{Electron transit time } (\tau)} \quad \text{---(1)}$$



First, transit time: $\tau_{sd} = \frac{\text{length of the channel}(L)}{\text{velocity}(v)}$

but velocity $v = \mu \cdot E_{ds}$

$\mu \rightarrow$ electron or hole mobility

$E_{ds} \rightarrow$ Electric field (d \rightarrow s)

$$\text{Now } E_{ds} = \frac{V_{ds}}{L}$$

$$\therefore v = \mu \cdot \frac{V_{ds}}{L}$$

$$\therefore \text{Thus } \tau_{sd} = \frac{L}{\mu \cdot V_{ds}/L} \Rightarrow \frac{L^2}{\mu \cdot V_{ds}} \quad \text{--- (2)}$$

Typical values of μ at room temp are

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

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

The non-saturated region:

charge induced in channel due to v_g is due to voltage difference b/w the gate and channel (v_{gs}) (assume substrate connected to source)

Now the voltage along the channel varies linearly with distance from the source due to IR drop in the channel.

and assume the device is not saturated then the avg value is $\frac{V_{ds}}{2}$.

Further more eff gate $V_g = V_{gs} - V_t$

$V_t \rightarrow$ Threshold voltage

Note that the charge/unit area = $E_g \epsilon_{ins} \epsilon_0$

Thus induced charge $Q_c = E_g \epsilon_{ins} \epsilon_0 WL$

where $E_g \rightarrow \bar{e}$ field

$\epsilon_{ins} \rightarrow$ relative permittivity of insulation b/w gate & channel

$\epsilon_0 \rightarrow$ Relative permittivity of free space

$W \rightarrow$ width of the gate

$L \rightarrow$ length of the channel

$(\epsilon_0 = 8.85 \times 10^{-14} \text{ F/cm})$

$(\epsilon_{ins} = 4 \text{ for SiO}_2)$

Now $E_g = \frac{[(V_{gs} - V_t) - \frac{V_{ds}}{2}]}{D}$

$D \rightarrow$ oxide thickness

Thus $Q_c = \frac{\epsilon_{ins} \epsilon_0 WL}{D} [(V_{gs} - V_t) - \frac{V_{ds}}{2}] \quad \text{--- (3)}$

now combining eq 2 & 3 in 1

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

$$\text{Let } k = \frac{\epsilon_{ins} \epsilon_0 \mu}{D}$$

$$B = \frac{kW}{L}$$

$$\therefore I_{ds} = B \left((V_{gs} - V_t) V_{ds} - \frac{V_{ds}^2}{2} \right) \quad \text{--- (4)}$$

And alternative form eq (4) in the form of gate/channel capacitance

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

$$\therefore k = \frac{\epsilon_{ins} \epsilon_0 \mu}{D} \Rightarrow \frac{C_g \mu}{WL}$$

$$I_{ds} = \frac{C_g \mu}{2L} \left((V_{gs} - V_t) V_{ds} - \frac{V_{ds}^2}{2} \right) \quad \text{--- (5)}$$

Sometimes it is convenient to use gate capacitance/unit area C_0 rather than C_g .

$$\therefore C_g = C_0 W L$$

$$I_{ds} = \frac{C_0 \mu W}{L} \left((V_{gs} - V_t) V_{ds} - \frac{V_{ds}^2}{2} \right) \quad \text{--- (6)}$$

The saturated region:

Saturation begins when $V_{ds} = V_{gs} - V_t$ since at this point the IR drop in the channel

equals effective gate to channel voltage at the drain and we may assume that the current remains fairly constant as V_{DS} increases further.

