



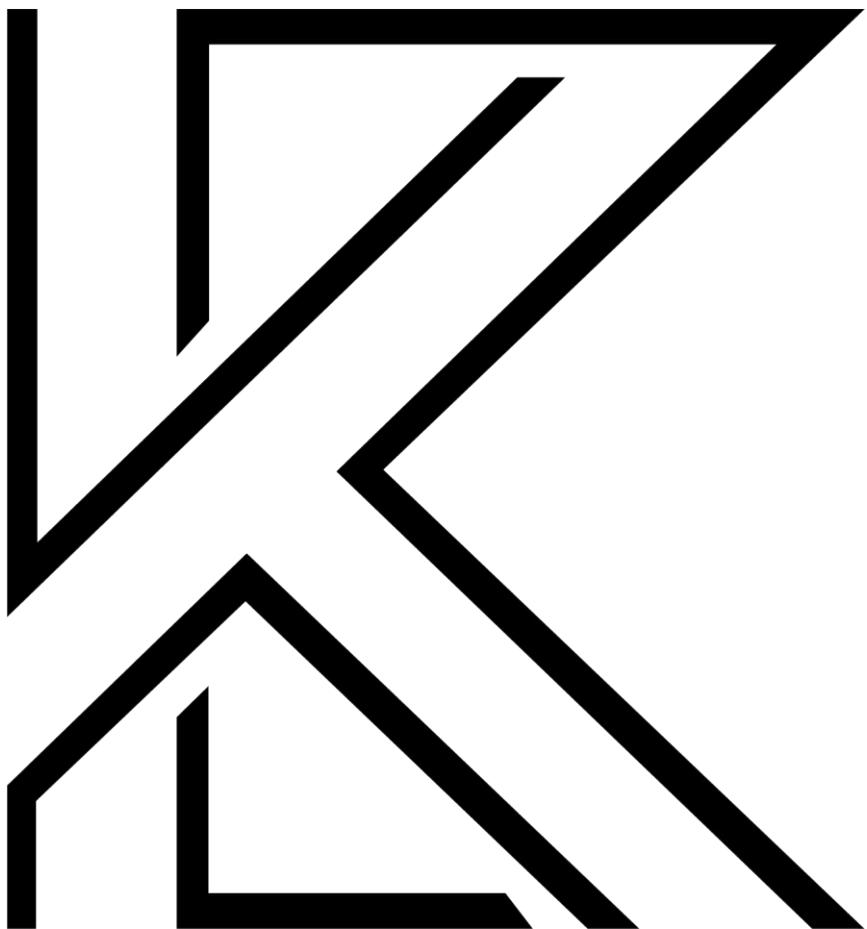
# VLSI Technology

وَمَا أَنْهَ فِي سُكُونٍ إِلَّا بِاللَّهِ<sup>ۖ</sup>

MY SUCCESS IS ONLY BY ALLAH

Quran 11:88

Course Teacher  
Arif Istiaque Rupom



Remember in Your Prayers

# INDEX

<b>Section A -----</b>	<b>5</b>
VLSI Introduction . . . . .	6
VLSI Design Approach . . . . .	9
MOS Technology . . . . .	10
NMOS Inverter . . . . .	15
Accumulation,Depletion and Inversion . . . . .	19
<b>CMOS Inverter . . . . .</b>	<b>20</b>
BiCMOS Inverters . . . . .	24
Stick Diagrams . . . . .	29
Array Subsystems . . . . .	35
6T SRAM Cell . . . . .	37
DRAM . . . . .	38
<b>Section B -----</b>	<b>39</b>
ROM . . . . .	40
FPGA . . . . .	40
HDL and VHDL . . . . .	41
<b>Fabrication . . . . .</b>	<b>47</b>
Wafer Processing . . . . .	48
Oxidation . . . . .	50
Lithography . . . . .	52
Epitaxial Growth . . . . .	54
Etching . . . . .	56
Wafer Cleaning . . . . .	61
Metalization & Passivation . . . . .	63
Diffusion & Ion Implantation . . . . .	63
<b>VLSI Physical Design Playlist . . . . .</b>	<b>65</b>
Y-Diagram . . . . .	67
VLSI Design Styles . . . . .	68
Floorplaning & Partitioning . . . . .	70
STA,Clock & Power Routing . . . . .	72
Placement & Routing . . . . .	73
The Clock Design in VLSI . . . . .	75
Pipelining . . . . .	77
Clock Skew and Jitter . . . . .	78
Designing Clock for VLSI . . . . .	81

# Section A

- VLSI Introduction
- MOS
- CMOS
- BiCMOS
- Array subsystems

## Basic knowledge:

Electrical: Deals with current through conductors like copper wires using AC/DC power for power transmission.

Electronics: Deals with current through semiconductors like diodes, transistors, for signal processing and control.

## History:

The history of VLSI began with diodes and vacuum tubes used in early electronics. In 1947, the invention of the transistor revolutionized electronics by replacing bulky tubes. In the 1960s, MOSFETs enabled miniaturization and low-power design. This led to the development of digital logic design DLD and control systems, which enabled automation and logic-based circuits. The 1970s introduced microcontrollers, integrating processor, memory and I/O on a single chip. These innovations culminated in VLSI technology, allowing millions of transistors on a chip. VLSI enabled the creation of powerful microprocessors and memory units, forming the backbone of modern computers and embedded systems.

## VLSI

Very Large Scale Integration is the process of integrating thousands to millions of transistors on a single microchip to create complex electronic circuits. It enables the design of compact, high-performance, and low-powered devices like microprocessors and memory chips. VLSI is the foundation of modern electronics such as smartphones, laptops and embedded systems.

The Intel Core i7 processor is a VLSI chip containing over a billion transistors integrated onto a single silicon die for performing advanced computing tasks.

## Advantages and goals of VLSI:

- High integration: Packs millions of transistors into a small chip.
- Reduced size: Minimizes the size of electronic devices.
- Low power consumption: Optimizes energy efficiency.
- High speed: Enhances processing speed and performance.
- Cost-effective: Reduces overall system cost with mass production.
- Improved reliability: Fewer external connections mean fewer failures.
- Goal: To design compact, efficient and powerful electronic system.

■ What is photo-lithography? Explain the fabrication process of VLSI.

■ Photolithography is the process used in VLSI manufacturing to transfer circuit patterns onto a silicon wafer using light-sensitive materials.

VLSI Fabrication Process: It begins with purifying silicon sand to create silicon ingots, sliced into wafers. The wafer is cleaned and oxidized, then coated with photoresist. A mask with the circuit pattern is placed over it, and UV light is shined, changing the exposed photo resist. The wafer is developed, etching away selected areas. Layers of metals, insulators and semiconductors are using repeated photolithography steps to build the circuit. After doping, layering and metalization, the wafer is tested, cut and packaged into micro-chips - ready for use in electronic devices.

■ What is integrated circuit? What benefits do they have over traditional circuits?

■ An Integrated Circuit (IC) is a complete electronic circuit fabricated onto a single piece of semiconductor material, usually Si. IC contains components like transistors, registers and capacitors all embedded on a single chip to perform logic or processing function.

## Benefits over traditional circuits:

- Compact size: Reduces space compared to bulky discrete components.
- High Reliability: Fewer connections means lower chances of failure.
- Lower cost: Mass production reduces manufacturing costs.
- High Speed: Shorter paths enable faster signal processing.
- Low power consumption: Smaller components use less power.
- Scalability: Enables the integration of millions of transistors for complex systems (VLSI)

■ Discuss about the choice of technology (BJT vs MOS) in VLSI?

■ MOS is preferred over BJT, because-

- Smaller size: MOS transistors occupy less silicon area than BJT.
- Lower Power: MOS circuits consume much less power than BJT.
- Higher Yield: Simpler structure results in fewer fabrication defects.
- Dynamic logic: only MOS supports dynamic logic for further area/power saving.
- Better integration: only MOS can realize true VLSI integration.
- Manufacturing simplicity: Easier and cheaper to fabricate than BJT.
- Improving speed: MOS speed increases with shrinking feature size; BJT speed is near its physical limit.
- CMOS Advantage: offers both low power and high speed ideal for VLSI.

## ■ Design approaches in VLSI

1. Off-the-shelf design: Uses pre-manufactured ICs (SSI, MSI, LSI). Limited to available chips; may need multiple ICs to achieve the function.  
Example: Using 7400-series TTL NAND gates to build an adder.
2. Full custom design: Every transistor, wire and layer is designed by the engineer for maximum efficiency in speed, power and area. Need High cost and time.  
Example: Designing a custom ARM Processor Core from transistor level.
3. Semi-custom design: (VLA) ASIC Pre-fabricated cells (logic gates, flip-flops) are customized by interconnection. faster and cheaper than full-custom but less flexible.  
Example: FPGA or Gate Array-based design for a single processor.

## ■ Hierarchical Design in VLSI - Top-Down Approach

VLSI design follows a top-down hierarchical process, where each level refines the design until fabrication. The key steps are:

System Specification: Defines system requirements and functionalities.

System Specification

System Design: Translates specifications into an architectural block diagram (registers, ALU, memory, data paths). Test strategy planned here.

System Design

Logic Design: Extends architecture into logic gates, control and timing. Verified by logic simulation.

Logic Design

Circuit Design: Converts logic into transistor-level circuits (stick diagrams), estimates size, speed and power.

Circuit Design

Geometric Layout: Creates mask-level layout of transistors and interconnections, ensuring design rule compliance.

Geometric Layout

Fabrication: Layout data is used to create masks, followed by manufacturing to produce the chip.

Fabrication

figure: Top-down design hierarchy

MOS conduction:  
A positive voltage is applied to the gate, with source and substrate at 0V. The oxide layer acts as an insulator between gate and (metal) substrate (semiconductor). forming a capacitor like structure.

Positive gate bias induces negative charge electrons in the p-type substrate under the oxide.

If the charge is large enough, it creates an n-type inversion layer, forming a conducting channel.

Current flows from drain to source when a drain bias is applied. The channel is isolated from the gate by oxide and from substrate by a depletion region.

### MOS Threshold voltage:

The minimum gate to source voltage  $V_{gs}$  needed to form an inversion layer and turn on the NMOS transistor, is known as the threshold voltage  $V_t$ . Below  $V_t$ , the transistor is OFF, no conduction occurs.

Above  $V_t$  an n-type channel forms, allowing current  $I_d$  to flow from drain to source.

At inversion charge balance gives,

$$Q_{ss} + Q_g = Q_d + Q_c \quad \text{where, } Q_{ss} = \text{Surface state charge.}$$

At threshold  $Q_c = 0$  and,

$$V_t = \frac{Q_g}{C_g} = \frac{Q_d - Q_{ss}}{C_g} + V_n + V_{ds}$$

$V_t$  can be adjusted by doping impurity levels.

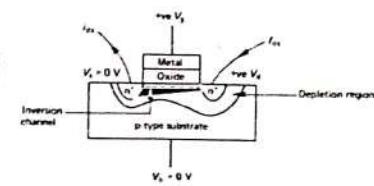


Figure 2.3 Conduction in an NMOS transistor

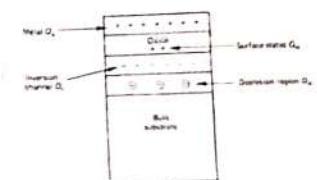


Figure 2.4 Extra charge within conducting NMOS transistor

$Q_g$  = Gate charge.

$Q_c$  = Induced charge in the inversion channel.

$Q_d$  = Impurity atom charge in the depletion

$C_g$  = Capacitor across the insulator.

$V_n$  = Voltage across depletion region just at inversion.

$V_{diff}$  = Voltage due to differences in between semiconductor and gate metal.

Body effect is the phenomenon where the threshold voltage  $V_t$  of a MOS transistor increases when the substrate (body) voltage differs from the source voltage.

When  $V_{SB}$  source to body voltage  $\neq 0$ , the depletion region widens increasing the charged  $Q_d$ , it rises  $V_t$  making the transistor harder to turn on. Common in NMOS where body is connected to ground and source is at higher potential.

Then modified threshold voltage is,

$$V_t = V_{to} + \gamma (V_{SB})^{1/2}$$

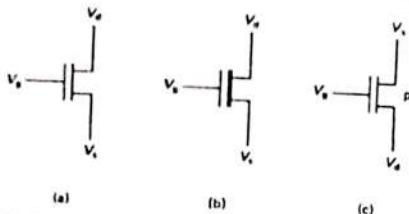


Figure 2.5 Symbols for MOS transistors: (a) enhancement mode NMOS, (b) depletion mode NMOS, (c) enhancement mode PMOS

where,

$V_t$  = threshold voltage modified

$V_{to}$  =  $V_t$  when  $V_{SB}=0$ .

$V_{SB}$  = source to body voltage

$\gamma$  = constant usually 0.5, depends on transistors parameters.

$I_{ds}$  vs  $V_{ds}$  characteristics of NMOS Devices:

when  $V_{gs} < V_t$  the transistor is OFF, regardless of the drain voltage. The device doesn't conduct and no current flows. The device conduct when  $V_{gs} > V_t$  and if a constant  $V_{gs}$  is applied then the resulting  $I_{ds}$  vs  $V_{ds}$  curve can be split into two regions.

(a) Resistive Region.  $V_{ds} < V_{gs} - V_t$

The device structure resembles an infinite number of capacitances between the drain and source, each one having a different voltage across it and therefore the different charge from its neighbours. The total charge induced in the channel is the sum of the charge included in each of these capacitances.

consider one of these capacitance of length  $dx$  is situated at a distance  $x$  meters from the drain (fig 2.6). The channel width and length are  $W$  &  $L$  respectively. Thus the capacitance,

$$C = \frac{AE}{D} = \frac{WE dx}{D} \quad [W dx = A] \quad \text{where,}$$

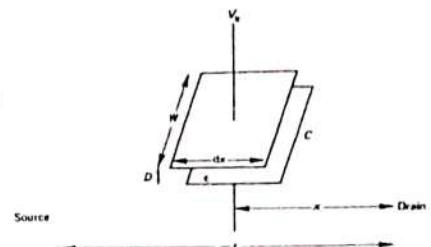


Figure 2.6 Elemental capacitor between source and drain

$\epsilon$  is the permittivity of insulator in F/m. and  $D$  is the thickness of the oxide.

The voltage  $V$  in excess of  $V_t$  across this capacitor is,

$$V = \left( V_{gd} + \frac{x}{L} V_{ds} - V_t \right) = \left( V_{gs} - V_{ds} + \frac{x}{L} V_{ds} - V_t \right)$$

Thus the charge induced on this capacitor is,

$$q = VC = \left( V_{gs} - V_{ds} + \frac{x}{L} V_{ds} - V_t \right) \cdot \frac{WE dx}{D}$$

The total charge  $Q$  induced in the channel is

$$Q = \int \frac{WE}{D} \left( V_{gs} - V_{ds} + \frac{x}{L} V_{ds} - V_t \right) dx = \frac{WEL}{D} \left[ (V_{gs} - V_t) - \frac{V_{ds}}{2} \right]$$

$$\text{Now, } I_{ds} = \frac{Q}{t} = \frac{WEL}{D} \left[ (V_{gs} - V_t) - \frac{V_{ds}}{2} \right] / \frac{L}{\mu_n V_{ds}} = \frac{4WEL\mu_n}{LD} \left[ (V_{gs} - V_t) - \frac{V_{ds}}{2} \right]$$

$$\therefore I_{ds} = \frac{WEL\mu_n}{LD} \left[ (V_{gs} - V_t) V_{ds} - \frac{V_{ds}^2}{2} \right]$$

$$I_{ds} \approx \frac{E W \mu_n}{L D} (V_{gs} - V_t) V_{ds}$$

giving a linear relationship between  $I_{ds}$  and  $V_{ds}$  for a constant  $V_{gs}$ .

$$(b) \text{ Saturation: } V_{ds} \geq V_{gs} - V_t$$

As the drain voltage rises, the voltage across the insulation at the drain drops, and at  $V_{ds} = V_{gs} - V_t$  it is  $V_t$ . This is the voltage necessary to just support inversion, and this point on the  $I_{ds}$  vs  $V_{ds}$  characteristic is called "pinch-off". At this point the inversion channel ends just at the drain. As  $V_{ds}$  increases beyond pinch-off, the point at which the inversion ceases moves away from the drain fig 2.8. The voltage difference along the inversion channel from the source to where it ceases is  $V_{gs} - V_t$  and the excess potential  $V_{ds} - V_{gs} + V_t$  is dropped between the end of the inversion channel and the drain. This creates a high electric field across this very short distance and the electrons from the inversion channel are quickly swept across this area to the drain.

At the above pinch off voltage between the source and the end of the inversion channel is constant at  $V_{gs} - V_t$ . Thus,

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

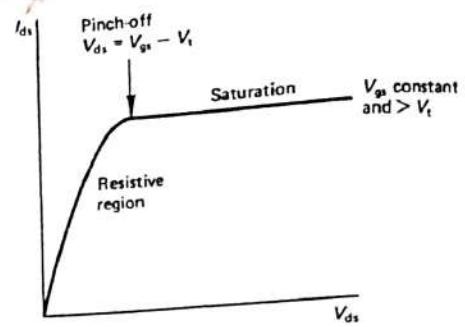


Figure 2.7  $I_{ds}$  versus  $V_{ds}$  for an NMOS transistor

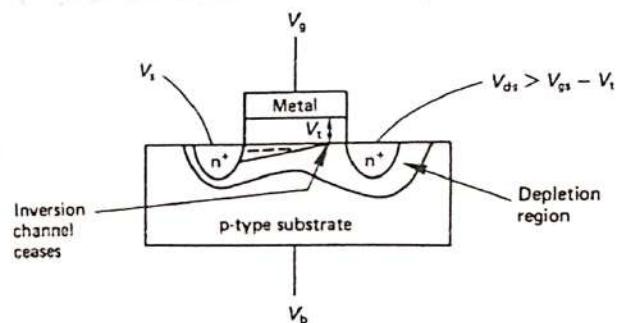


Figure 2.8 Inversion channel beyond pinch-off

### NMOS

Resistive current equation,

$$I_{ds} = \frac{\epsilon W M_n}{L D} (V_{gs} - V_t) V_{ds}$$

Saturation current equation,

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

### PMOS

Resistive current equation,

$$I_{sd} = \frac{\epsilon W M_p}{D L} (V_{sg} - V_t) V_{sd}$$

Saturation current equation,

$$I_{sd} = \frac{\epsilon W M_p}{2 D L} (V_{sg} - V_t)^2$$

Principles of inverters:

Inverters are the circuit that gives output high for low input and vice-versa.