$$\text{Thus } I_{DS} = \frac{kW}{L} \frac{(V_{GS} - V_t)^2}{2}$$

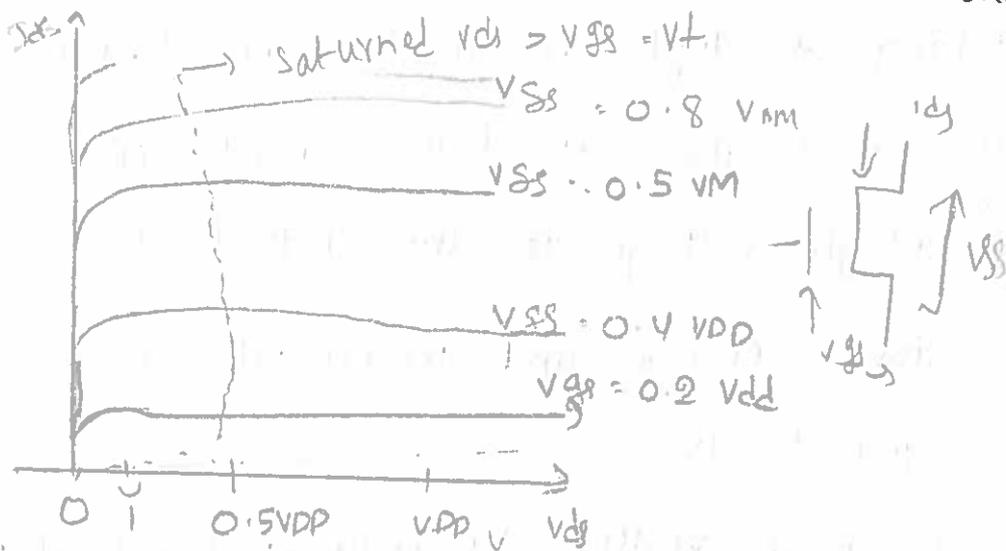
we may write

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

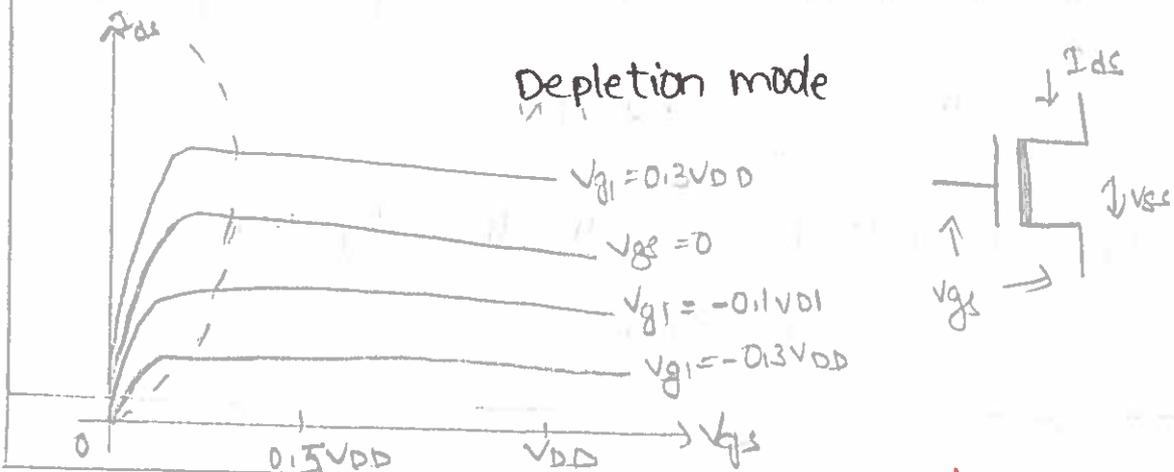
we may also write

$$I_{DS} = C_0 \mu \frac{W}{2L} (V_{GS} - V_t)^2 \quad \text{--- (A)}$$

Typical characteristics:- Enhancement mode



Depletion mode



Aspects of MOS transistor threshold voltage V_t .

The gate structure of a MOS transistor consists, electrically, 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 on state consists in applying sufficient gate voltage to neutralize these charges and enable the Si to inversion due to \bar{E} 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

$$V_t = \phi_{ms} \cdot \frac{Q_B - Q_{ss}}{C_0} + 2\phi_{FN}$$

$Q_B \rightarrow$ charge/unit area in the depletion layer beneath the oxide

Q_{ss} → charge density at Si: SiO₂ interface

C_0 → capacitance / unit gate area

ϕ_{ms} → work function difference b/w gate & Si.

Now, for polysilicon gate and Si substrate, the value of ϕ_{ms} is negative but negligible.

and the magnitude and sign of V_t are thus determined by the balance b/w the remaining negative term - $\frac{Q_{ss}}{C_0}$ and other two terms, both of which are +ve.

To evaluate V_t each term is determined as follows:

$$Q_B = \sqrt{2 \epsilon_0 \epsilon \text{ sign } (2 \phi_{fn} + V_{SB})} \text{ coulomb/m}^2.$$

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

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

depending on crystal orientation and where

V_{SB} = substrate bias voltage (-ve w.r.t source for nmos, +ve for pmos)

$$q = 1.6 \times 10^{-19} \text{ C}$$

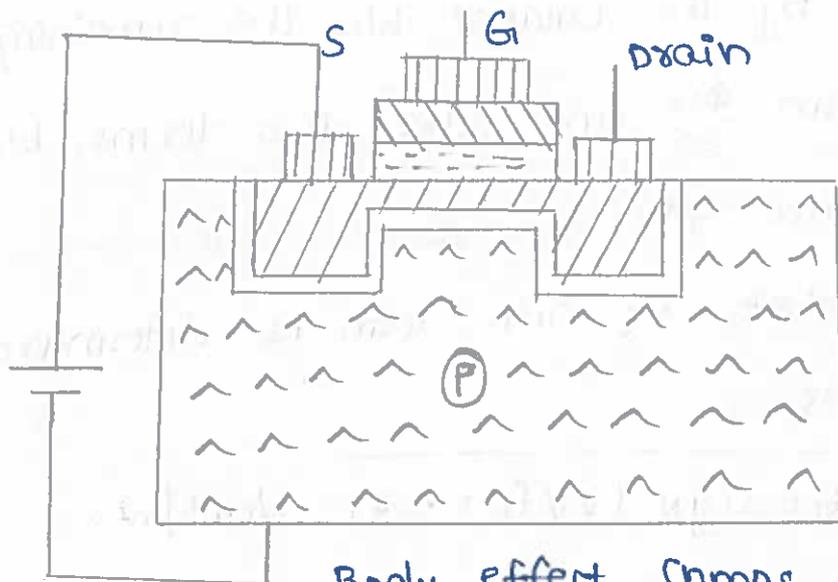
$N =$ impurity concentration in the substrate.

(N_A or N_D as appropriate).

$\epsilon_{Si} =$ relative permittivity of Si = 11.7.

$n_i =$ intrinsic electron concentration ($1.6 \times 10^{10} / \text{cm}^3$ at 300 K)

$k =$ Boltzmann's constant = 1.4×10^{-23} Joule/K.



Body effect (nmos device)

The body effects may also be taken into account since the substrate may be biased w.r.t to source.

Increasing V_{SB} causes the channel to be depleted charge carriers thus the V_t is raised

change in V_t is given by $\Delta V_t = \gamma (V_{SB})^{1/2}$

where $\gamma = \text{constant}$.

which depends on substrate doping so that more lightly doped the substrate, the smaller will be the body effect.

Alternatively, we may write.

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

where $V_t(0)$ is threshold voltage for $V_{SB} = 0$

To establish the magnitude of such effects, typical figures for V_t are as follows

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

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

MOS transistor trans conductance (g_m) and output conductance (g_{ds}).

Trans conductance expresses the relationship b/w output current I_{ds} and the input voltage V_{gs} and is defined as

$$g_m = \frac{\delta I_{ds}}{\delta v_{gs}} \Big|_{v_{ds} = \text{constant}}$$

To find an expression for g_m in terms of circuit and transistor parameters, consider that the charge in the channel Q_c is such that

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

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

Thus change in current $\delta I_{ds} = \frac{\delta Q_c}{\tau}$

$$\text{Now } \tau_{ds} = \frac{L^2}{\mu \cdot v_{ds}}$$

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

but change in charge $\delta Q_c = c_g \delta v_{gs}$

$$\text{So that } \delta I_{ds} = \frac{c_g \delta v_{gs} \mu v_{ds}}{L^2}$$

$$\text{Now } g_m = \frac{\delta I_{ds}}{\delta v_{gs}} = \frac{c_g \mu v_{ds}}{L^2}$$

In saturation $v_{ds} = v_{gs} - v_t$

$$g_m = \frac{c_g \mu}{L^2} (v_{gs} - v_t) \quad \text{--- (1)}$$

and substituting for $c_g = \frac{\epsilon_{ins} \epsilon_0 W L}{D}$

$$g_m = \frac{\mu \epsilon_{ins} \epsilon_0}{D} \frac{W}{L} (V_{gs} - V_t) \quad \text{--- (2)}$$

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

It is possible to increase the g_m of a MOS device by increasing its width. However, this will also increase the i/p capacitance and area occupied.

A reduction in the channel length results in an increase in w_0 owing to the higher g_m .

However, the gain of the MOS device decreases owing to the strong degradation of the o/p resistance = $1/g_{ds}$.

The o/p conductance g_{ds} can be expressed by

$$g_{ds} = \frac{\delta I_{ds}}{\delta V_{gs}} = \lambda \cdot I_{ds} \propto \left(\frac{1}{L}\right)^2$$

Here the strong dependence on the channel length is demonstrated as $\lambda \propto \left(\frac{1}{L}\right)$ and $I_{ds} \propto \left(\frac{1}{L}\right)$ for the MOS device.

MOS transistor figure of merit ω_0 :

An indication of frequency response may be obtained from the parameter ω_0 where

$$\omega_0 = \frac{g_m}{C_g} = \frac{\mu}{L^2} (V_{gs} - V_t) \left(= \frac{1}{\tau_{sd}} \right) \quad - \textcircled{1}$$

This shows that switching speed depends on gate voltage above threshold and on carrier mobility and inversely as the square of channel length.

As fast circuit requires that g_m be as high as possible. Electron mobility on a (100) oriented n-type inversion layer surface (μ_n) is larger than that on a (111) oriented surface, and is in fact about three times as large as hole mobility on a (111) oriented p-type inversion layer. Surface mobility is also dependent on the effective gate voltage $(V_{gs} - V_t)$.

For faster nmos circuits, then one would choose a (100) oriented p-type substrate.

in which the inversion layer will have a surface carrier mobility $\mu_n = 650 \text{ cm}^2/\text{V sec}$ at room temp.

compare this with the typical bulk mobilities

$$\mu_n = 1250 \text{ cm}^2/\text{V sec}$$

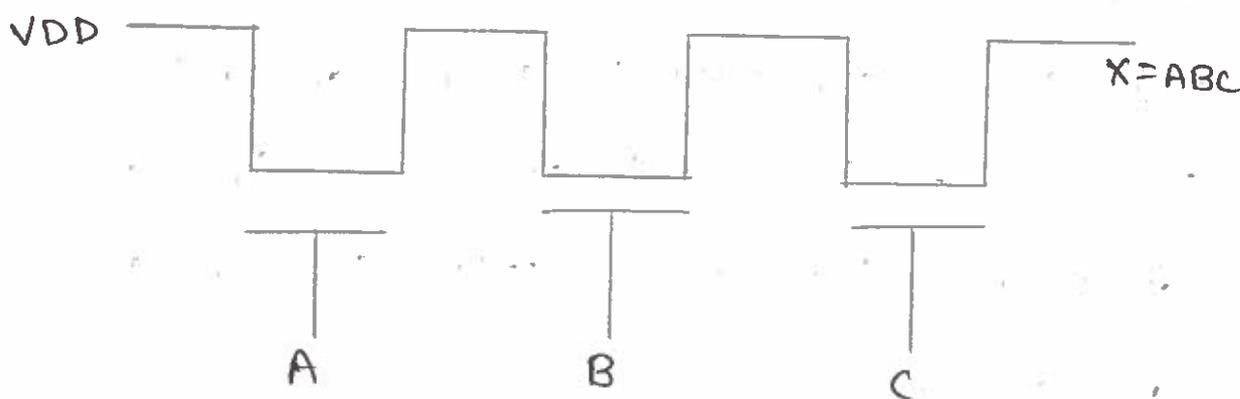
$$\mu_p = 480 \text{ cm}^2/\text{V sec}$$

From which it will be seen that $\frac{\mu_s}{\mu} = 0.5$ where $\mu_s =$ surface mobility and $\mu =$ bulk mobility

The pass transistor:

MOS transistors can be used as switches by virtue of the isolated nature of the gate.

The application of the MOS device is called the pass transistor and switching logic arrays.



→ Pass transistor used as switch and can transfer logic levels from one point to another

→ used as Transmission gate (Application)

→ MUX

• → XOR

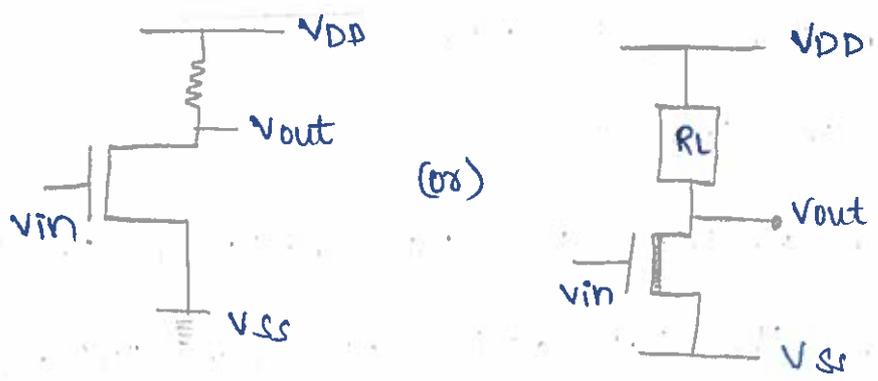
→ DFIF

Inverters

Nmos Inverter:

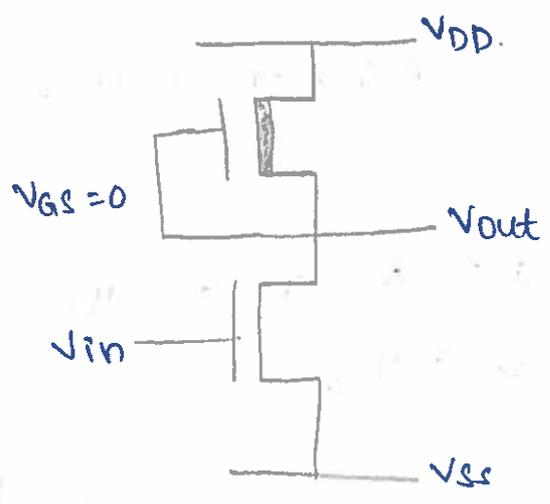
→ A basic requirement for producing a complete range of logic circuits is the inverter. This is needed for restoring logic levels for NAND and NOR gates, sequential and memory circuits of various forms.