Referring to the fig 1:

- For  $V_{in}$  high, switch S closes, output pulled down to 0V.
- For  $V_{in}$  low, switch S open, output pulled high  $V_p$ .
- NMOS transistor used as switch:  $V_{gs} < V_t \rightarrow \text{off}$ ,  $V_{gs} > V_t \rightarrow \text{on}$ .
- Basis of all mos logic circuits.
- For resistive load R chosen so  $V_{out}$  sufficiently low for noise margin.

Resistive load is typically impractical since it uses large area (300X).

Using the equation,

$$\begin{aligned} I_{ds} &= \frac{\epsilon W M_n}{L D} \left[ (V_{gs} - V_t) V_{ds} - \frac{V_{ds}^2}{2} \right] \\ &= 30 \left( \frac{1}{1} \right) \left[ (5 - 1) \cdot (0.3) - \frac{(0.3)^2}{2} \right] \\ &= 34.7 \mu A. \end{aligned}$$

$$\text{Hence, } R = \frac{V_p - V_{out}}{I_{ds}} = \frac{5 - 0.3}{34.7 \mu A} = 135.4 \text{ k}\Omega$$

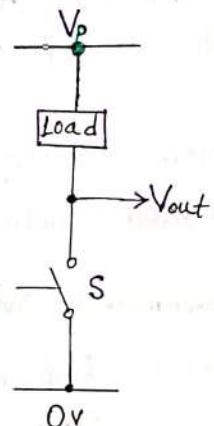


Fig 1. Inverter principle

Where,

$$\frac{\epsilon W M_n}{D} = 30 \mu A/V^2$$

$$\text{Aspect ratio } \frac{W}{L} = \frac{1}{1}$$

$$V_{gs} = 5V; V_t = 1V$$

$$V_{ds} = 0.3V$$

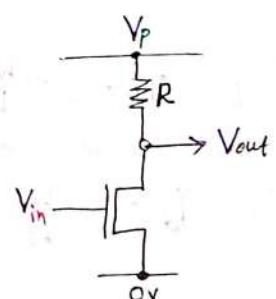


Fig 2: NMOS inverter with R load.

The silicon area required to implement this resistor is far larger than that for the transistor (300X). Thus it is not practical to use a resistor as a load and so an MOS device is used instead.

NMOS inverter with an NMOS Enhancement type-transistor Load.

Uses two NMOS enhancement type transistors where  $T_1$  driven (below),  $T_2$  load (top).  $T_2$  node is tied to  $V_{DD}$ - $V_p$ , always biased for maximum  $V_{GS}$ .

$T_1$  driver (below),  $T_2$  load (top).  $T_2$  node is tied to  $V_{DD}$ - $V_p$ , always biased for maximum  $V_{GS}$ .

- For  $V_{in}$  low ( $< V_{th}$ ):  $T_1$  off, small leakage through  $T_1$ ,  $V_{out} = V_{DD} - V_{th} \approx 3.12V$ , with body effect.

- For  $V_{in}$  high ( $> V_{th}$ ):  $T_1$  on, conducts current  $V_{out}$  pulled low ( $\sim 0.3 V_{th}$ )

- Proper sizing  $\frac{W}{L}$  balances current of  $T_1$  and  $T_2$ , inverter ratio  $K \approx 12$ .

Improvement: over resistive load, smaller area, better integration.

Limitations: Reduces high output voltage  $V_{OH}$  due to body effect, slow rising edge because  $T_2$  turns off as  $V_{out}$  closes  $V_{OH}$ .

Using the equation,

$$I_{ds} = \frac{EW_{eff}}{L_D} \left[ (V_{GS} - V_t) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

for

$$T_1: I_{ds} = 30 \left( \frac{W_1}{L_1} \right) (3.12 - 1) 0.3 - \frac{0.3^2}{2} = 17.73 \frac{W_1}{L_1} = 17.73 \left( \frac{W_1}{L_1} \right).$$

and using the equation,  $I_{ds} = \frac{EW_{eff}}{L_D} (V_{GS} - V_t)^2$

$$\text{for } T_2: I_{ds} = 30 \left( \frac{W_2}{L_2} \right) (4.7 - 1)^2 = 205.4 \left( \frac{W_2}{L_2} \right)$$

$$\text{inverter ratio, } K = \frac{17.73 W_1 / L_1}{205.4 W_2 / L_2} = \frac{W_1 / L_1}{W_2 / L_2} = \frac{205.4}{17.73} = 11.6 \approx 12$$

Aspect ratio of  $T_1$

Aspect ratio of  $T_2$

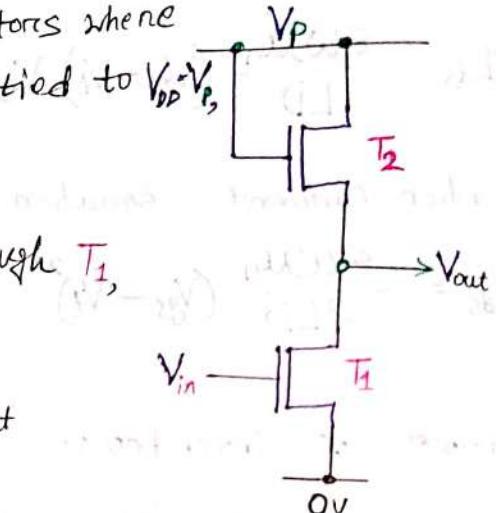


Fig 3: NMOS inverter with NMOS enhancement load

✓ NMOS inverter with an NMOS Depletion Transistor load:

Load device  $T_2$  is depletion type NMOS, negative  $V_{th} = -4V$  always on.

- For  $V_{in}$  low ( $< V_{th}$ ) of  $T_1$ :  $T_1$  off,  $T_2$  conducts leakage,  $V_{out} \approx V_{DD}$  (no  $V_{OH}$  loss).

- For  $V_{in}$  high ( $> V_{th}$ ): both  $T_1$  and  $T_2$ ,  $T_1$  is in resistive mode,  $V_{out}$  pulled to  $\sim 0.3 V_{th}$ .

Depletion load requires smaller inverter ratio ( $K \approx 6$ ) saving silicon area.

Improvement: over enhancement load  $V_{OH} \approx V_{DD}$ , faster rising edge  $T_2$  never cuts off.

Limitations: needs special doping to fabricate depletion NMOS.

Overall better speed, higher noise margin and reduced area than enhancement type load inverters.

Using the equation,

$$I_{ds} = \frac{E W M_n}{L D} \left[ (V_{ds} - V_t) V_{ds} - \frac{V_{ds}^2}{2} \right]$$

For,  $T_1$ :  $I_{ds} = 30 \left( \frac{W_1}{L_1} \right) (5-1) 0.3 - \frac{0.3^2}{2} = 34.65 \left( \frac{W_1}{L_1} \right) \mu A$

For depletion type transistor

$$I_{ds} = \frac{E W M_n}{2 L D} (V_{ds} - V_t)^2 \quad \left| \frac{E M_n}{D} = 25 \right.$$

For  $T_2$ :  $I_{ds} = \frac{25}{2} \left( \frac{W_2}{L_2} \right) [0 - (-4)]^2 = 200 \left( \frac{W_2}{L_2} \right) \mu A$

inverter ratio,  $K = \frac{W_1 L_1}{W_2 L_2} = 5.8 \approx 6$

For convenient - thus inverter ratio would be taken as 6 and split as

$$\frac{W_1}{L_1} = \frac{3}{1} \text{ and } \frac{W_2}{L_2} = \frac{1}{2}$$

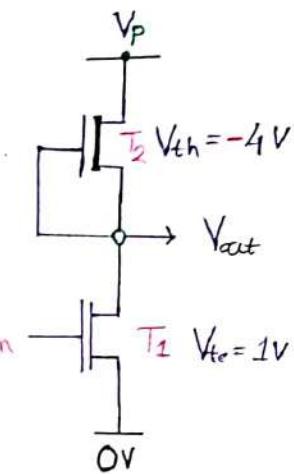


Fig 4: NMOS inverter with NMOS depletion type load.

⑤ Edge time for NMOS inverter with depletion load

Figure 2.13 represents a practical inverter for NMOS technology. The edge time response of this inverter is worth examining as this will indicate the maximum speed of operation.

The maximum current for  $T_2$  is already found to be  $200 \left( \frac{W_2}{L_2} \right) \mu A$ , the maximum current through  $T_1$  occurs when the gate source voltage is maximum 5V and the device is saturated.

maximum  $I_{ds}$  of  $T_1$ :

$$I_{ds} = \frac{EWUn}{PL} (V_{gs} - V_t)^2 = \frac{30}{2} \left( \frac{W_1}{L_1} \right) (5-1)^2 = 240 \left( \frac{W_1}{L_1} \right) \mu A.$$

Thus,

$$\frac{\text{Max } I_{ds} \text{ of } T_1}{\text{Max } I_{ds} \text{ of } T_2} = \frac{6 \left( \frac{W_1}{L_1} \right)}{5 \left( \frac{W_2}{L_2} \right)} = \frac{6}{5} K$$

Load capacitance  $C_{out}$  comes from gate, source and drain parasitics typically  $\sim 0.1 \mu F$ . (Fig 2.14)

Rise time determined by current from  $T_2$  charging  $C_{out}$ .  $T_2$  saturates briefly ( $0.8 \sim 1$ ) V, then enters resistive region, rise time is dominated by  $R_{C_{out}}$  charging. Approximately.

Now calculating  $R$  for figure 2.16(b).

$$R = \frac{\text{Pinch-off voltage}}{\text{pinch-off current}}$$

$$= \frac{(V_{gs} - V_t)}{\frac{EWUn}{2D} \left( \frac{W}{L} \right) (V_{gs} - V_t)^2}$$

$$= \frac{1}{\frac{EWUn}{2D} \left( \frac{W}{L} \right) (V_{gs} - V_t)} = \frac{1}{\frac{25}{2} \left( \frac{W_2}{L_2} \right) (0-(-4))} M\Omega = \frac{20}{\left( \frac{W_2}{L_2} \right)} k\Omega$$

$$\therefore R = \frac{20}{\left( \frac{W_2}{L_2} \right)} k\Omega$$

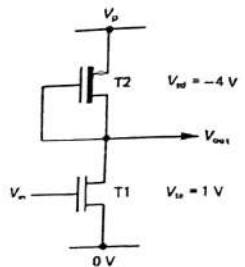


Figure 2.13 NMOS inverter with NMOS depletion load

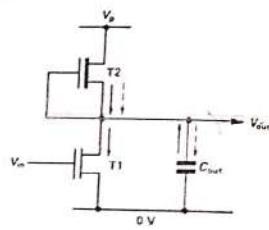


Figure 2.14 Current flow during edge times: + rising edge current flow, ↓ current flow during falling edge

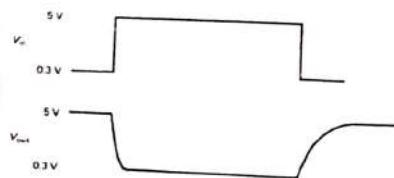


Figure 2.15 Output response to an input step

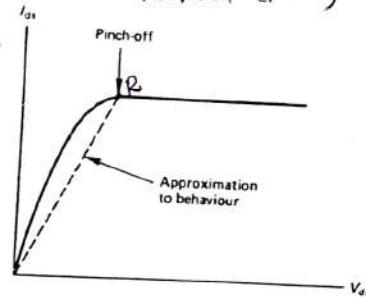


Figure 2.17 Characteristic approximation

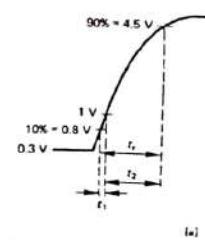
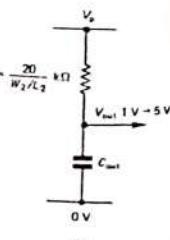


Figure 2.16 Rise time behaviour: (a) rising edge, (b) circuit approximation



The output voltage at fig 2.16(B) is.

$$V_{out} = V_p - (V_p - V_i) \exp\left[-\frac{t}{RC_{out}}\right]$$

where,

$$V_p = 5V$$

$$V_i = 1V$$

$$V_{out} = 4.5V$$

The time for the output to reach 4.5V (50%) with initial voltage of 1V

$$4.5 = 5 - (5-1) \exp\left[-\frac{t_2}{RC_{out}}\right]$$

$$\therefore t_2 = 9.08 RC_{out}$$

If  $C_{out}$  is in pF and  $R$  is in kilo-ohm, then  $t_2$  is given in nano-seconds. Substituting  $T_2$ 's value of  $R$ , the rise time is

$$t_n = t_2 = \frac{42 C_{out}}{W_2/L_2} \text{ ns}$$

the fall time,

$$t_f = \frac{5 t_n}{6K} = \frac{5}{6} 42 \frac{C_{out}}{W_2/L_2} = \frac{35 C_{out}}{W_2/L_2} \text{ ns}$$

The expression for the fall time agrees with results obtained by simulating the inverter circuit. However due to the 'body effect' the expression for the rise time is not so accurate. As the output rises, the threshold of  $T_2$  rises, reducing the current in  $T_2$ . This increases the time necessary to charge the output capacitance. Simulation results of the inverter shows that,

$$t_n = \frac{60 C_{out}}{W_2/L_2} \text{ ns.}$$

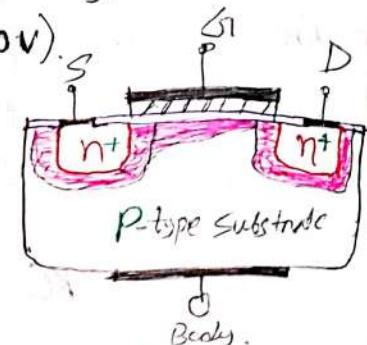
As the rise time scales with  $C_{out} \approx 0.15$ ,  $t_n = 12 \text{ ns} \times 1.5 = 18 \text{ ns}$  and  $t_f = 1.2 \text{ ns}$

$$f_{max} = \frac{1}{t_n + t_f} = \frac{1}{(18+1.2) \text{ ns}} = \frac{1}{19.2 \text{ ns}} = \frac{1}{19.2} \times 10^9 \approx 50 \text{ MHz.}$$

a lots of approximation!!!

✓ Why NMOS substrate is connected to ground, explain with example:

□ NMOS transistor is built on P-type substrate. To keep PN junction reverse biased (source-substrate and drain substrate), the substrate must be tied to the lowest potential (ground 0V). If not grounded, the substrate could float, which forward biased the junctions, hence leakage current flows or latching-up occurs.



Example: In NMOS with  $V_{DD} = 5V$  source 0V, drain = up to 5V substrate = 0V. This ensures the drain-substrate on source-substrate junctions remain reverse biased, preventing unwanted current.

✓ Explain Accumulation mode, Depletion mode and inversion mode of MOS

□ Accumulation mode: Fig 1

- Occurs when  $V_{GS} < 0$ , for NMOS on P-type substrate.

- Negative gate voltage attracts holes (majority carriers) to the oxide-semiconductor interface.

- Increases majority carrier density 'Accumulation'

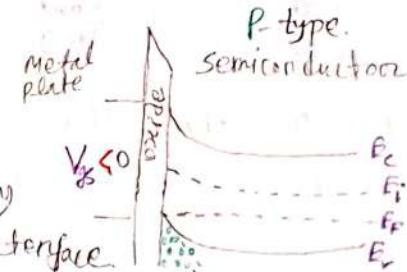
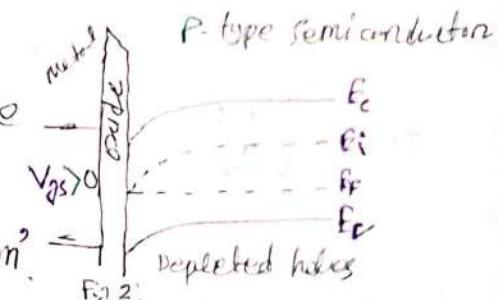


Fig 1: Accumulation of holes

Depletion Mode: Fig 2

- Occurs when  $0 < V_{GS} < V_{th}$ , positive gate voltage repels holes from the surface, leaving behind fixed negative ions, depletion region.

- Surface carrier concentration decreases.



Inversion Mode: Fig 3

- Occurs when  $V_{GS} \geq V_{th}$ .

- Strong positive gate voltage attracts electrons (minority carriers) to the surface, forming a conducting N-channel.

- MOS acts as transisistor (switch on state for NMOS)

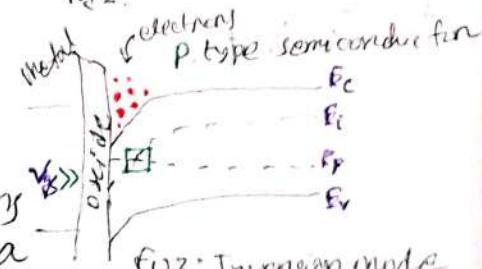


Fig 3: Inversion mode

✓ What is contention current? how is it generated and how it be mitigated?

Contention current is the short circuit current that flows when both NMOS and PMOS are on simultaneously in CMOS circuits.

Generation: Occurs during input transitions (e.g. inverter switching)  $V_{in}$  at intermediate voltage makes NMOS & PMOS conduct together, current flows directly from  $V_{DD}$  to GND.