* The basic inverter circuit requires a transistor with source connected to ground and a load resistor of some sort connected from the drain and the input applied between gate and ground.



Resistors are not conveniently produced on the 'Si' substrate even modest values and it occupies large space.

Therefore some other form of RL is required and one of the convenient way to solve this problem is to use a D.M transistor as the load



fig(2): nmos inverter.

Now,
 → with no current drain wfrom o/p, the Ids for both transistor must be equal.

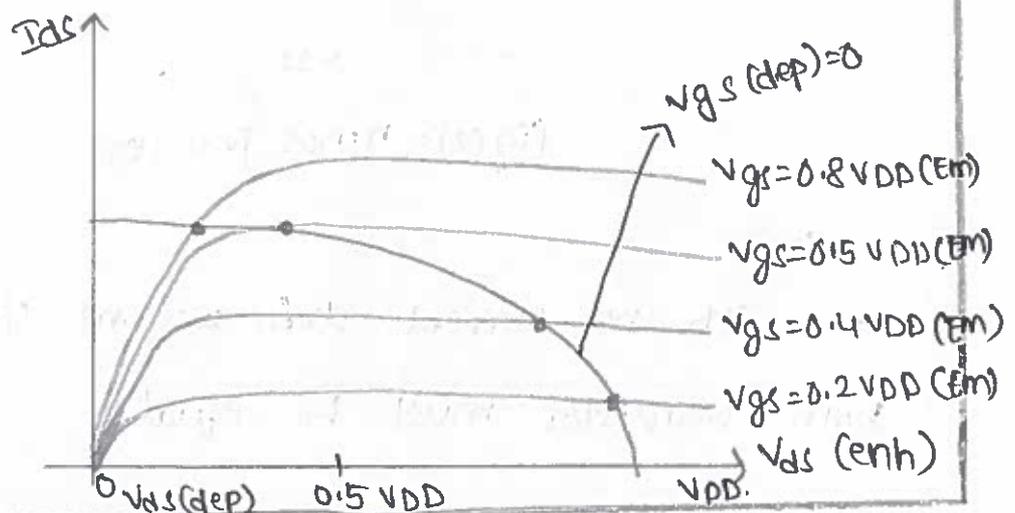
→ For D.M transistor $G \rightarrow$ is connected to source so it is always on ($v_{gs}=0$)

→ In this configuration the D.M device is called Pull-up (PU) and the E.M device is called pull-down transistor

→ To obtain the inverter transfer characteristics we superimpose the $v_{gs}=0$, D.M characteristics curve on the family of enhancement mode characteristics curve.

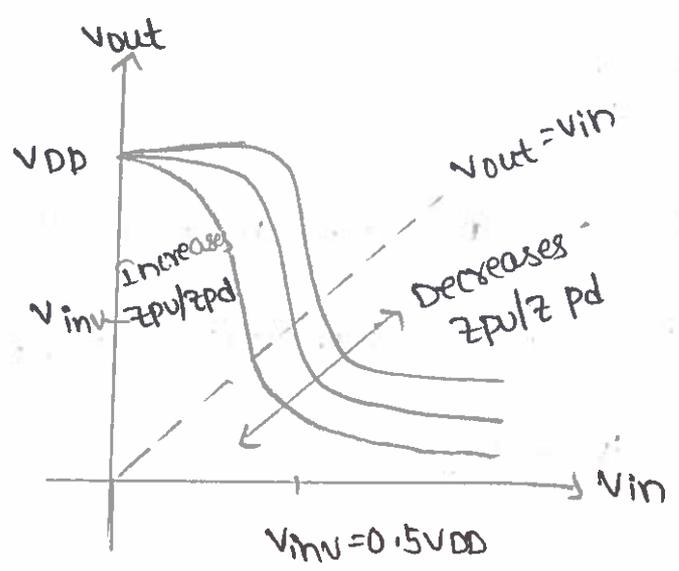
Note that the Max voltage across EM ~~compose~~ corresponds to minimum voltage across D.M transistor.

→ The points of intersection of the curves gives transfer characteristics.



$V_{ds} (EM) = V_{DD} - V_{ds} (dep) = V_{out}$

$V_{gs} (Enh) = V_{in}$



Fig(4) nmos inverter ch transfer characteristics

- Note that as $V_{in} = V_{gs}$ (P.d transistor) exceeds V_{th} current begins to flow.
- The V_{out} thus decreases and subsequent increases in V_{in} will cause the P.d transistor to come out of sat and become resistive.
- Note that P.u transistor is initially resistive as the P.d turns on
- During the transition, the slope of characteristics determines $Gain = \frac{\delta V_{out}}{\delta V_{in}}$

The point at $v_{in} = v_{out}$ is denoted as v_{inv} shifted by pull-up and pull-down resistances

$$\frac{z_{PU}}{z_{PD}} \Rightarrow z = \frac{L}{W}$$

* Determination of pull-up to pull-down ratio $\left(\frac{z_{PU}}{z_{PD}}\right)$ for an nmos inverter driven by another nmos inverter.



Fig: nmos inverter driven directly by another inverter.

consider an arrangement, in which an inverter, is driven from the output of another similar inverter

consider the D.M (Depletion mode) transistor for which $v_{gs} = 0$. under all conditions, further assume that in order to cascade inverters without degradation of levels we are aiming to meet the requirement.

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

For equal margins around the inverter threshold we set $V_{inv} = 0.5 V_{DD}$.

At this point both transistors are in

Saturation and
$$I_{ds} = k \cdot \frac{W}{L} \frac{(V_{gs} - V_t)^2}{2}$$

In the D.M $V_{gs} = 0$

$$\therefore I_{ds} = \frac{k \cdot W_{PU}}{L_{PU}} \left(\frac{-V_{td}}{2} \right)^2$$

and in the E.M

$$I_{ds} = k \cdot \frac{W_{PD}}{L_{PD}} \left(\frac{V_{inv} - V_t}{2} \right)^2 \quad (V_{gs} = V_{inv})$$

Equating (since currents are the same) we have

$$k \cdot \frac{W_{PD}}{L_{PD}} \left(\frac{V_{inv} - V_t}{2} \right)^2 = k \cdot \frac{W_{PU}}{L_{PU}} \left(\frac{-V_{td}}{2} \right)^2$$

where W_{PD} , L_{PD} , W_{PU} , L_{PU} are the widths and lengths of pull down and pull-up transistors respectively.

$$Z_{PD} = \frac{L_{PD}}{W_{PD}}, \quad Z_{PU} = \frac{L_{PU}}{W_{PU}}$$

we have
$$\frac{1}{Z_{PD}} (V_{inv} - V_t)^2 = \frac{1}{Z_{PU}} (-V_{td})^2$$

Now substitute $V_{in} = V_{inv} = 0.5 \cdot V_{DD}$

$$V_t = 0.2 V_{DD}$$

$$V_{td} = -0.6 V_{DD}$$

$$\frac{1}{z_{pd}} (V_{inv} - V_t)^2 = \frac{1}{z_{pu}} (-V_{td})^2$$

$$\frac{(V_{inv} - V_t)^2}{z_{pd}} = \frac{(-V_{td})^2}{z_{pu}}$$

$$\text{Hence } V_{inv} = V_t - \frac{V_{td}}{\sqrt{z_{pu}/z_{pd}}}$$

$$V_t = 0.2 V_{DD} \quad V_{td} = -0.6 V_{DD}$$

$$V_{inv} = 0.5 V_{DD}$$

$$0.5 = 0.2 + \frac{0.6}{\sqrt{z_{pu}/z_{pd}}}$$

$$\sqrt{z_{pu}/z_{pd}} = 2$$

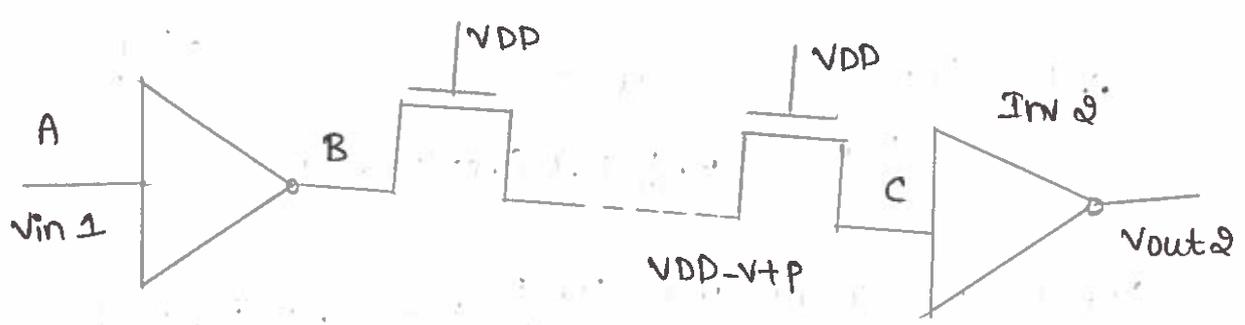
$$\text{and } \sqrt{z_{pu}/z_{pd}} = \frac{z_{pu}}{z_{pd}} = \frac{4}{1}$$

For an inverter directly driving another inverter by $\frac{4}{1}$.

* Pull-up to pull-down ratio for an nmos inverter driven through one or more pass transistors.

consider the arrangement, in which the input to inverter 2 comes from the output of inverter 1 but passes through one or more nmos transistors used as switches in series (called pass transistors).

connection of pass transistors in series will degrade the logic 1 level into inverter 2 so that output will not be a proper logic '0' level. The critical condition is when point A is at 0 volts & B is at VDD, but the voltage into inverter 2 at point 'c' is now reduced from VDD by threshold voltage of the series pass transistors



* Pull-up to pull down ratio for inverting logic coupled by pass transistor. with all pass transistor gates connected to VDD, there is a loss of V_{tp} , therefore, the i/p voltage to inv 2 is

$$V_{in\ 2} = V_{DD} - V_{tp}$$

$V_{tp} \rightarrow$ Threshold voltage for a Pass transistor.

we must now ensure that for this i/p voltage we get out, the same voltage as would be the case for inverter 1 driven with $i/p = V_{DD}$.

Consider inverter 1 with $i/p = V_{DD}$

If the i/p is at V_{DD} , then the P-D transistor T_2 is conducting but a low voltage across it.

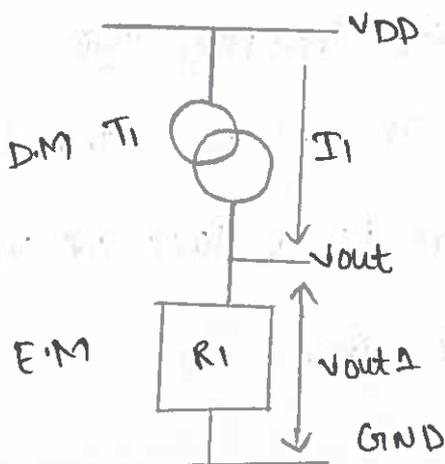
\therefore It is in resistive region represented by R_1 .

Mean while, the P-U transistor T_1 is in saturation and is represented as current source.

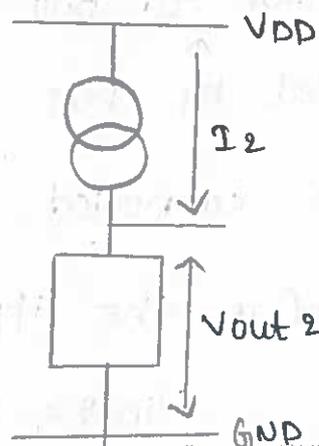
For the P-d transistor, $I_{D2} = k \cdot \frac{W_{Pd}}{L_{Pd}} (V_{DD} - V_t) V_{ds1}$

$$\therefore R_1 = \frac{V_{ds1}}{I_{D2}} \Rightarrow \frac{V_{ds1}}{k \cdot \frac{1}{2} P_{d1} (V_{DD} - V_t - \frac{V_{ds1}}{2})}$$

V_{ds1} is small and $\frac{V_{ds1}}{2}$ may be ignored



InV 1 with $v_{in} = V_{DD}$



InV 2 with $i/p = V_{DD} - V_{tp}$

Thus $R_1 = \frac{1}{k} z_{pd1} \left(\frac{1}{V_{DD}-V_t} \right) \quad \text{--- (1)}$