In CMOS inverters, when  $V_{in} = V_{DD}/2$ , both transistors partially on, contention current occurs.

Mitigation:

- Used proper transistor sizing (ratiobalanced design).
- Reduce input transition time (fast switching).
- Apply dynamic logic or design techniques to avoid overlap.

### Ratiobalanced Vs Ratioballess Design:

Ratiobalanced: Output levels depend on the W/L aspect ratio of load driver transistors (e.g. NMOS with depletion/enhancement load). Proper ratio sizing is required for correct logic levels.

Ratioballess Design: Uses complementary switches (like CMOS) where only one transistor conducts at a time, ensuring full logic levels independent of transistor sizing.

- Widths (W/L) can still be varied for speed or power, but logic correctness does not rely on ratio. Hence called 'ratioballess' because functionality, unlike ratiobalanced logic, is not ratio-dependent.

### CMOS Inverter:

A CMOS inverter uses an NMOS transistor  $T_1$  (driver) and a PMOS transistor (load) in a complementary arrangement.

When  $V_{in} = 0V$ , NMOS  $T_1$  is 'on', PMOS  $T_2$  is 'on'  $V_{out} = V_{DD}$  (logic high).

When  $V_{in} = V_{DD}$ , NMOS  $T_1$  is 'off', PMOS  $T_2$  is 'off'  $V_{out} = 0V$  (logic low).

Only leakage current flows in steady states, almost zero static power dissipation.

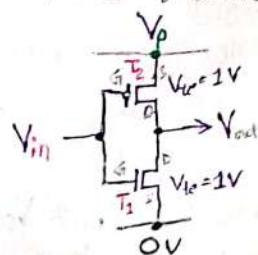


fig: CMOS inverter

Rise time: ( $0 \sim 5V$ ): Load capacitance charges via PMOS

Fall time: ( $5V \sim 0V$ ): Load capacitance discharges via NMOS.

Balanced edge time requires  $W_p = 2W_n$  (PMOS wider, we to lower hole mobility)

$$t_r = t_f = \frac{35 C_{out}}{W_1/L_1}$$

Draw and explain the I-V curves for CMOS inverters different values of  $V_{in}$ ?

Refering to the figure of emos inverter on the previous page, and figure 2.20

- For low  $V_{in}$  NMOS  $T_1$  is 'OFF', PMOS  $T_2$  is 'on',  $V_{out} = V_{DD}$ , current  $\approx 0$ . As  $V_{in}$  increases  $T_1$  turns on and  $T_2$  weakens  $V_{out}$  starts dropping.
- Around  $V_{in} = \frac{V_{DD}}{2}$ , both  $T_1$  and  $T_2$  conducts (Saturation region), sharp transition, maximum current
- For high  $V_{in}$ ;  $T_1$  is 'on',  $T_2$  is 'off',  $V_{out} \approx 0$ , current  $\approx 0$
- Current peaks only during switching explains low static power, dynamic power dissipation.

Explain CMOS inverter DC transfer curve with different region.

The CMOS inverter DC transfer curve shows how output voltage  $V_{out}$  changes with input voltage  $V_{in}$ .

- Region 1:  $V_{in} \approx 0$  to low PMOS  $T_2$  is on, NMOS  $T_1$  off,  $V_{out} = V_{DD}$  (logic high).
- Region 2: transition at  $V_{in} = \frac{V_{DD}}{2}$ : Both NMOS  $T_1$  and PMOS  $T_2$  conduct, causing a sharp fall in  $V_{out}$ , current peaks here, defining the switching region.
- Region 3:  $V_{in} \approx V_{DD}$  to high: NMOS  $T_1$  is 'on' PMOS  $T_2$  is 'off',  $V_{out} \approx 0$  (logic low).

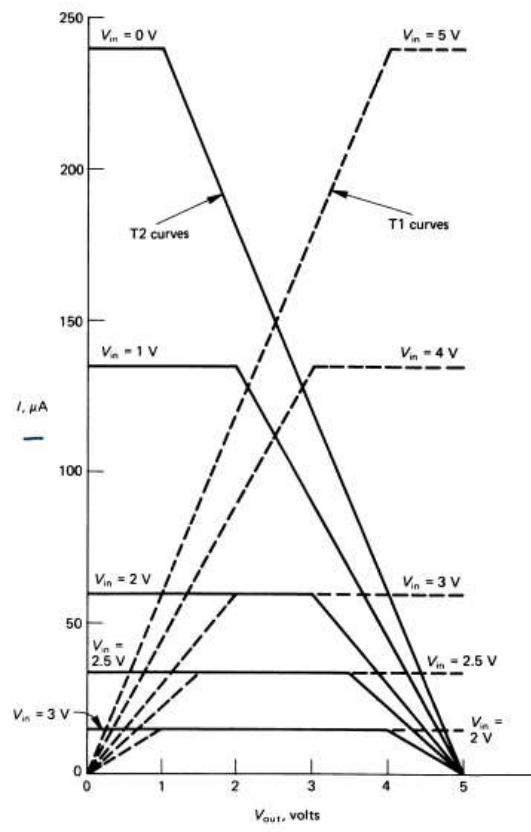


Figure 2.20  $I$  versus  $V_{out}$  curves for  $T_1$  and  $T_2$

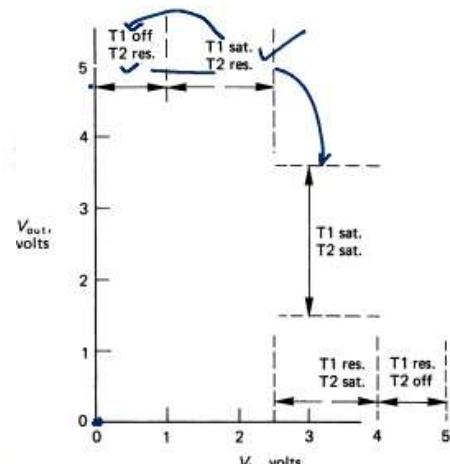


Figure 2.21 Characteristic curves for the CMOS inverter: (a) transfer characteristic, (b) current versus  $V_{in}$

Explain CMOS power dissipation curve  
 The CMOS inverter power curve shows power dissipation versus input voltage  $V_{in}$ .

Region 1: ( $V_{in}$  low,  $V_{out}$  high): one transistor 'OFF', the other 'ON', negligible static power.

Region 2: Transition region around  $V_{in} = V_{DD}/2$ : Both NMOS and PMOS conduct, maximum short circuit current, peak power dissipation

Region 3: ( $V_{in}$  high,  $V_{out}$  low): Again one transistor 'off', the other 'on', negligible static power.

Thus, power mainly dissipates during switching and total power depends on supply voltage  $V_p$  and switching frequency.

NMOS Pass transistors:

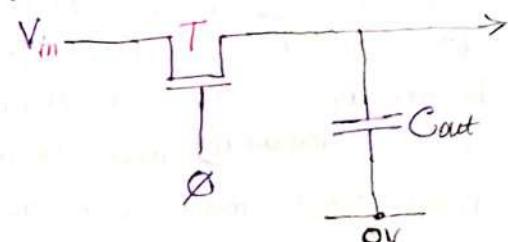
An NMOS pass transistor works as a voltage controlled switch.

When control  $\phi = 0$ , transistor off, output holds charge on  $C_{out}$ .

When  $\phi = V_p$  transistor 'ON', input is passed to output.

Limitation: Passes strong 0 (v) but weak 1 ( $V_p - V_{th}$ ).

Use-case: compact implementation of AND-like control, saves area, 18x than conventional NMOS logic. Common in multiplexers, dynamic logic and transmission gates.



CMOS Pass gate:

Uses parallel NMOS and PMOS transistors controlled by complementary signals ( $\phi$  &  $\bar{\phi}$ ).  $\phi = 0 \rightarrow$  both OFF, output holds previous value.  $\phi = V_p$  both ON passes input to output.

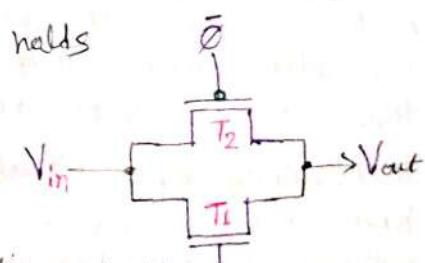
cases:

1.  $V_{in} = V_{out} = 0$ , No current,  $V_{out}$  stays 0

2.  $V_{in} = V_{out} = V_p$ , No current,  $V_{out}$  stays  $V_p$

3.  $V_{in} = V_p$ ,  $V_{out} = 0$ ,  $C_{out}$  charges NMOS turns ON,  $V_{out} = V_p$

4.  $V_{in} = 0$ ,  $V_{out} = V_p$ ,  $C_{out}$  discharges, PMOS turns ON,  $V_{out} = 0$ . Figure: CMOS pass gate.



## ■ NMOS Ratioless Inverter:

The ratioless NMOS inverter uses pass transistors and phased clocks ( $\phi_1, \phi_2$ ) to operate with a fixed transistor ratio.

$\phi_1$  high:  $T_3$  on,  $T_2$  off.  $V_{out}$  charges via  $T_3$  to  $V_p - V_{th}$ , pre-charging output high.

$\phi_1$  low and  $\phi_2$  not yet high: all OFF.  $V_{out}$  held by change on  $C_{out}$ .

$\phi_2$  high:  $T_2$  on,  $T_3$  off. • If  $V_{in} = 0$ ,  $T_1$  OFF.  $V_{out}$  stays high.

• If  $V_{in} = 1$ ,  $T_1$  &  $T_2$  on,  $C_{out}$  discharges  $V_{out} = 0$  after  $\phi_2$  all OFF.  $V_{out}$  held dynamically on  $C_{out}$ .

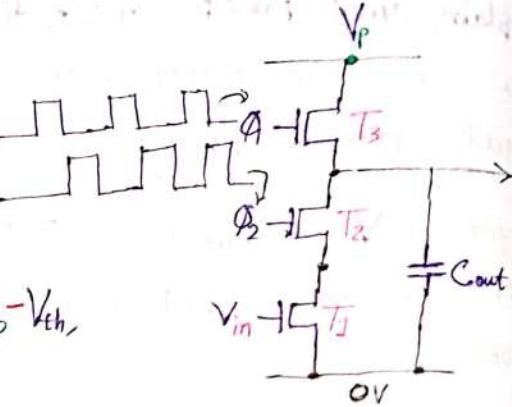


Fig: NMOS ratioless inverter.

Key point: Inversion achieved dynamically, output valid only during  $\phi_2$ , saves area by using minimum geometry NMOS devices.

## ■ NMOS Buffer circuit:

Buffer is formed by connecting two inverters in series.

- The first inverter inverts the input signal.
- The second inverter re-inverts it, restoring the original logic level.

This double inversion provides isolation, stronger drive capability and signal regeneration without changing logic.

It ensures clean transitions, improves noise immunity and allows the signal to drive larger loads. Result, output = input but with higher strength and stability.

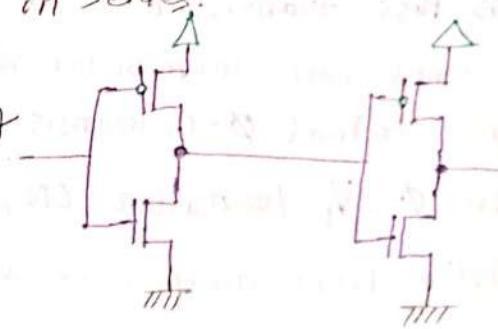
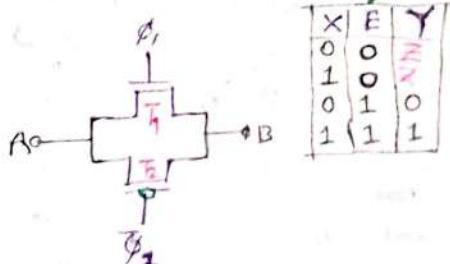


Fig: Buffer circuit using 2 inverters

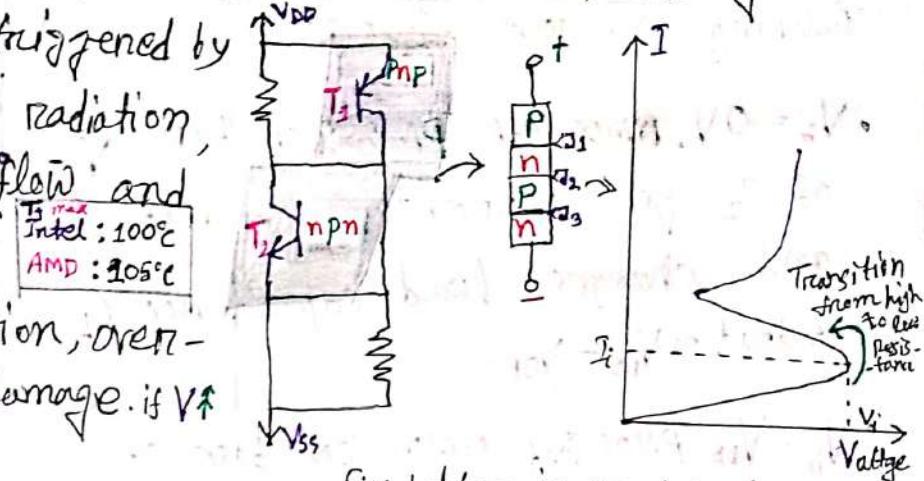
## ■ Transmission gate:

A CMOS transmission gate is bidirectional digital switch formed by a parallel combination of an NMOS and PMOS transistors. It acts as an ideal switch transferring input signals to the output when enabled and entering a high impedance state when disabled.



✓ Explain Latch-up in cmos circuit. Book puchnell \$2.13

Latch-up occurs in CMOS due to parasitic 'pnp' and 'npp' transistors forming a positive feedback path between  $V_{DD}$  and  $V_{SS}$ , creating a low resistance short. It is triggered by supply glitches, noise or radiation, causing large current flow and potential device failure.



**Problems:** High power dissipation, over-heating and permanent damage if  $V_i$  ↑

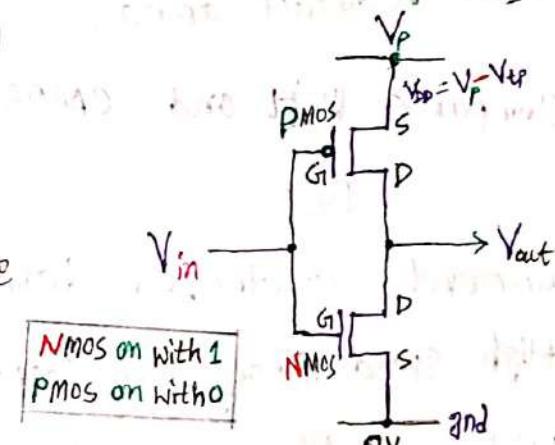
**Solution:**

- Bi-CMOS uses bipolar transistors for base/discharge paths, reducing substrate current injection.
- Improves latch-up immunity by bypassing parasitic conduction paths.
- Provides higher speed and robustness compared to pure CMOS.

✓ Why PMOS transistor passes 1's well but 0's poorly? and vice-versa for NMOS

PMOS Pass transistor passes logic 1 well because source at  $V_{DD}$ , so  $V_{out}$  can rise close to  $V_{DD}$ . Passes logic zero(0) poorly since  $V_{ds}$  must exceed  $N_{tp1}$ ; output stops at  $N_{tp1}$  above 0.

NMOS passes logic 0 well because source at gnd, so  $V_{out}$  falls to 0. Passes logic 1 poorly since  $V_{out}$  stops at  $V_{DD} - V_{th}$ .



Limitation comes from threshold  $V_t$  voltage during conduction.

Fig: CMOS inverter

**Bi-CMOS:**

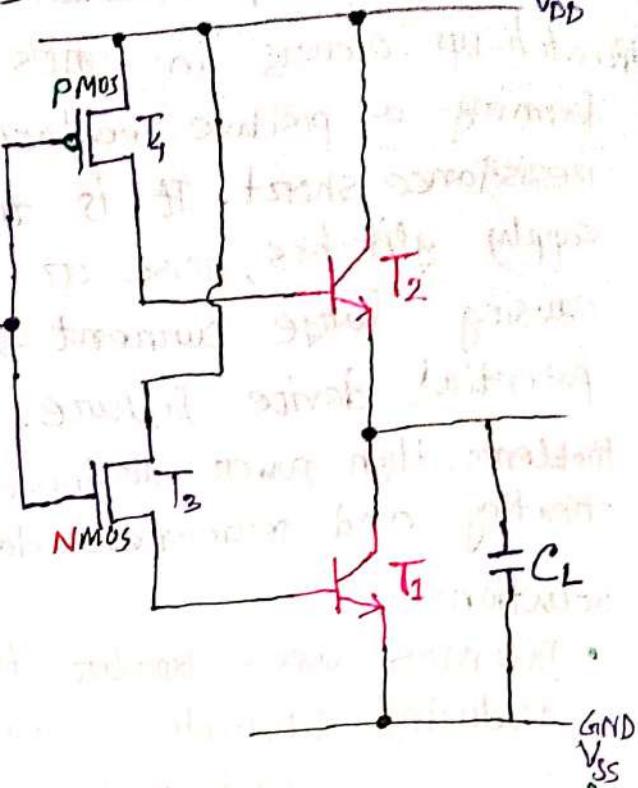
Combines CMOS logic (low static power, high input impedance, scalability) with BJT drive (high speed, low output resistance, large current drive)

Result: circuit with low power + high speed, ideal for driving large capacitive loads.

Very high current driving capability with small area, faster than CMOS-only inverters.

- ✓ Explain the operation of a Bi-CMOS inverter and its features.
- ⊕ Bi-CMOS inverter combines MOS logic control with BJT drive strength
- Referring to the figure:

- $V_{in} = 0V$ , NMOS 'off', PMOS 'ON', base of  $T_2$  gets current,  $T_2$  conducts and charges load capacitance  $C_L$ . Output  $\approx V_{DD} - V_{BE T_2}$
- $V_{in} = V_{DD}$  PMOS 'off', NMOS 'ON', base of  $T_1$  gets current,  $T_1$  conducts discharges load. Output  $\approx V_{CE T_1 (\text{sat})}$ .



Features:

- High input impedance CMOS,
- low output impedance BJT.
- High current drive
- fast charging discharging
- Noise margin close to rail voltages
- compact area.

- ✓ Compare BJT and CMOS technology.

Fig: A simple Bi-CMOS inverter.

Drawback: There is a direct DC path from  $V_{DD}$  to GND through  $T_2$  &  $T_3$  when  $V_{in} = V_{DD}$ . Which causes static power consumption. Also there is no discharging path for current from the base of either BJT when it is being turned off.

BJT	CMOS
<ul style="list-style-type: none"> <li>• current controlled device</li> <li>• High Speed, strong drive capability</li> <li>• Higher static power dissipation</li> <li>• Better analog performance (gain, linearity)</li> <li>• Large area, more complex fabrication.</li> <li>• Faster switching for high current loads.</li> </ul>	<ul style="list-style-type: none"> <li>• Voltage-controlled device</li> <li>• Low power consumption, scalable.</li> <li>• Almost zero static power dissipation.</li> <li>• Better digital performance (noise margin, density)</li> <li>• Smaller area, high integration density.</li> <li>• Slower for heavy loads but energy efficient.</li> </ul>

Q What do you understand by lambda based design rules? Explain how this affects VLSI

A  $\lambda$  is a scalable design parameter representing half the minimum feature size of a technology. It defines minimum width, spacing and overlap rules for IC layout independent of absolute dimensions.

Example: If  $\lambda = 0.125 \mu\text{m}$ , minimum channel length =  $2\lambda = 0.25 \mu\text{m}$ . Metal-metal spacing may be  $3\lambda$ , poly width  $2\lambda$  etc.

#### Impact on VLSI:

- Simplifies design portability across technologies (only  $\lambda$  value changes, layout remains same).
- Reduces design complexity and supports design reuse.
- Ensures manufacturability by accounting for lithography limits, misalignment, and over-etching.
- However  $\lambda$ -based rules are conservative and may not fully exploit advanced technology scaling.

B An alternative BiCMOS inverter with no static current flow.

An improved version of BiCMOS inverter, where DC path through  $T_3$  &  $T_4$  is eliminated but the O/P voltage swing is now reduced.

Hence  $T_2$  is in floating, when  $T_4$  is off.

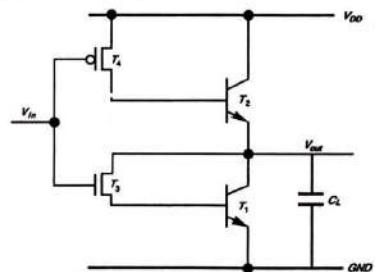


FIGURE 2.18 An alternative BiCMOS Inverter with no static current flow.

Q) Improved BiCMOS inverter using MOS transistors for base current discharge.

Hence the added MOS transistors  $T_5$  and  $T_6$  acts as base discharge transistors, providing low resistance path to discharge the base currents of  $T_2$  and  $T_1$  quickly improving the switching speed.

- $V_{in} = 0V$   $T_3$  'off'  $T_1$  non-conducting,  $T_4$  on, supplies current to base of  $T_2, T_6$  turn 'on' and discharge the base current of  $T_1$  when it is off.
- $V_{in} = V_{dd}$ ,  $T_4$  is off  $T_3$  'on' and supplies base current to  $T_1$ .  $T_5$  is turned on and discharge the base of  $T_2$  to ground.  $T_2$  is turned 'off'

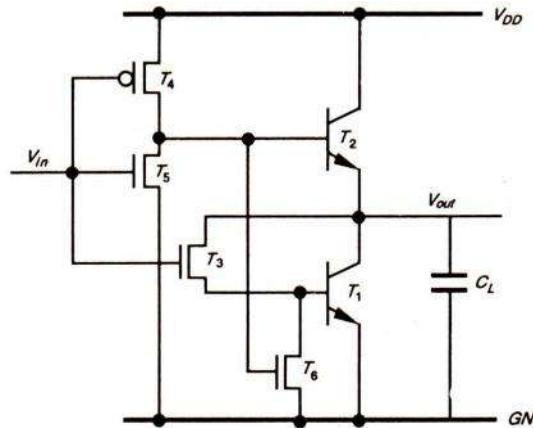


FIGURE 2.20 An Improved BiCMOS Inverter using MOS transistors for base current discharge.

Q: What are the core technological differences between NMOS, CMOS, and BiCMOS?

Feature / Tech	NMOS	CMOS	BiCMOS
Power Consumption	High (static current flows)	Very low (only switching power)	Moderate (lower than NMOS, higher than CMOS)
Speed	Fast but limited by resistive loads	Moderate speed, lower drive capability	High speed (better drive strength from bipolar)
Integration Density	High but limited by power dissipation	Very high (dominant in VLSI)	Lower than CMOS due to added bipolar devices
Noise Margin	Poor	Excellent	Good (improved over NMOS, less than CMOS)
Fabrication Complexity	Simple	Moderate (needs n & p MOSFETs)	Complex (MOS + Bipolar integration)

Q) Explain the following:

(i) Hot carrier effect:

- High electric field near the drain accelerate carriers (electrons/holes)
- Some carriers gain enough energy ('hot') to get trapped in gate-oxide.
- Leads to threshold voltage shift, reduce mobility, device degradation.

(ii) Short - channel Effect:

- In scale down mosfets, channel length is comparable to depletion region width.
- Gate loses control over channel, lower  $\rightarrow$  threshold voltage, higher leakage currents.

(iii) Drain Induces Barrier lowering:

- At high drain voltage, depletion region extends toward the source.
- Reduces Source-channel barrier, threshold voltage decreases.
- Causes subthreshold leakage.

(iv) Punch through:

- At very short channel lengths source and drain depletion region merge.
- Current flows uncontrollably, bypassing gate control.
- Leads to loss of transistor action.

(v) Body effect:

- Substrate bias changes threshold voltage.
- When source to body voltage increases,  $V_{th}$  increases
- Reduces drive current, affects switching speed.

(vi) Latch up:

- In CMOS parasitic pnp - npn transistors form SCR (thyristor).
- A small trigger causes short circuit between  $V_{dd}$  & gnd.
- Leads to device failure unless mitigated (guard rings, well tops)

# Stick Diagrams



## Stick Diagram:

A stick diagram is a simplified, color-coded sketch of a VLSI layout showing relative placement of transistors, interconnections and layers without exact dimensions.

It is used for quick visualization before detailed layout.

'Stick' diagrams are means of capturing topography and layer information using simple diagrams'

Example: Drawing PMOS and NMOS with polysilicon gate, diffusion and metal lines to represent a CMOS inverter.

### Advantages:

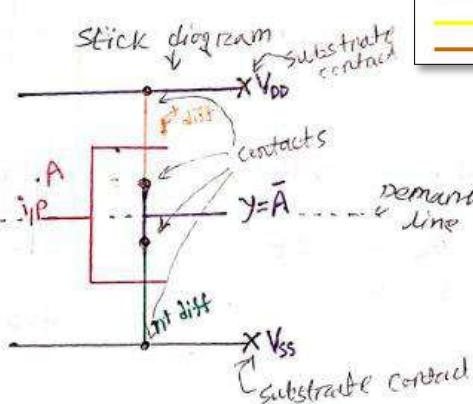
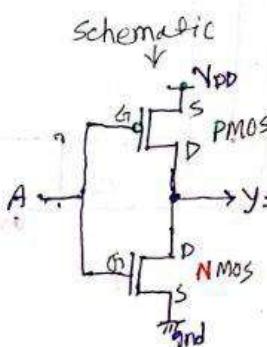
- Quick layout planning.
- Easy to understand and modify
- Shows layer interaction clearly

### Disadvantages:

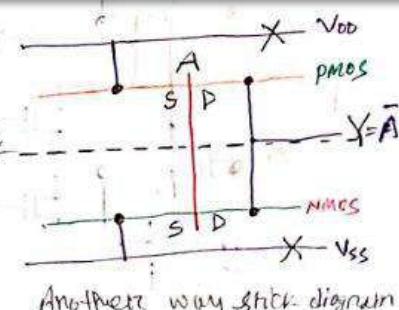
- Not to scale (no exact sizes)
- Cannot predict parasitic effects
- Needs detailed layout for fabrication.

#### Stick diagram of inverter circuit.

Input, A : output  $\bar{A}$



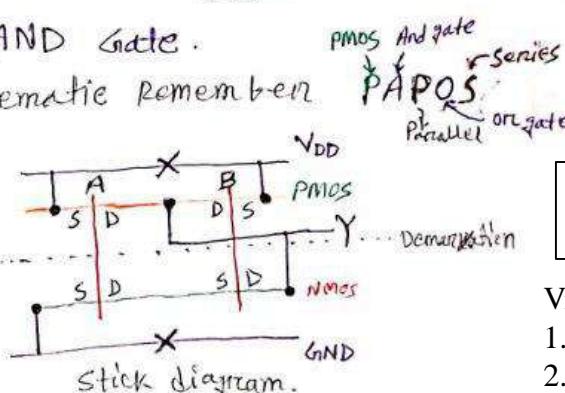
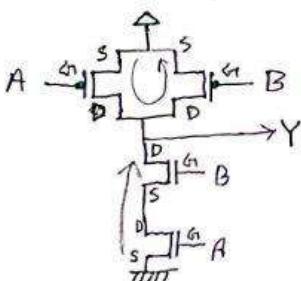
diffusion (device well, local interconnect)
polysilicon (gate electrode, interconnect)
metal (contact, interconnect)
contact windows
depletion implant
P well (CMOS devices)



Another way stick diagram

#### Stick diagram of NAND gate.

Output,  $Y = \overline{A \cdot B}$ , For schematic remember



PMOS And gate  
Parallel series  
Parallel or gate

NMOS series  
NAND-Parallel  
And gate or gate

Never use Red color ink pen in the exam

Video Link You can click

1. [Inverter Stick](#)

2. [NAND Stick](#)

3. [NOR Stick](#)

4. [\(A\(B+C\)+DE\)' Stick](#)

✓ Stick diagram for NOR gate

Output equation:  $Y = \overline{A+B}$

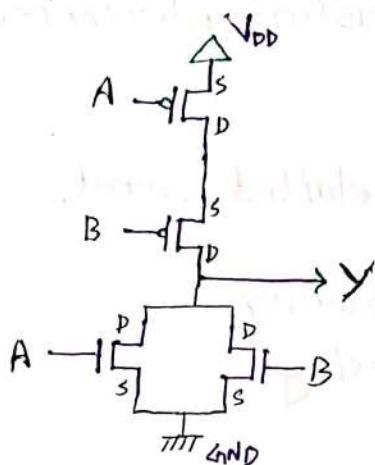


fig: Schematic of  $\overline{A+B}$

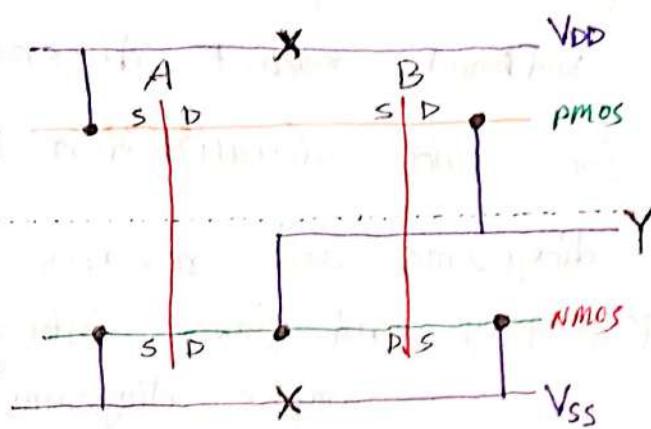


fig: Stick diagram of  $\overline{A+B}$

\* Draw the stick diagram for the boolean function  $Y = \overline{A(B+C)} + DE$

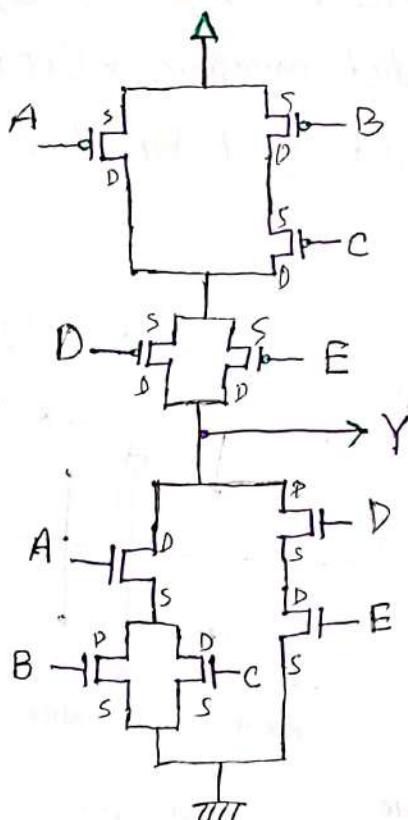


fig: Schematic of Y

Euler's path: CBADE

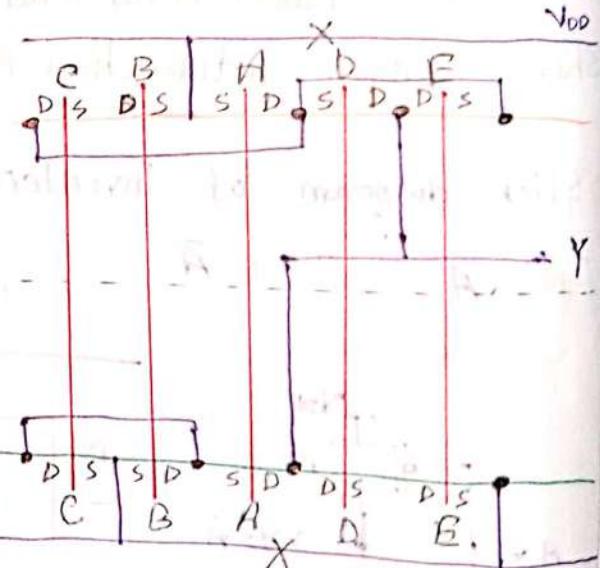
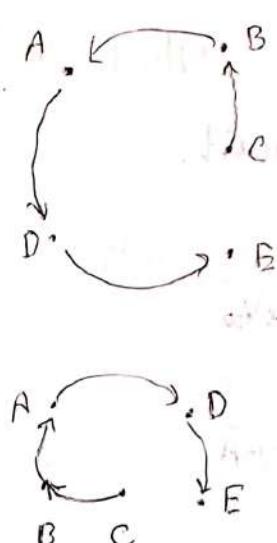
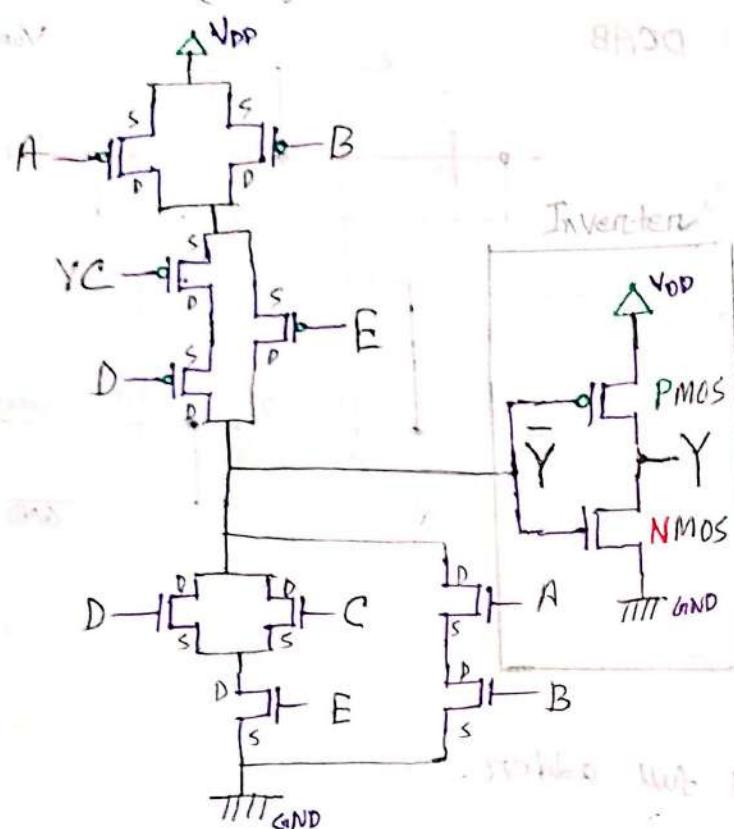


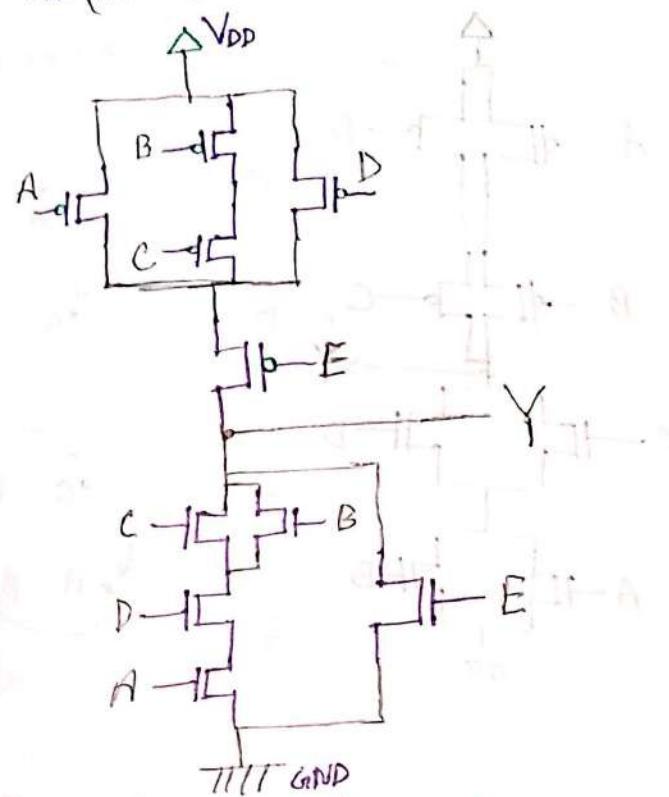
fig: Stick Diagram of Y.