Now, for DM P.U in saturation with $V_{gs}=0$

$I_1 = I_{ds} = k \cdot \frac{W_{PU1}}{L_{PU1}} \frac{(V_{td})^2}{2} \quad \text{--- (2)}$

The product $I_1 R_1 = V_{out1}$

$V_{out1} = \frac{z_{pd1}}{z_{pu1}} \frac{(V_{td})^2}{(V_{DD}-V_t)} \quad \text{--- (3)}$

consider Inv2.

when i/p $V_{in} = V_{DD} - V_{tp}$

AS for inverter 1

$R_2 = \frac{1}{k} z_{pd2} \frac{1}{(V_{DD}-V_{tp})-V_t}$

$I_2 = k \cdot \frac{1}{z_{pu2}} \frac{(V_{td})^2}{2}$

$\therefore V_{out2} = I_2 R_2 = \frac{z_{pd2}}{z_{pu2}} \left(\frac{(V_{td})^2}{2(V_{DD}-V_{tp}-V_t)} \right) \quad \text{--- (4)}$

If inv2 is to have same o/p voltage under these conditions then $V_{out1} = V_{out2}$

That is $I_1 R_1 = I_2 R_2$

$\therefore \frac{z_{pu2}}{z_{pd2}} = \frac{z_{pu1}}{z_{pd1}} \frac{(V_{DD}-V_t)}{(V_{DD}-V_{tp}-V_t)}$

Here substitution typical values of $V_T = 0.2 V_{DD}$

$$V_{TP} = 0.3 V_{DD}$$

$$\therefore \frac{z_{PU2}}{z_{PD2}} = \frac{z_{PU1}}{z_{PD1}} \left(\frac{V_{DD} - 0.2 V_{DD}}{V_{DD} - 0.3 V_{DD} - 0.2 V_{DD}} \right)$$

$$= \frac{z_{PU1}}{z_{PD1}} \left(\frac{0.8}{0.5} \right)$$

$$= 2 \left(\frac{z_{PU1}}{z_{PD1}} \right)$$

$$= 2 \times \frac{4}{1} = \frac{8}{1}$$

$$\therefore \frac{z_{PU1}}{z_{PD1}} = \frac{4}{1}$$

Summarizing for CMOS inverters

→ An inverter driven directly from o/p of another, should have a z_{PU}/z_{PD} ratio of $\geq 4:1$

→ An inverter driven through one or more pass transistors, should have a $\frac{z_{PU}}{z_{PD}}$ ratio of $\geq 8:1$.

Alternative forms of pull-up (various pull-ups)

Up to now we have assumed that the inverter ckt has a depletion mode pull-up transistor as its load. There are 4 least possible arrangements.

1) Load resistance (R_L): This arrangement is not often used because of the large space requirement of resistors produced in a silicon substrate.

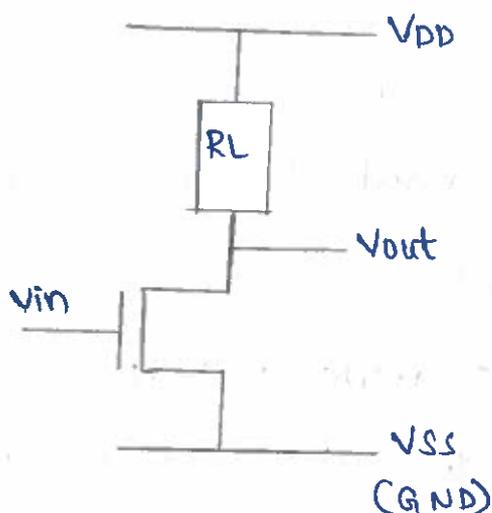


Fig (i) Resistor pull-up.

2) nmos depletion mode transistor pull-up

(a) Dissipation is high since rail to rail current flows when $V_{in} = \text{logical } 1$

(b) Switching of o/p from 1 to 0 begins when V_{in} exceeds V_t of P.d device

(c) when switching the o/p from 1 to 0, the PMOS device is non-saturated initially and this presents lower resistance through which to charge capacitive loads.

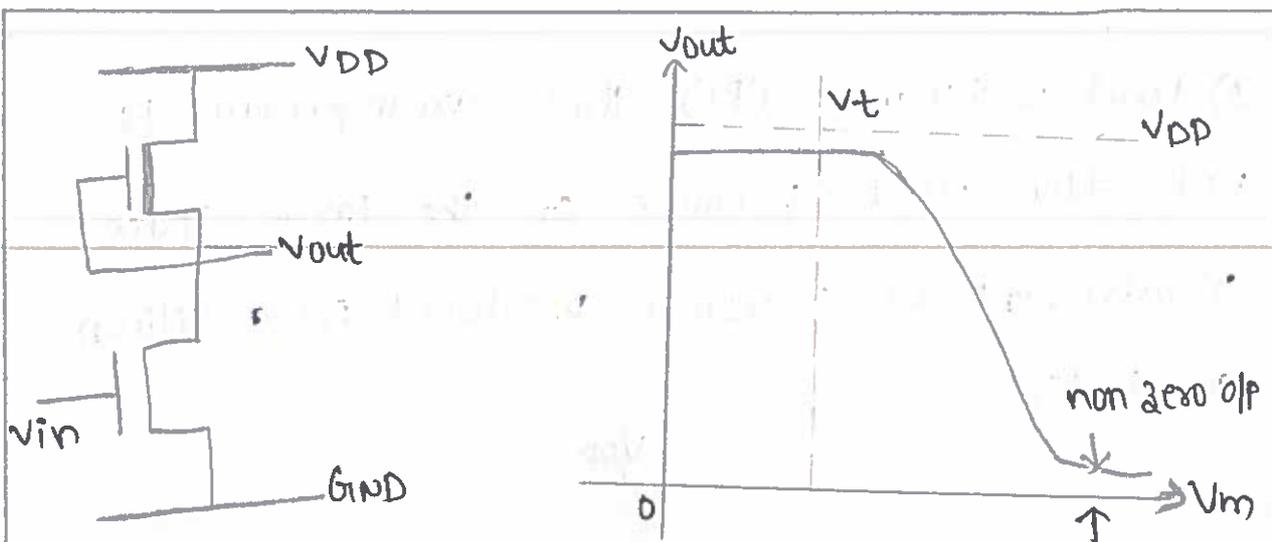


Fig. 2. nmos depletion mode transistor pull-up and transistor characteristics.