✓ Draw the schematic for the following boolean expression using CMOS

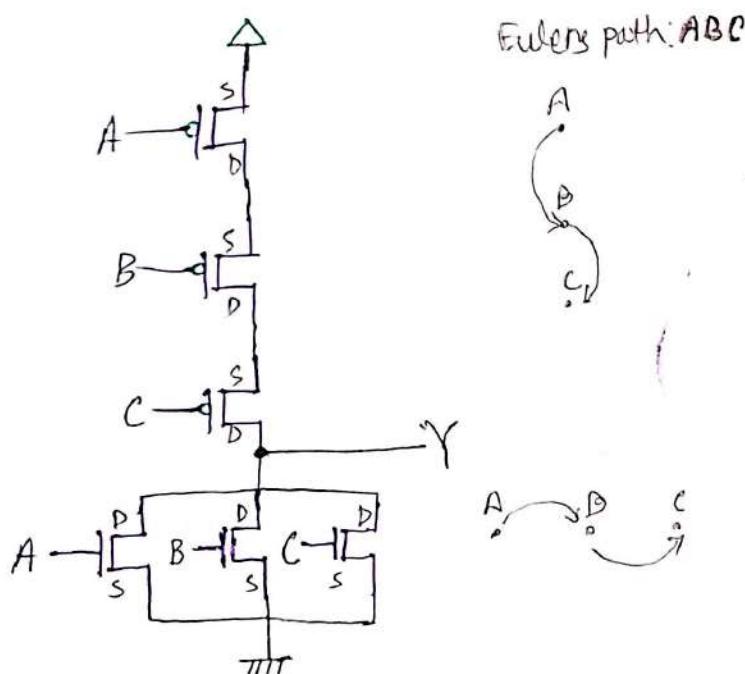
$$(i) AB + (C+D)E$$



$$(ii) (A(B+C)D + E)'$$



✓ Draw the stick diagram of the equation  $Y = \overline{(A+B)} + C$



Faulkes path: ABC

Verified

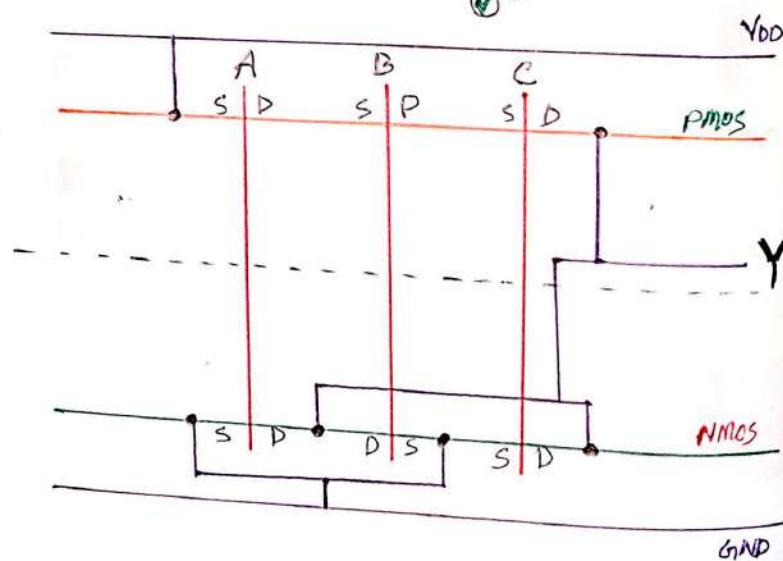
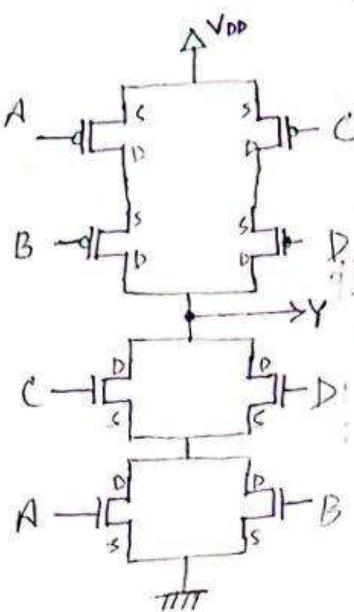


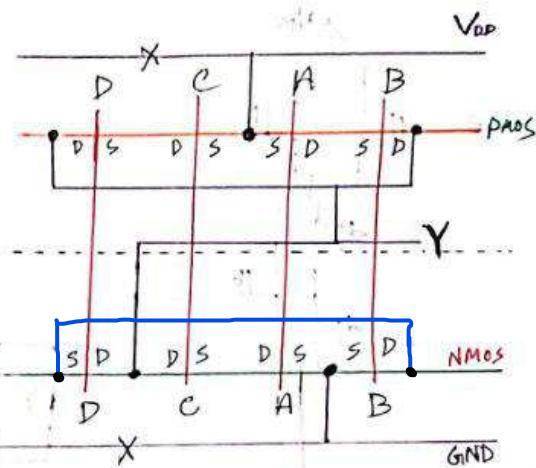
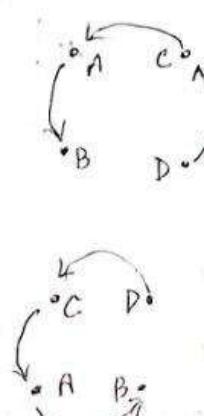
Fig: Stick diagram of  $\overline{(A+B)} + C$

Figs Schematic of  $Y = \overline{AB} + C$

④ Draw the stick diagram of  $((A+B)(C+D))$ .



Failure path DCAB



⑤ Draw the stick diagram of a 1 bit full adder.

#### Website Link

Full Adder Boolean Equations:

- Sum (S) =  $A \oplus B \oplus Cin$
- Carry (Cout) =  $(A \cdot B) + (Cin \cdot (A \oplus B))$

Stick Diagram Description:

- Use two XOR gates for Sum implementation ( $A \oplus B$ , then  $\oplus Cin$ ).
- Use two AND gates + one OR gate for Carry.
- In stick diagram:
  - Polysilicon lines (red) cross diffusion (green) → MOSFETs.
  - Metal lines (blue) for interconnections.
  - PMOS network at top, NMOS at bottom (CMOS style).
  - Sum path uses cascaded XOR transistor networks.
  - Carry path uses parallel-series NMOS/PMOS implementing  $(A \cdot B + Cin)$ .

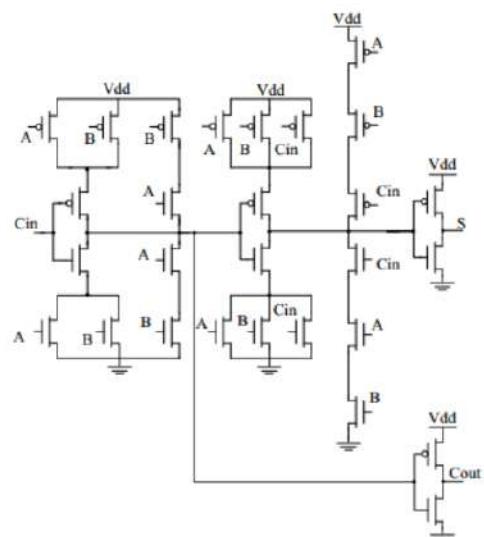


Fig: Schematic of 1 bit Adder

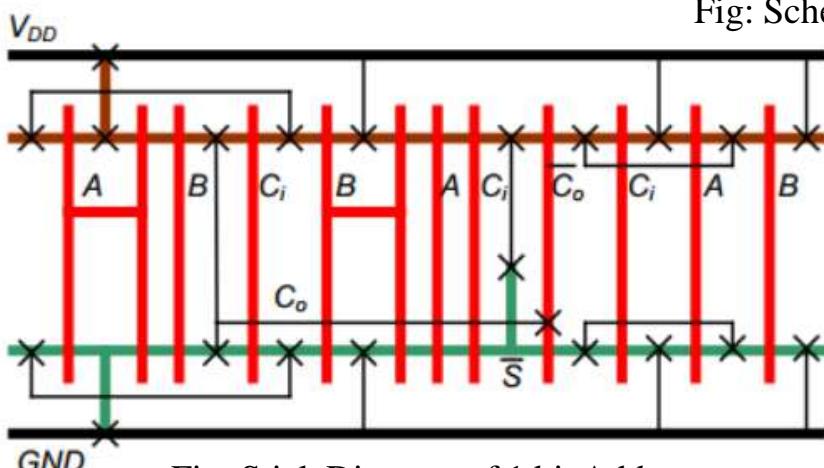


Fig: Stick Diagram of 1 bit Adder

Never use Red color ink pen in the exam

**Q: How do you improve switching speed within a specific gate?**

**Ans:**

Switching speed can be improved by:

1. Reducing load capacitance – minimize transistor and wiring capacitances.
2. Using smaller channel lengths (L) – shorter channels increase current drive.
3. Increasing transistor width (W) – enhances drive strength and reduces delay.
4. Optimizing threshold voltage ( $V_{th}$ ) – lower  $V_{th}$  increases speed (at the cost of power).
5. Reducing supply voltage variation – ensures faster transitions.
6. Using proper transistor sizing – balance rise/fall times for symmetry and speed.

**Q: What are the advantages of dynamic logic over static logic? Also write down its disadvantages.**

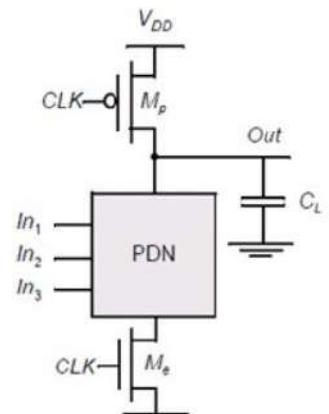
**Ans:**

**Advantages:**

1. Requires fewer transistors → smaller area and higher density.
2. Faster operation due to reduced load capacitance.
3. Better performance in high-speed circuits.
4. Simplifies complex logic functions using fewer stages.

**Disadvantages:**

1. Sensitive to noise and charge leakage → reduced reliability.
2. Requires clocking and careful timing (precharge and evaluate phases).
3. Higher power consumption due to switching and clock load.
4. More difficult to design and test than static logic.

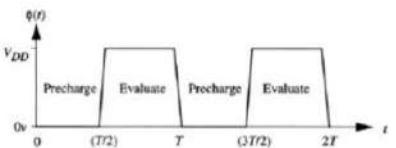


Phase	Clk	Inputs	Outputs
Precharge	Low	Don't care	High
Evaluation	High	Valid	Valid

**Q: Explain Precharge and Evaluate phases (with diagram).**

**Ans:**

Dynamic logic circuits operate in two phases controlled by a clock:



- **Precharge Phase (Clock = 0):** The output node is precharged to a high logic level (usually  $V_{DD}$ ) through a pMOS transistor.
- **Evaluate Phase (Clock = 1):** The pull-down nMOS network is activated; if the logic condition is true, it discharges the output to low.

This approach improves **speed** and **reduces transistor count**, but requires careful **timing** to avoid charge sharing or leakage.

## Array Subsystems

Array subsystems are fundamental building blocks of memory (RAM/ROM) arranged in rows (word lines) and columns (bit lines) to store large amount of data efficiently.

Example: • In DRAM, each cell = 1 transistor + 1 capacitor

- Cells are arranged in  $M \times N$  matrix.

- A row decoder selects a word line and a column decoder selects bit lines to read/write data.

■ Why not sequential circuits for storage?

■ Sequential circuits (flip-flops, latches) store bits but are area and power-inefficient for large data (need multiple transistors/bits)

- RAM/ROM always scale better, enabling millions of bits in compact space with less power and cost.

■ What do you understand by volatile and non-volatile type memory?

■ Volatile Memory: Requires continuous power supply to retain stored data. When power is turned off data is lost. Usually faster and used for temporary storage (working memory). Example: SRAM used in cache, DRAM used in main RAM.

Non-Volatile memory: Retains data even without power. Used for permanent and long term storage. Slower compared to volatile memory but essential for program/data storage.

Example: ROM, Flash memory, EEPROM, SSD.

In short: **Volatile:** temporary, fast, power dependent.

**Non-Volatile:** Permanent, slower, power independent.

Different types of Memory for storage:

RAM: Data lost when power off; volatile memory.

- Types
- SRAM: Static feedback (6T cell), fast, stable but large area.
  - DRAM: 1T + 1C cell, stores charge, compact but needs refresh.

ROM: Retain data without power, non-volatile memory.

Types: Mask ROM: Programmed at fabrication.

PROM: One time programmable (fuse).

EPROM: UV erasable.

EEPROM: Electrically erasable.

Flash: Block erase, dense widely used.

Different types of storage elements are used for balancing speed, density, power and cost across applications.

Example: SRAM for cache, DRAM for main memory, Flash/EEPROM for firmware storage.

## ■ Cybersecurity Perspective

- RAM critical for malware analysis (often hides in volatile).
- ROM/Flash can store secure boot code but also exploited for firm-ware level attacks.
- Backups in non-volatile memory ensures recovery after ransomware or system crashes.

SAM: Serial access Memory.

- Data is accessed in fixed sequence (not random).
- Slower than RAM, but simple and area-efficient.

Used in PISO, SISO, LIFO, FIFO type applications. in buffers, registers, stacks and queues for sequential data processing.

✓ Explain the operating principle for a 6T SRAM cell along with read and write operations and cell stability conditions.

④ 6T SRAM cell consists of 6 transistors, of them

- Two cross-coupled inverters form a latch to store data nodes  $Q$  and  $Q-b$

- Two access transistors  $A_1$  and  $A_2$

Connect cell to bitlines via wordline.

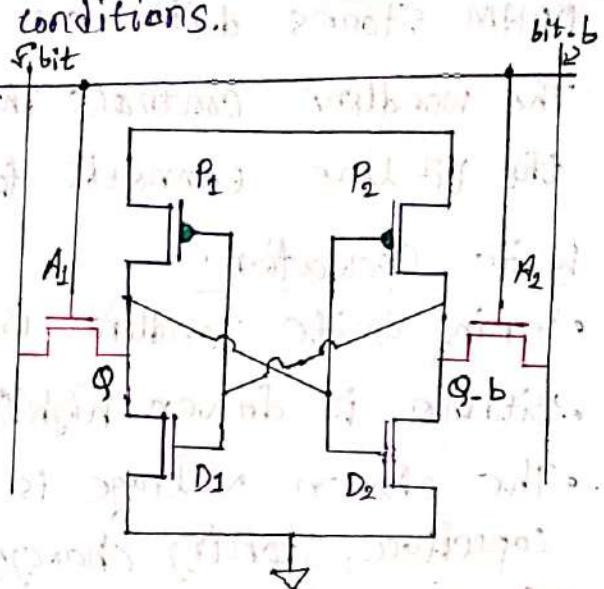


Figure 1: 6T SRAM cell

### Read operation:

- Wordline is raised  $\rightarrow$  access transistor on.
- Both bitlines precharged high.
- Since  $Q=0$ ,  $Q-b=1$  node  $Q$  slightly rises but stays below threshold. Read Stability.
- Bitline (bit) starts discharging through  $Q=0$  side  $\rightarrow$  bit goes low.
- Bit-b remains high, sense amplifier detects differences  $\rightarrow$  cell value '0' is read correctly.

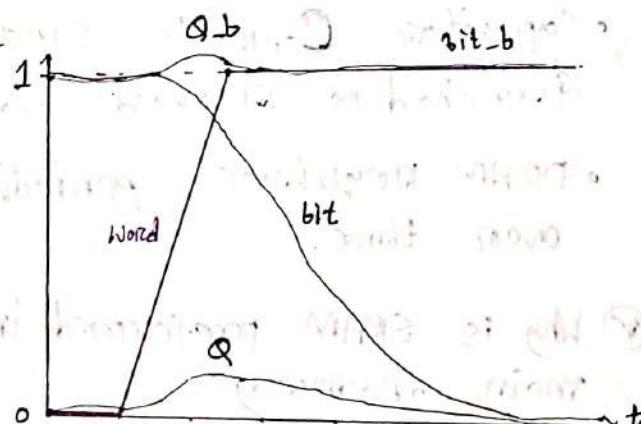


Fig 2: Read operation of 6T SRAM cell

### Write operation:

- Wordline raised, write drivers force values on bit / bit-b.
- Assume writing 1 : bit-b driven LOW, bit Kept HIGH.
- This pulls  $Q-b \rightarrow 0$ , breaking feedback latch.
- As  $Q-b$  falls,  $Q$  flips to 1 and stabilizes.
- Access transistor must overpower PMOS pull-up  $\rightarrow$  ensures writability.

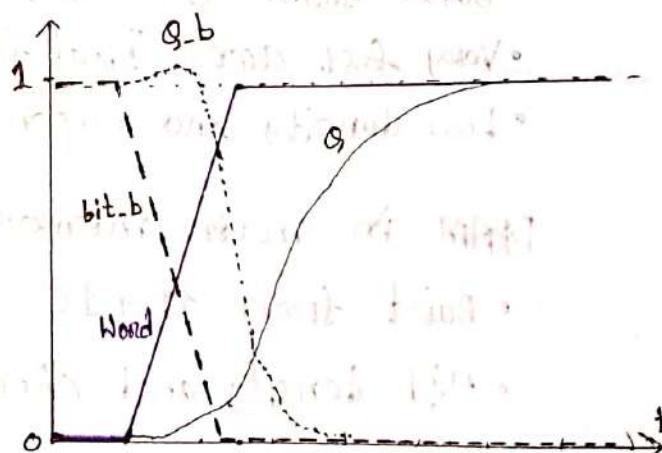


Fig 3: Write operation of 6T SRAM cell

DRAM: DRAM stores data as charge on a capacitor ( $1T+1C$ ) cell. The wordline controls the access transistor, the bitline connects to the capacitor.

### Write Operation:

- During write Wordline is active (on).
- Bitline is driven high(1) or low (0)
- The chosen voltage is forced onto the cell capacitor, storing charge for 1) or discharging (for 0).
- Capacitor  $C_{cell}$  is small and often 3D structured (trenched or stacked) for higher density.
- DRAM requires periodic refreshing since charge leaks over time.

Q Why is SRAM preferred in cache memory, while DRAM is used in main memory?

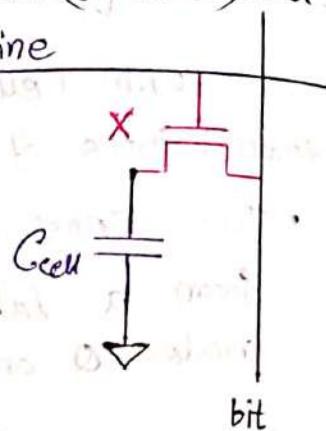
A SRAM in cache memory because:

- Built from 6T transistor cells, hence no refresh needed.
- Very fast access, ideal for CPU cache where speed is critical.
- Low density and expensive, so only small capacity used.

DRAM in main memory because:

- Built from  $1T+1C$  cells, hence requires refresh.
- High density and cheaper, allows storing GB's of data.
- Slower than SRAM, but acceptable for main memory where capacity > speed.

In on-line cache demand speed, SRAM is used while main-memory demands capacity & cost-effectiveness DRAM



# Section B

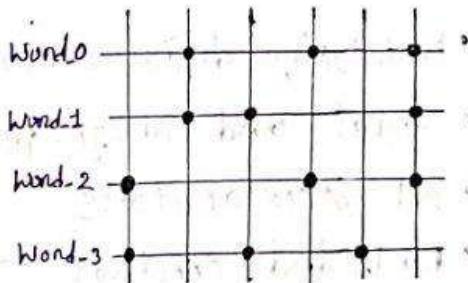
- Fabrication
- HDL/VHDL
- VLSI Physical Design Course Playlist

## ④ ROM

Read only Memory Rom is a non-volatile memory that permanently stores data even without power.

- Built using 1T transistors per bit, typically in a NOR array configuration.
- Each cell has an NMOS transistor connected to a bitline, its gate connects to wordline.
- A dot in the array represent the presence of a transistor (logic 1).

Word-0 → 010101  
Word-1 → 011001  
Word-2 → 100101  
Word-3 → 101010



Reading: When a wordline is activated, all transistors in that row conduct according to stored dots, pulling corresponding bitlines low (0).

- Non-dotted positions remain high (1).

The rom internally stores '1' as dot, but outputs 0s for those cells because of the NOR, that gives complement.



## ⑤ FPGA

FPGA is a configurable integrated circuit that can be programmed after manufacturing to implement any digital logic function.

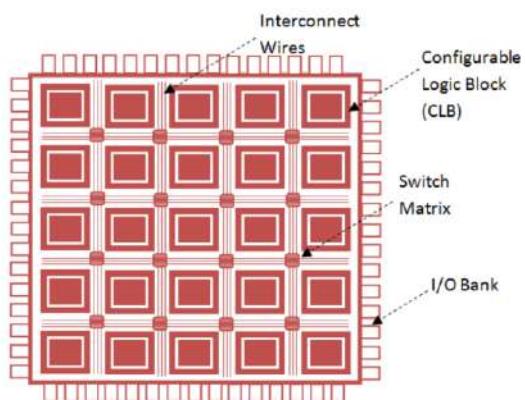
Architecture:

1. CLB: Configurable logic block contains LUTs, look-up tables, flip-flops and multiplexers to perform logic and storage

2. Programmable interconnects: Allow flexible routing between CLBs and I/O blocks.

3. I/O blocks: Interface FPGA with external devices.

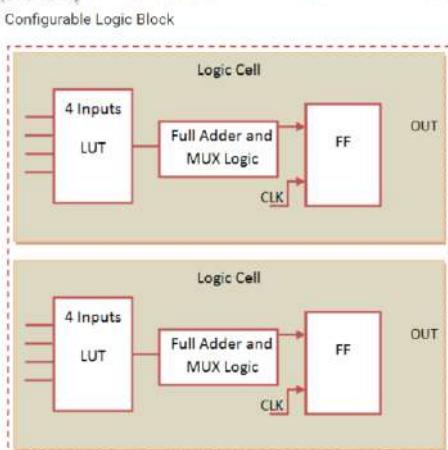
4. Configuration memory: Stores programming data defining logic and routing.



Working: The design is written in HDL (Verilog/VHDL) synthesized and loaded into the FPGA, configuring the interconnections and logic functions.

Use cases:

- Prototyping digital circuits.
- Signal and image processing.
- AI accelerators.
- Embedded control systems.
- Communication hardware (routers, modem).



### ✓ HDL:

Hardware Descriptive Language is used to describe, design and simulate digital hardware at various abstraction levels (behavioral, structural or RTL).

Languages like Verilog and VHDL defines how circuits behave and interconnect.

In VLSI automation, HDL code is the first step - it is synthesized into gate level netlists, then mapped, placed and routed automatically to generate the final chip layout.

HDLs enable automation, verification and reusability of complex VLSI designs, reducing manual circuit design time and errors.

### ✓ What does VHDL stands for? Explain the reason for using VHDL?

✓ VHDL-Stands for Very High Speed Integrated circuit Hardware description language. It is used for

1. Describes digital systems at multiple abstraction levels (behavioral, structural, RTL)
2. Allows simulation and verification before hardware fabrication.
3. Supports concurrent processing, matching real hardware behavior

4. facilitates design portability across different technologies and tools.
  5. facilitates automation of synthesis, optimization and layout in VLSI.
  6. promotes reusability of modular and parameterized designs.
- Write down the complete VHDL code for single bit half adder with enable pin.

```
library IEEE;
use IEEE.STD_LOGIC_1164.ALL;
```

#### -- Entity Declaration

```
entity half_adder_en is
```

```
Port(
```

```
    a : in STD_LOGIC;
    b : in STD_LOGIC;
    en : in STD_LOGIC;
    sum : out STD_LOGIC;
    carry : out STD_LOGIC;
```

```
);
```

```
end half_adder_en;
```

#### -- Architecture definition

```
architecture Behavioral of half_adder_en is
```

```
begin
```

```
process (a,b,en)
```

```
begin
```

```
    if en=1 then
```

```
        sum <= a XOR b;
```

```
        carry <= a AND b;
```

```
    else
```

```
        sum <= '0';
```

```
        carry <= '0';
```

```
    end if;
```

```
end process;
```

```
end Behavioral
```

④ Write VHDL code for 4 to 1 MUX.

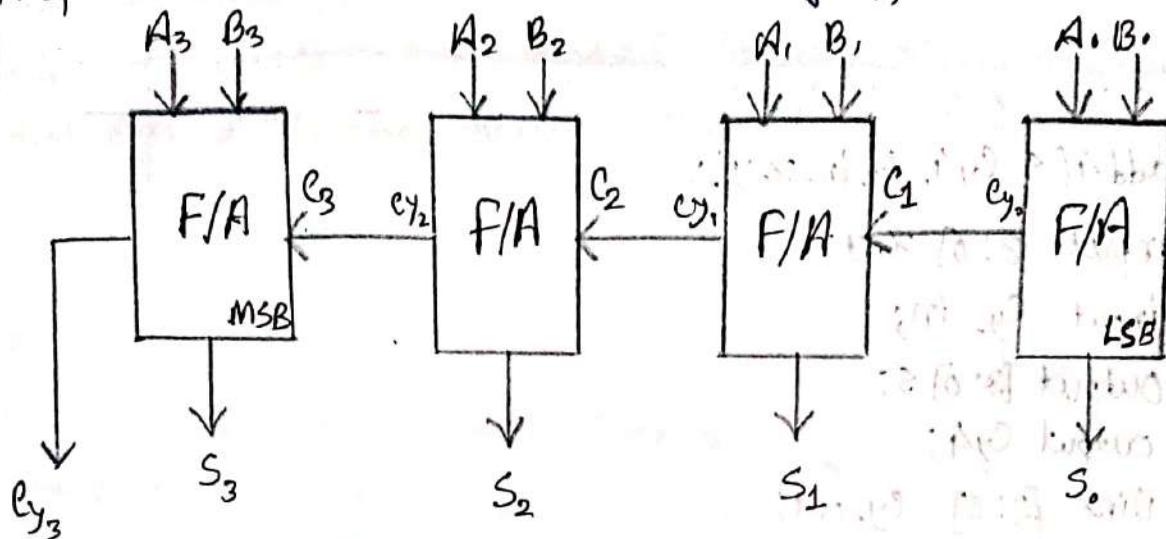
```
④ library IEEE;
use IEEE.STD_LOGIC_1164.ALL;
entity mux4to1 is
    Port (
        d0, d1, d2, d3 : in STD_LOGIC;
        sel : in STD_LOGIC_VECTOR (1 downto 0); -- 2bit select
        y : out STD_LOGIC);
end mux4to1;
architecture Behavioral of mux4to1 is
begin
    process (d0, d1, d2, d3, sel)
    begin
        case sel is
            When "00" => y <= d0;
            When "01" => y <= d1;
            When "10" => y <= d2;
            When "11" => y <= d3;
            When others => y <= '0';
        end case;
    end process;
end Behavioral;
```

⑤ What is the significance of simulation in HDL based design?

④

- Allows verification of logic before hardware fabrication, saving cost and time.
- Detects functional errors, timing issues, and logical bugs early.
- Helps test design under different input pattern and corner cases.
- Ensures compliance with specifications and expected behavior.
- Supports iterative debugging and refinement of the design.
- Reduces risk of hardware failure and increases design reliability.

A 4 bit adder construction by cascading 4, 1-bit adders.



HDL code:

An algorithmic level description of  $C_y$

module Carrying ( $C_y, a, b, c$ );

input  $a, b, c$ ;

output  $C_y$ ;

assign

$$C_y = (a \& b) | (b \& c) | (c \& a);$$

end module

primitive Carrying ( $C_y, a, b, c$ );

input  $a, b, c$ ;

output  $C_y$ ;

table

//	$a$	$b$	$c$	$C_y$
1	1	?	:	1
1	?	1	:	1
?	1	1	:	1
0	0	?	:	0
0	?	0	:	0
?	0	0	:	0

end table

end primitive

## Structural Representation.

module add4(s, Cy4, Cy-in, x, y);

input [3:0] x, y;

input Cy-in;

output [3:0] s;

output Cy4;

wire [2:0] Cy-out;

add B0 (Cy-out[0], s[0], x[0], Y[0], e1);

add B1 (Cy-out[1], s[1], x[1], Y[1], Cy-out[0]);

add B2 (Cy-out[2], s[2], x[2], Y[2], Cy-out[1]);

add B3 (Cy-out[3], s[3], x[3], Y[3], Cy-out[2]);

end module

module add(Cy-out, sum, a, b, Cy-in);

input a, b, Cy-in;

output sum, Cy-out;

sum S1 (sum, a, b, Cy-in);

carry C1 (Cy-out, a, b, Cy-in);

end module

module sum (sum, a, b, Cy-in);

input a, b, Cy-in;

output sum;

wire t;

xor x1(t, a, b);

xor x2(sum, t, Cy-in);

end module

module carry(Cy-out, a, b, Cy-in);

input a, b, Cy-in;

output Cy-out;

wire t1, t2, t3;

and g1(t1, a, b);

and g2(t2, a, b);

and g3(t3, b, c);

or g4(Cy-out, t1, t2, t3);

end module



## GPT Generated Version

```
1 // 1-bit full adder (reusable)
2 module full_adder (
3     input wire a,
4     input wire b,
5     input wire cin,
6     output wire sum,
7     output wire cout
8 );
9     assign sum = a ^ b ^ cin;
10    assign cout = (a & b) | (b & cin) | (a & cin);
11 endmodule
12
13 // Parameterized N-bit ripple carry adder (N=4 by default)
14 module ripple_adder #(
15     parameter N = 4
16 )(
17     input wire [N-1:0] a,
18     input wire [N-1:0] b,
19     input wire         cin,
20     output wire [N-1:0] sum,
21     output wire         cout
22 );
23     // internal carry chain: carry[0] = cin, carry[N] = final cout
24     wire [N:0] carry;
25     assign carry[0] = cin;
26
27     genvar i;
28     generate
29         for (i = 0; i < N; i = i + 1) begin : FA_CHAIN
30             full_adder fa (
31                 .a    (a[i]),
32                 .b    (b[i]),
33                 .cin  (carry[i]),
34                 .sum  (sum[i]),
35                 .cout (carry[i+1])
36             );
37         end
38     endgenerate
39
40     assign cout = carry[N];
41 endmodule
42
43 // Example instantiation for a 4-bit adder:
44 // ripple_adder #(.N(4)) adder4 (.a(a), .b(b), .cin(cin), .sum(sum),
45 // .cout(cout));
```

# SEMICONDUCTOR WAFER FABRICATION PROCESS



## INGOT GROWTH

A seed crystal is used as a template to grow large cylindrical crystal silicon ingots in a hot quartz crucible.

## SLICING

The ingot is precisely sliced into thin disc-shaped silicon wafers using a specialized diamond saw.



## LAPPING

Rotating abrasive plates grind the wafer surfaces flat and parallel to each other.



## POLISHING

A silica-based slurry combined with mechanical polishing creates smooth, damage-free wafer surfaces.



## EDGE GRINDING

Cracked edges from slicing are removed by mechanically grinding the wafer's periphery.



## ETCHING

Wet chemicals evenly remove surface damage and contaminants, exposing a clean surface layer.



## CLEANING

The wafers undergo chemical baths to remove particles and impurities, qualifying them for device fabrication.

## Reference Videos:

1. [Inside Micron Taiwan's Semiconductor Factory](#)

Branch education

1. [CPU Manufacturing](#)
2. [Transistor Making](#)
3. [Transistor in CPU](#)

Overall Once Again

1. [How are microchips made?](#)
2. [Lithography Process](#)
3. [RCA Cleaning](#)

## Reference Websites

1. [What is Semiconductor?](#)
2. [The Process step by Step](#)
3. [Plasma induced damage](#)



## Fabrication Process:

VLSI fabrication is the process of creating millions of transistors and inter-connections on a single silicon chip. It involves multiple sequential steps to transform a silicon wafer into an integrated circuit.

### Steps:

1. Wafer preparation: silicon ingot grown (Cz method), sliced into wafers.
2. Oxidation:  $\text{SiO}_2$  grown as insulating layer.
3. Photolithography: pattern transfer using light and masks.
4. Etching: Remove unwanted oxide areas.
5. Doping (Ion Implantation/Diffusion): Control conductivity of regions.
6. Metallization: Deposit metal for inter-connection.
7. Testing and Packaging: final inspection, packaging into chips.

### Advantages:

- High device density
- Low cost per function
- Faster performance

### Limitations:

- High fabrication cost
- Complex process control.
- Susceptible to defects

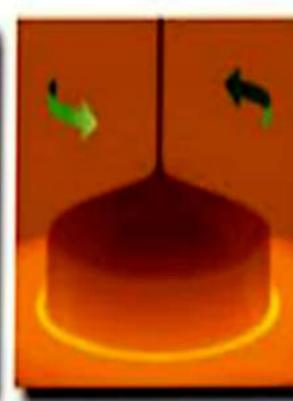
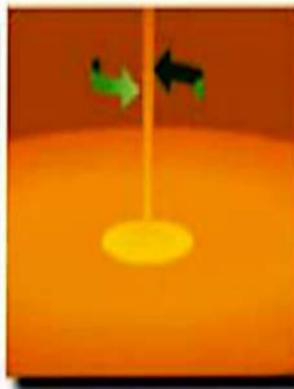
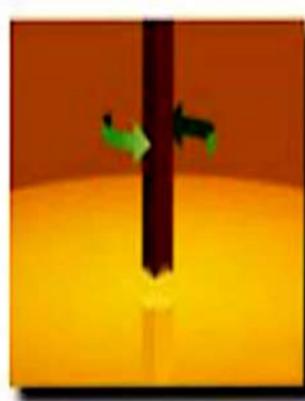
### Wafer Preparation:

The wafer must be created with a very high quality, it must be grown as single crystal silicon body having a very small number of defects and proper type and level of doping must contain. This is accomplished by the 'Czochralski' method. The method is described below.

Raw material preparation: High-purity polycrystalline silicon is melted in a quartz crucible.

Seed Crystal: A small, defect-free silicon seed crystal is dipped into the molten silicon.

# Wafer Processing...



Single Crystal Silicon Ingot

Crystal pulling: The seed crystal is slowly pulled upward while rotating, allowing molten silicon to solidify onto it.

Controlled growth: Temperature, pulling rate and rotation speed are precisely controlled to maintain a uniform cylindrical shape.

Ingot formation: A large single crystal silicon ingot (boule) is formed as the seed pulls more material.

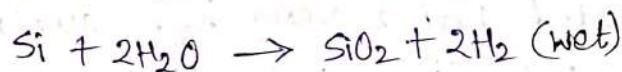
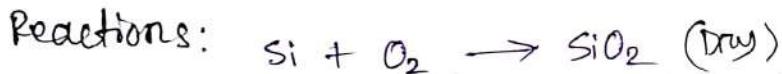
Ingot shaping: The ingot is ground to uniform diameter and its surface defects are removed.

Wafer slicing: The ingot is sliced into thin wafers using a diamond saw, then polished for semiconductor device fabrication.

This method ensures large defect free single crystal wafers for IC manufacturing.