3) Nmos Enhancement mode pull-up

(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 $V_{GG} = V_{DD}$

(c) V_{GG} may be derived from a switching source, EX: One phase of 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)

(a) NO current flow either logic 0 or logic 1 I/B

(b) Full logic 1 and 0 levels are presented at the o/p

(c) For devices of similar dimensions the p-channel is slower than the n-channel device

* The CMOS inverter:-

The general arrangement and characteristics are illustrated below with current/voltage relationships for the MOS transistor may be written.

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

in the resistive region, or

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

in saturation.

In both cases the factor k is a technology-dependent parameter such that

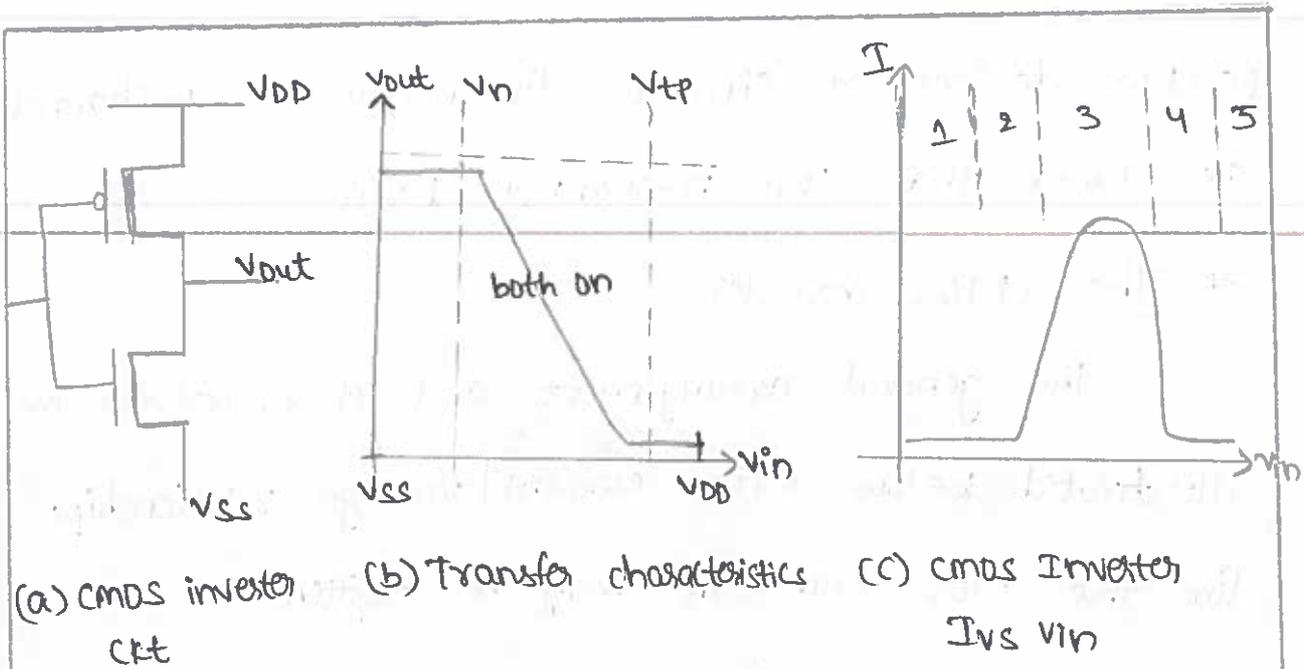
$$k = \frac{\epsilon_{ins} \epsilon_0 \mu}{D}$$

The factor W/L is of course, contributed by the geometry and it is common practice to write

$$\beta = k \cdot \frac{W}{L}$$

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

where $\beta_n = k \cdot \frac{W_n}{L_n}$, $\beta_p = k \cdot \frac{W_p}{L_p} \Rightarrow W_n \& L_n, W_p \& L_p$ are the n & p-transistors dimensions respectively.



Operation: CMOS inverters have five distinct regions of operation. Consider static condition 1st, it may be seen that in

Region 1: For which $V_{in} = \text{logic } 0$, so the p-transistor is fully turned on while the n-transistor is fully turned off. Thus no current flows through the inverter and the o/p is directly connected to V_{DD} through the P-transistor.

$\therefore V_{out} \Rightarrow$ Present good logic 1 o/p

In region 5:- when $V_{in} = \text{logic } 1$, the n transistor is fully on while the p-transistor is fully off. Again no current flows and a good logic 0 appears at the o/p.

In region 2: The i/p voltage (v_{in}) has increased to a level which just exceeds the threshold voltage of the n-transistor.

The n-transistor conducts and has a large voltage b/w source and drain. So it is in Saturation region.

The p-transistor is also conducting but with only a small voltage across it, it operates in unsaturated or resistive region.

A small current flows through the inverter from VDD to VSS. If we want to analyze the behaviour of this region, equating p-device resistive region I_{ds} = n-device saturation region (I_{ds}). Thus obtain v_{ds} & I_{ds} relationships.

Region 4: It is similar to region 2 but with the roles of the P & n transistors reversed (P → Saturation)
n → resistive)

current magnitudes are small

Region 3: in which the inverter exhibits gain and in which both transistors are in saturation

Therefore currents in each device must be same. Since the transistors are in series, so we may write

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

$$I_{dsp} = \frac{\beta_p}{2} (v_{in} - V_{DD} - V_{tp})^2 \quad (\because v_{gs} = v_{in} - V_{DD} \text{ for p-mos})$$

$$I_{dsn} = \frac{\beta_n}{2} (v_{in} - V_{th})^2 \quad (\because v_{gs} = v_{in} \text{ for nmos})$$

From which we can express v_{in} in terms of β ratio

$$v_{in} = \frac{V_{DD} + V_{tp} + V_{th} (\beta_n / \beta_p)^{1/2}}{1 + (\beta_n / \beta_p)^{1/2}}$$

→ so both are act as current source so that the equivalent ckt in this region is two current sources in series b/w V_{DD} & V_{SS} .