#### ■ Oxidation:

Oxidation is the process of reacting Si with  $O_2$  or water vapor  $H_2O$  at high temperature  $700\text{--}1200^\circ C$  to form a thin layer of  $SiO_2$  on the wafer's surface.



#### ✓ Why oxidation is essential in IC fabrication?

##### ■ Oxidation is essential because

- Provides  $SiO_2$  mask against diffusion/ implantation.
- Used as MOS gate dielectric
- Ensures surface passivation reduces defects.
- Provides device isolation
- Enables multilevel metallization insulation.

⑤ Why is wet oxidation rate is higher than dry oxidation rate?

The values of  $C^*$  (concentrations) for  $O_2$  and  $H_2O$  species in  $SiO_2$  at  $1000^\circ C$  are  $5.2 \times 10^{16}$  and  $3.0 \times 10^{19}$  respectively. Thus  $H_2O$  has higher concentration than  $O_2$  at  $1000^\circ C$ .

Again flux of oxidant is proportional to  $C^*$ , which is almost three orders of magnitude greater for  $H_2O$ , than that of  $O_2$ .

Wet oxidation uses water vapor, which diffuses faster than dry oxidation.

These are the reasons for faster growth rate of wet oxidation than that of dry oxidation.

Though wet oxidation has faster growth rate but dry oxidation gives better electrical characteristics.

⑥ What are the factors that affect the oxide growth rate?

The factors that affect the oxide growth rate are:

1. Crystal orientation of Si affects the oxide growth rate, because reaction rate constant depends on crystal structure of Si surface.
2. Any unintentional moisture accelerates the dry oxidation rate.
3. High concentration of sodium enhances the oxidation rate by changing the bond structure in the oxide. Enhancing the diffusion and concentration of oxygen molecules in the oxide.
4. The common dopant elements of group III and V can enhance the oxidation behaviour.
5. Typically 1~5% addition of dry HCl with dry  $O_2$  increases the oxidation rate.
6. Oxide growth rate is enhanced with temperature and pressure.

① Write down the 'typical' properties of oxide?

ii) The typical properties of oxide include:

- Susceptibility to defect.
- Oxide composition
- Leakage current through oxide.
- Density
- Stress within thin films.
- Etch rate in acid bases.
- Dielectric breakdown strength.
- Refractive index 1.46 for  $\text{SiO}_2$
- Oxide charges ( $Q_{it}$ ,  $Q_f$ ,  $Q_{st}$ ,  $Q_m$ )
- Masking property of  $\text{SiO}_2$

② Describe various types of oxide charge at Si-SiO<sub>2</sub> interface

iii) The oxide charges are.

Interface trapped charges ( $Q_{it}$ ): from defects/impurities at interface, neutralized by H<sub>2</sub> anneal.

Fixed oxide charge ( $Q_f$ ): Positive, near interface, reduced by control cooling/annealing.

Oxide trapped charges ( $Q_{st}$ ): Trapped carriers (holes/electrons), caused by radiations/injection.

Mobile Ionic charge ( $Q_m$ ): due to alkali ions (Na<sup>+</sup>, K<sup>+</sup>), reduced by Chlorine et cleaning or  $\text{Si}_3\text{N}_4$  barriers.

iv) Lithography:

Lithography is a pattern transfer technique that selectively wafer regions so process steps (etching, doping, deposition) apply only where needed.

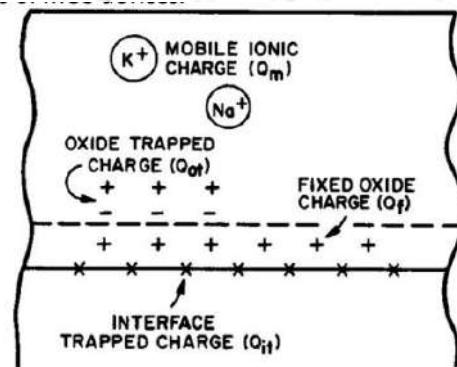


Fig.- Charges in thermally oxidized Silicon

## Types:

- photo-lithography: UV light + mask - the dominant high-throughput method
- Electron-beam lithography: maskless, very high resolution, slow - Used for mask writing & prototypes.
- X-ray lithography: Higher resolution than optical uses X-ray masks
- Ion-radiation/beam lithography: direct write, nano scale capability

✓ with neat sketches, show the step by step lithography process for manufacturing NMOS?

The lithography steps explained below.

1. Wafer prep/clean: remove contaminants so resists adheres and features are defect free.

2. Photo resist coating: apply positive/negative resist uniformly (typ. ~1μm); soft-bake to remove solvent.

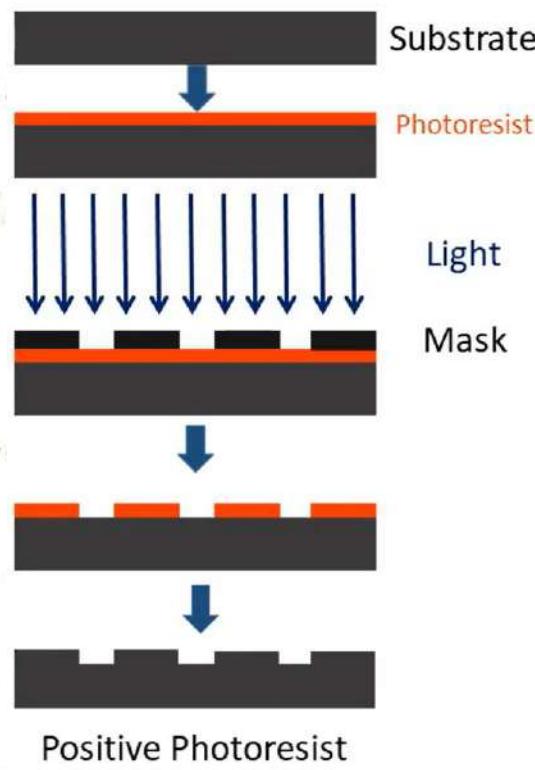
3. Mask alignment and exposure: (stepper) align photomask; expose UV through transparent areas; mask patterns define features.

4. Post-exposure Bake: (PEB) stabilized chemical changes in resist for improved resolution.

5. Photo resist development: dissolve exposed (positive) or unexposed (negative) resist; rinse → pattern resist remains

6. Hard bake: hardens resist for subsequent processing (etch/implant)

7. Etching: Remove underlying material where resist is open; wet chemical or dry plasma are used.



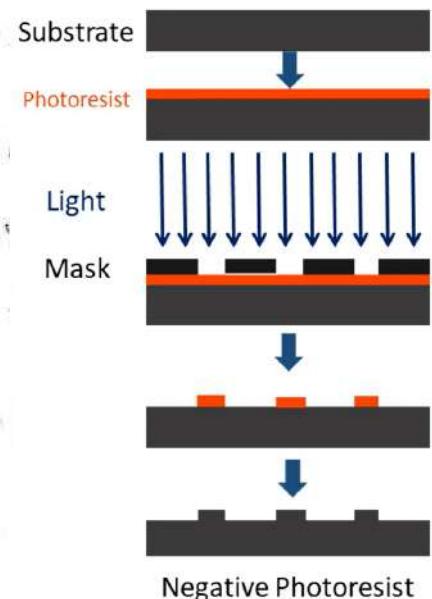
8. Spin Rinse and Dry: A special tool called (SRD) is used to clean the wafer with de-ionized water and dries it with nitrogen.

9. Photo resist removal/ashing: plasma

or solvents strip remaining resist; clean residues.

10. Development of gate: After this aluminium contacts are made from source, drain and gate.

Write down the differences between positive and negative photoresist?



#### Positive Photoresist

- Exposed area becomes soluble in developer.
- Masks transparent region = removed resist.
- Provides better resolution and accuracy
- More widely used in modern IC fabrication
- More expensive than negative resist.

#### Negative photoresist

- Exposed area become insoluble in developer.
- Masks transparent region = protected resist.
- Provides lower resolution.
- Used in older/larger geometry processes.
- Cheaper compared to positive resist.

#### Epitaxy and Epitaxial growth:

Epitaxy is the ordered growth of a mono-crystalline layer that bears a definite relation to the underlying mono-crystalline substrate.

The process of growing an oriented, single crystal layer on a substrate wafer is called epitaxial growth.

## Advantages:

- Provides high-purity, defect-free crystalline layers on substrate.
- Enables precise control over thickness, doping and composition.
- Improves electrical performance of devices (mobility, speed).
- Essential for heterostructures (e.g. GaAs on Si, SiGe layers).
- Reduces defects like lattice mismatch dislocations.
- facilitates scaling and integration in modern VLSI fabrication.

Q) What are the basic difference between homo epitaxy and heteroepitaxy?

Homo-epitaxy	Hetero-epitaxy
<ul style="list-style-type: none"><li>• Layers and substrate are chemically similar.</li><li>• Maintains same crystal structure and lattice.</li><li>• Easier growth, fewer defects.</li><li>• Used for improving substrate quality.</li><li>• Example: P-Si on Si</li></ul>	<ul style="list-style-type: none"><li>• Layer and substrate are different in chemistry/crystal properties.</li><li>• May differ in structure, symmetry, or lattice parameter.</li><li>• More complex growth, risk of lattice mismatch.</li><li>• Used for new material properties integration.</li><li>• Example: AlGaAs on GaAs, GaN on SiC</li></ul>

Q) Briefly describe the CVD, LPE and MBE methods for performing epitaxy.

### ■ Chemical Vapor Deposition (CVD) / Vapor Phase Epitaxy (VPE):

- Uses chemical vapors to deposit single-crystal epitaxial layers.  
Example:  $\text{SiCl}_4 + \text{H}_2 \rightarrow \text{Si} + \text{HCl}$ , where Si depends on heated wafer.
- Wafer placed on heated graphite susceptor in a reaction chamber.
- Doping gases can be added for controlled impurities.

Advantages: Low temperature reduces impurity migration, thin Si on insulators possible.

## Liquid phase Epitaxy (LPE):

- Grows crystal from a liquid solution at temperature below melting point.
- Wafer dipped in a graphite slider into semiconductor melt pool.
- After growth, surface wiped and can be moved for multiple layer growth.

**Advantage:** Lower processing temperature, simpler than CVD.

## Molecular Beam Epitaxy (MBE):

- High vacuum technique: atomic/molecular beams ( $Al, Ga, As$ , dopants) directed at substrate.
- Growth rate are precisely controlled  $\rightarrow$  high crystal quality.
- Used for complex heterostructures (e.g.  $AlGaAs$  on  $GaAs$ ).

**Limitation:** Expensive, requires sophisticated setup.

## Etching:

Etching is the process of removing material from wafer regions not protected by photoresist. It transfers resist pattern into underlying device layers to form circuit features.

### Wet etching:

- Uses liquid chemicals (acids, bases, caustics)
- Example:  $HCl + NH_4F$  used to etch  $SiO_2$
- Widely used, but less precise (isotropic)

### Dry etching:

- Wafer placed in vacuum chamber (10 m Torr, 100°C). Plasma of  $Cl_2 + BCl_3$  used to etch.
- Provides anisotropic etching with sharp vertical patterns.

# Comparison between Wet and Dry Etching

	Wet	Dry
Method	Chemical Solutions	Ion Bombardment or Chemical Reactive
Environment and Equipment	Atmosphere, Bath	Vacuum Chamber
Advantage	<ul style="list-style-type: none"><li>1) Low cost, easy to implement</li><li>2) High etching rate</li><li>3) Good selectivity for most materials</li></ul>	<ul style="list-style-type: none"><li>1) Capable of defining small feature size (&lt; 100 nm)</li></ul>
Disadvantage	<ul style="list-style-type: none"><li>1) Inadequate for defining feature size &lt; 1um</li><li>2) Potential of chemical handling hazards</li><li>3) Wafer contamination issues</li></ul>	<ul style="list-style-type: none"><li>1) High cost, hard to implement</li><li>2) low throughput</li><li>3) Poor selectivity</li><li>4) Potential radiation damage</li></ul>
Directionality	Isotropic (Except for etching Crystalline Materials)	Anisotropic

## ✓ Si oxidation model:

model assumptions: this model is valid for a temperature range of  $700\text{--}1300^\circ\text{C}$ ,  $\text{O}_2 \sim 1 \text{ atm}$  pressure and oxide thickness of  $300\text{--}20,000 \text{ \AA}$ .

The oxidizing species -

- are transported from bulk gas phase to gas-oxide interface with flux  $F_1$  (flux means no. of species crossing a unit area per unit time)
- are transported across the existing oxide towards the silicon with flux  $F_2$ .
- React at the  $\text{Si-SiO}_2$  interface with the silicon  $\text{Si}$  with flux  $F_3$
- at steady state,  $F_1 = F_2 = F_3$

where,

$$F_1 = h(C^* - C_o)$$

$$F_2 = D(C_o - C_i)/d$$

$$F_3 = k_s C_i$$

$C_o$ : equilibrium concentration in the oxide at outer surface.

$C^*$ : equilibrium bulk concentration in the oxide.

$h$ : gas-phase mass transfer coefficient

$C_i$ : oxidizing species concentration in the oxide adjacent to the  $\text{Si-SiO}_2$  interface.

$d$ : oxide thickness and

$k_s$ : reaction rate constant.

$D$ : diffusion coefficient.

After solving at steady state, values of  $C_o$  and  $C_i$  can be obtained

### Limiting cases:

- When  $D$  is very small  $C_i \rightarrow 0$  and  $C_o \rightarrow C^*$  (diffusion controlled case)
- Oxidation rate depends on the supply of oxidant to the interface through the oxide layer.
- When  $D$  is large,  $C_i = C^*$  (reaction controlled case).
- Oxidation rate depends on the reaction rate constant and  $C_o$  or  $C_i$  ( $C_i = C_o$  because  $D$  is high)

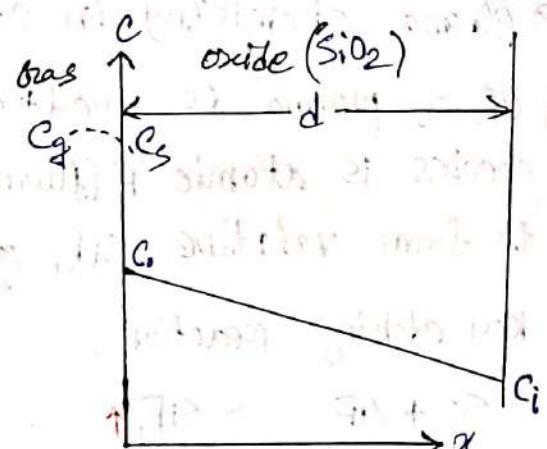
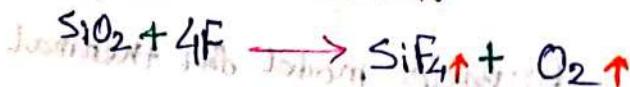


Fig: Basic model for thermal oxidation of silicon.

## ✓ Plasma chemistry in etching:

In  $\text{CF}_4\text{-O}_2$  plasma is used to etch Si or  $\text{SiO}_2$ . The active etching species is atomic F (fluorine) which reacts with Si and  $\text{SiO}_2$  to form volatile  $\text{SiF}_4$  gas that can be easily removed.

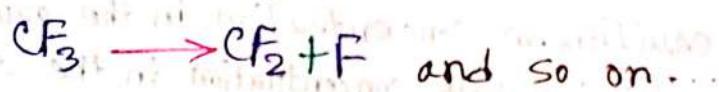
Key etching reaction:



In both cases,  $\text{SiF}_4$  is volatile and escapes, leaving the etched surface behind.

When only  $\text{CF}_4\text{-O}_2$  is used there is no appreciable etching because:

$\text{CF}_4$  molecules undergo dissociative collisions:



and creates F atoms.

However no combination reactions also occurs at the same time.



which reduces concentration of free F atoms.

So. Etching rate depends on F generation and recombination.

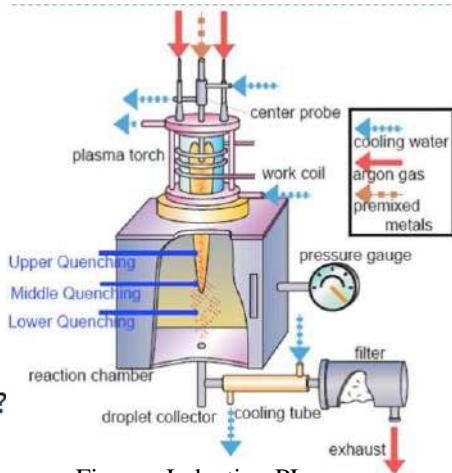
## Q1: What is Plasma? How is it Produced? Explain Induction Plasma.

**Q** Plasma = 4th state of matter, mixture of ions, electrons, neutrals, excited species; electrically neutral but good conductor.

- Produced when external field energizes electrons, causing ionization, excitation & dissociation → self-sustained equilibrium.

### • Induction Plasma:

- Electrodeless, contamination-free discharge (very clean).
- Highly reactive, used in material processing.
- No electrode corrosion → long apparatus life.
- Operates from soft vacuum to several atm (pressure flexibility).
- High-enthalpy, radiative plasma source.



## Q2: What are the Effects of O<sub>2</sub> and H<sub>2</sub> Addition on Etching Rate?

### **Q** O<sub>2</sub> Addition:

- Forms COF<sub>2</sub>, CO, CO<sub>2</sub> → reduces CFx.
- Less CFx-F recombination → more free F atoms.
- Result: Etching rate increases.

### • H<sub>2</sub> Addition:

- With Si → forms fluorocarbon/hydrocarbon polymers on surface.
- Si etching rate decreases due to polymer deposition.
- For SiO<sub>2</sub>, no polymerization → H<sub>2</sub> has negligible effect.



## ■ Wafer cleaning!

Wafer cleaning is an important part of an VLSI fabrication.

### ✓ Explain why wafer cleaning is necessary?

■ Wafer cleaning is necessary, because.