If $\beta_n = \beta_p$ & if $V_{th} = -V_{tp}$

$$v_{in} = \frac{V_{DD}}{2} = 0.5 V_{DD}$$

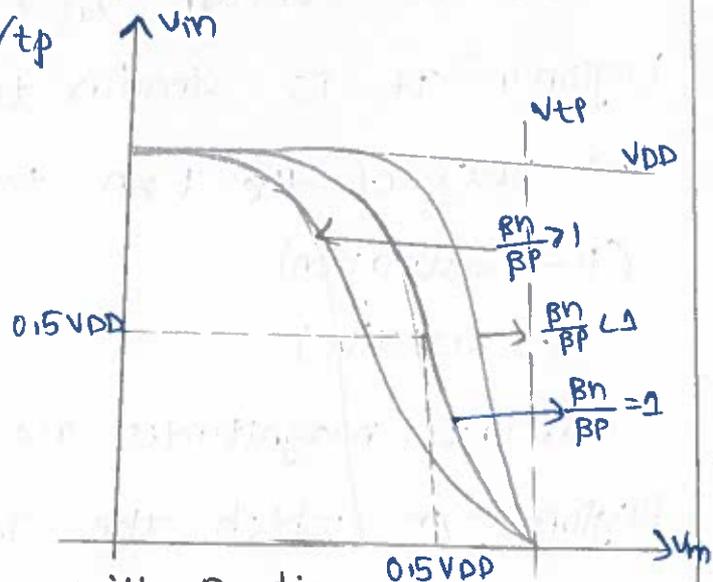


Fig: characteristics with β ratio

The transfer characteristics are observed at

$$V_{in} = 0.5 V_{DD}$$

$$V_{in} = V_{out} = V_{inv} = 0.5 V_{DD}$$

Here $\beta_n = \beta_p$ in terms of mobility

$$\mu_p \frac{W_p}{L_p} = \mu_n \frac{W_n}{L_n}$$

$$\frac{W_p}{L_p} = 2.5 \frac{W_n}{L_n}$$

so length & width ratio of p device is 2 (or) 3 times that of n-device

$$\mu = \mu_2 \left(1 - \frac{1}{\phi (V_{gs} - V_t)} \right) \quad \phi \rightarrow \text{constant} = 0.05$$

$\mu_2 \rightarrow$ zero transverse field.

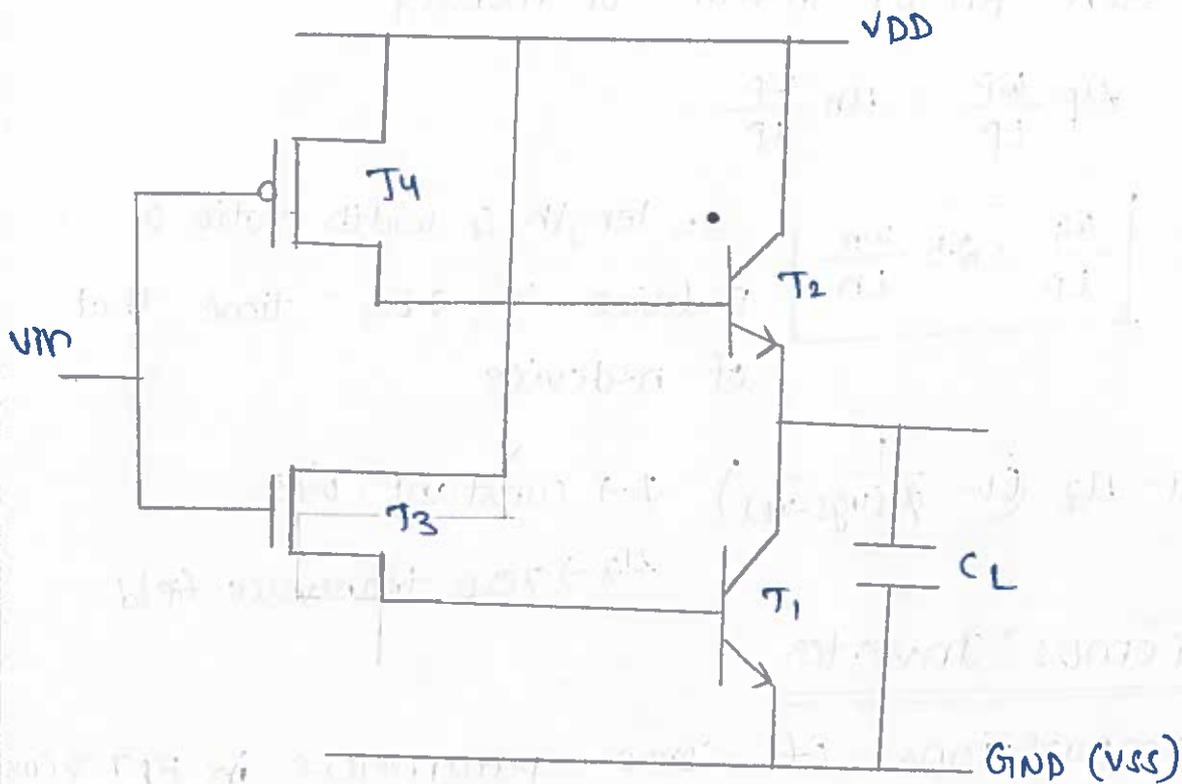
Bicmos Inverter

comparisons of some parameters in BJT & CMOS

CMOS	Bipolar
1) $I_{ds} = \frac{\mu C_0}{2} \frac{W}{L} (V_{gs} - V_t)^2$	1) $I_c = I_s e^{qV_{be}/kT}$
2) $g_m = (2\beta)^{1/2} (I_{ds})^{1/2}$	2) $g_m = \frac{I_c}{kT/2}$
3) $I_{ds}/A = \frac{\mu C_0}{2L^2} (V_{gs} - V_t)^2$	3) $I_c/A = \frac{1}{R_B \mu T_B}$
$\frac{I_{ds}}{A}, \frac{I_c}{A}$ current/Area and	$R_B \rightarrow$ base resistance $T_B \rightarrow$ base transit time (10-30ps)

~~Q.1~~ → we may see that I/A for bipolar is 5 times better than that of CMOS.

Bicmos Inverters:-



A simple Bicmos inverter.

As in nmos and cmos logic circuitry, the basic logic element is the inverter ckt.

When designing with Bicmos in mind, the logical approach is to use Mos switches to perform the logic function and bipolar transistors to drive the o/p loads.