- ICs are highly sensitive to contamination.
- FEOL cleaning; removes impurities from Si or SiO<sub>2</sub> surfaces to ensure device quality.
- BEOL Cleaning; removes particles from metal interconnect layers.
- Prevents defects, yield loss and device failure.

### ✓ Differentiate between hydrophilic and hydrophobic surfaces.

Hydrophilic	Hydrophobic
<ul style="list-style-type: none"><li>• Easily wet by cleaning solution</li><li>• Particles remain in solution until removed.</li><li>• Contact angle (CA) &lt; 90°</li><li>• High surface energy</li><li>• Easier to clean during wafer process.</li><li>• Example SiO<sub>2</sub> surface.</li></ul>	<ul style="list-style-type: none"><li>• Difficult to wet by cleaning solution.</li><li>• Solutions bead up, leaving particles behind.</li><li>• Contact Angle (CA) &gt; 90°</li><li>• Low surface energy</li><li>• Harder to clean, risk of contamination.</li><li>• Example Si without oxide.</li></ul>

## Explain RCA Clean step by step



### RCA Wafer Clean — step-by-step (to the point):

- Purpose: Industry-standard FEOL clean to remove organic residue, particles, metals and native oxide; protects device yield.

#### Original (basic) RCA sequence:

1. SC-1 (clean 1) — Mix  $H_2O : H_2O_2(30\%) : NH_4OH(29\%) = 5 : 1 : 1$ , heat  $70-80^\circ C$  (typ. 10–20 min).
  - Removes organics & particles by forming/dissolving hydroxide films.
  - Rinse thoroughly in ultrapure water.
2. SC-2 (clean 2) — Mix  $H_2O : H_2O_2(30\%) : HCl(37\%) = 6 : 1 : 1$ ,  $\sim 70^\circ C$  (typ. 10–15 min).
  - Removes metallic/alkali contaminants (e.g., Na, Fe, Cu).
  - Rinse + dry (ulpure water, SRD).

#### Common extended variant (pre-clean):

- SPM (piranha) —  $H_2SO_4(98\%) : H_2O_2(30\%) = 4 : 1$ ,  $\sim 130^\circ C$ , 10–15 min — bulk organic removal & surface oxidation. → Rinse.
- DHF dip —  $H_2O : HF(49\%) \approx 50 : 1$ , room temp, ~10 s — remove chemical oxide formed by SPM. → Rinse, then proceed SC-1 → SC-2.

Notes: every chemical step is followed by ultrapure-water rinsing and drying; SC2 may leave Cl residues if not rinsed.

## Explain IMEC Clean step by step



### IMEC Clean — step-by-step (to the point)

#### 1. Baseline (RCA) — starting point

- SPM ( $H_2SO_4 + H_2O_2$ ) → rinse → DHF (HF dip) → SC-1 ( $NH_4OH + H_2O_2$ ) → rinse → SC-2 ( $HCl + H_2O_2$ ) → megasonic rinse → Marangoni dry.
  - Removes organics, oxides, particles and metals.

#### 2. Dilute clean

- Same sequence but using dilute chemistries (lower concentrations).
- Cuts chemical consumption, cost and environmental impact while remaining effective.

#### 3. Reduced clean (IMEC)

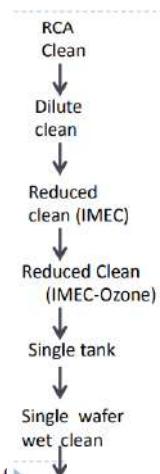
- Replace/modify aggressive steps:  $H_2SO_4 + O_3$  (ozonated sulfuric) used instead of hot SPM in a single-tank flow; follow with HF and dilute SC steps; rinse & dry.
- Fewer baths and lower chemical loads.

#### 4. Reduced clean (IMEC-Ozone / single-tank)

- Use  $H_2O + O_3$  (water/ozone) for organic removal and oxide regrowth, then HF/HCl as needed; supports single-tank single-wafer processing.

#### 5. Next-generation cleans

- Hybrid wet/dry cluster-tool approaches, single-wafer processing, and ozone-based chemistries for minimal footprint and emissions.



#### Why IMEC?

- Progressive reduction of chemical use, cost and environmental impact while maintaining particle/metal removal.

Contaminant classes targeted: particles, metals, organics, native oxide, micro-roughness.

✓ Write short notes on (i) Metallization (ii) Passivation.

### Metallization:

- Process of forming interconnects (wires) between devices on a chip.
- Conventionally uses aluminum due to good conductivity and ease of processing.
- Evaporation: High current heats Al wire in vacuum, vapor deposits on wafer.
- Sputtering: RF/DC plasma ions strike Al target, atoms dislodged and deposit on wafer.

### Passivation:

- Final protective glass layer over the IC surface.
- Prevents contaminants, moisture, and mechanical damage.
- Openings left for I/O pads and test probes.
- Enables further steps like bumping for direct circuit board connections.

✓ Write short notes on Diffusion and Ion implantation.

### Diffusion:

Process of moving impurity atoms through silicon crystal lattice. Used to form bases, emitters, collectors, resistors and sources/drain polysilicon doping.

- Performed in furnaces at  $1000\text{--}1200^\circ\text{C}$ .
- Dopant depth depends on time and temperature.
- Common dopants: Boron (P-type), Phosphorus, Arsenic (N-type).

### Ion Implantation:

Introduces dopants by accelerating ions into silicon with electric field. Depth depends on ion beam energy; dose controlled by beam current.

- Allows accurate and reproducible doping profiles.
- Can be done at room temperature, unlike diffusion.
- Widely used in precise doping in modern IC fabrication.

## Q: What are the merits and demerits of Ion Implantation?

Merits	Demerits
Precise control of dose and depth	Equipment is very expensive
Low process temperature	Causes lattice damage, needs annealing
Can form shallow junctions	Limited throughput (slow process)
Compatible with fine geometry devices	Channeling effects may occur

## Q: What are the merits and demerits of Diffusion?

Merits	Demerits
Simple, low-cost process	Poor control over junction depth
Suitable for large-scale doping	High process temperature required
High throughput, easy for batch processing	Lateral diffusion reduces resolution
No lattice damage introduced	Less suitable for modern small-geometry VLSI

## Q: Why is ion implantation preferred over diffusion for impurity doping?

Ans:

Ion implantation offers precise control over **dopant dose, depth, and location**, enables **low-temperature processing**, ensures **sharper junctions**, and provides **better uniformity and reproducibility** compared to diffusion.

## Mask Alignment and Its Importance in Multi-Layer Fabrication

- **Concept:** Mask alignment is the precise positioning of a photolithography mask over a wafer to define circuit patterns layer by layer.
- **Importance:**
  - Ensures different layers (metal, oxide, polysilicon, etc.) align correctly.
  - Prevents misalignment errors that cause short circuits or open connections.
  - Critical for device performance, yield, and reliability.
  - Modern VLSI requires sub-micron accuracy using automatic alignment systems.

## VLSI Physical Design Playlist



Clock To Play See the Playlist

## ▶ VLSI. Physical Design:

The process of converting circuit netlist into a final chip layout (placement, routing, timing verification) for fabrication, using CAD tools to handle design complexity is known as physical design.

Moore's Law: The empirical observation that the number of transistors on a integrated circuit doubles approximately every 18 months. Leading to exponential growth.

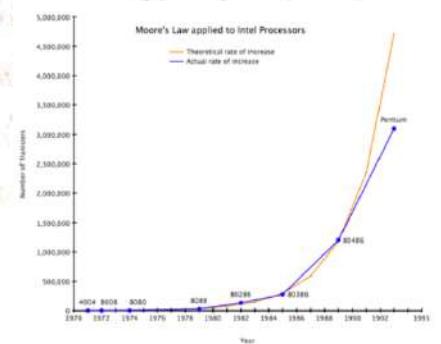
① Explain the importance of physical verification, also explain how it is achieved?

② Physical verification ensures the layout is manufacturable and functionally correct before tape-out. Prevents fabrication faults, rule violations, mismatches. Physical verification achieved by.

DRC: Design rule check: Ensures layout obeys boundary rules (spacing, width, overlaps, density) to guarantee manufacturability.

LVS: Layout Vs. Schematic: Verifies that the layout netlist matches the schematic netlist, ensuring correct connectivity and functionality.

RCX: Resistance-Capacitance Extraction/GRC: Extracts parasitic resistance and capacitance from the layout to generate an accurate post-layout netlist for timing and power analysis.



The three basic DRC checks

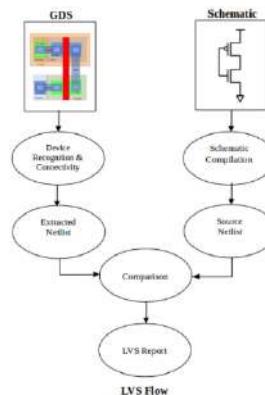
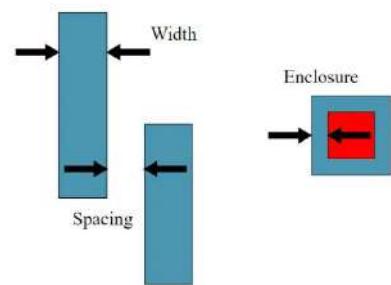
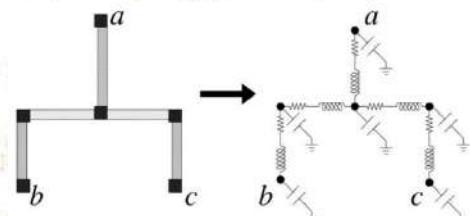


Fig6: LVS Flow



RCX Checks

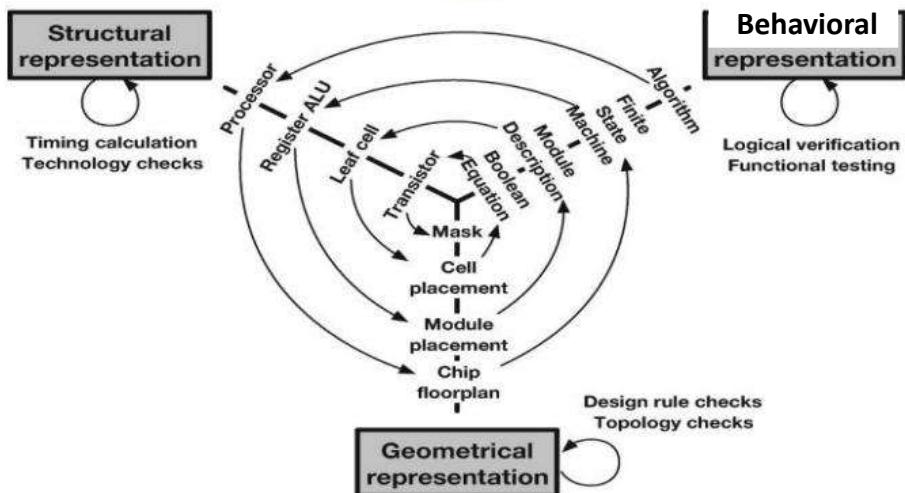
## D Design Representation:

✓ Explain simplistic view of design flow, briefly explain the behavioral, physical and structural domain of representation?

(i) The simplistic view of VLSI design can be represented using Y diagram.

- That shows three views (behavioral, structural, physical) converging.
- Design proceeds top-down (specification  $\rightarrow$  layout) and verified bottom up (layout  $\rightarrow$  function)
- Ensures smooth transition from abstract function to manufacturable chip.

Y Chart



Y-diagram domains:

(i) Behavioral Domain (what it does):

- Describes functionality: algorithms, Boolean equations, truth tables, FSMs.
- Top down: high level specification  $\rightarrow$  synthesis into logic.

(ii) Structural Domain (how it is built):

- Specifies inter connections of modules, gates, or transistors.
- Example: CPU made of ALU, registers, memory, a 4-bit adder built from full adders.
- Bottom up: leaf cells/modules combined into larger blocks.

(iii) Physical Domain (where it is implemented):

- Actual layout: geometrical placement, floorplanning, routing.
- Includes masks, polygons, and fabrication layers for manufacturing.

✓ VLSI Design Styles:

VLSI chips can be designed using different design styles depending on cost, performance, time-to-market and application. Each style offers a trade-off between flexibility, automation, performance and area optimization.

### Full Custom Design:

- Every transistor and wire is custom designed.
- Highest performance, lowest power and smallest area.
- very time consuming and expensive.
- Application: High volume, performance-critical ICs (CPUs, GPUs, DRAM)

### Standard Cell Design:

- Uses a library of pre-designed, pre-characterized logic cells.
- Good balance of performance, area, and design automation.
- Widely used in ASIC and SoC development.
- Application: Consumer electronics, networking chips.

### Gate Array (semi-custom):

- Wafer based layers pre-diffused, only metal layers customized.
- Faster turn around, less expensive than full-custom, but less efficient.
- Application: Medium-volume design, where quick delivery matters.

### FPGA (Field-programmable Gate Array):

- Completely programmable, reconfigurable even after manufacturing.
- High flexibility, fast prototyping, but poor power/efficiency.
- Application: prototyping, low/medium volume systems, adaptive or research projects.

✓ What do you understand by standard cell approach? Explain its cell characteristics and draw a multirow standard cell layout showing various wire routing.

⊖ Standard cell use a library of pre-defined cells (logic gates, FF) with fixed height and width. Cells abut side by side in rows; power rails run along rows, and row based routing is simplified.

### Characteristics:

1. Fixed height  $\rightarrow$  simplifies row placement and alignment.
2. Multiple drive strengths/variants (current vs speeds).
3. Fully characterized timing, power and layout views.
4. Pin accessibility constraints considered in cell layout.
5. DRCL design rule compliant, routable pin configuration.

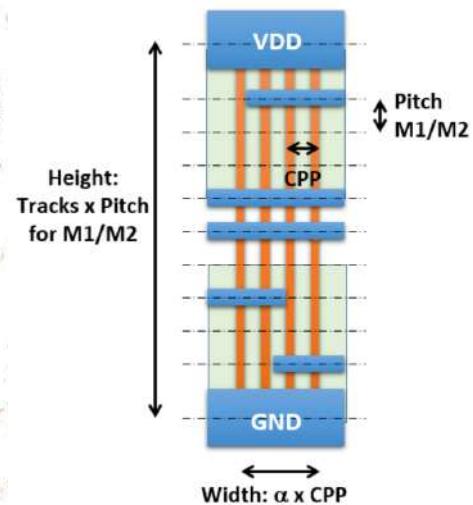


Fig : An Standard Cell with routing paths

Drawing: multi-row standard cell layout would show several rows, horizontal routing tracks, vertical intra-row local routing and power ground rails.

✓ Why Standard cell approach is suitable for design Automation.

⊖ Standard cell approach is suitable for design automation because

- Pre-designed, pre-characterized cells  $\rightarrow$  easy integration in CAD tools.
- Fixed height, consistent layout  $\rightarrow$  Simplifies placement and routing automation.
- Library cells come with timing, power and area models  $\rightarrow$  compatible with synthesis & STA.
- Supports design scalability and reuse across projects.
- Reduces design cycle time and human error.

① Briefly describe the following.

## ② Floorplanning:

- Floorplanning determines chip/corve shape, macro(large blocks) placement, I/O pad locations and blockages.
- It fixes the top-level outline and decides where standard-cell rows will go, how many to include and wiring channels.
- Good floorplan minimizes inter-connect length, congestion, timing path, IR drop.
- Poor floorplan leads to routing congestion, timing violations, difficult clock/power routing.
- It's the first physical step and determines later success of placement and routing.

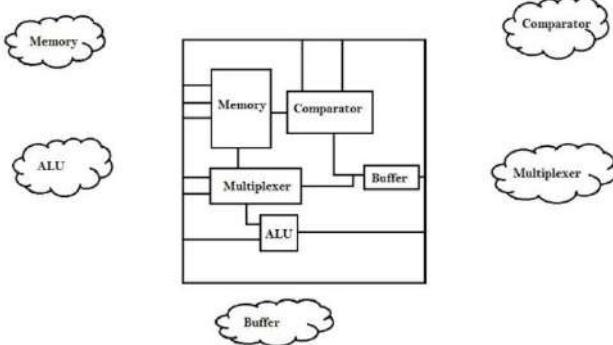
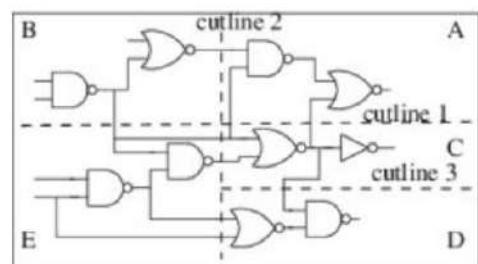


Fig : Floorplan

## ③ Partitioning:

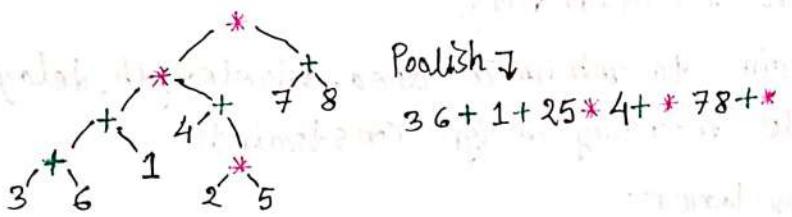
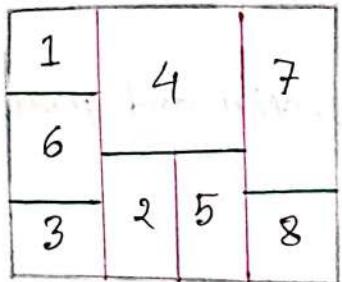
- Partitioning divides the design netlists into subblocks or modules that are manageable physically.
- It reduces complexity by localizing connectivity, improving locality.
- Helps in hierarchical layout, floor planning and reduces inter-block wire lengths.
- Partitions often aim to minimize cut size (nets crossing partitions) and balance area.
- Good partitioning improves placement, routing efficiency, and timing closure.



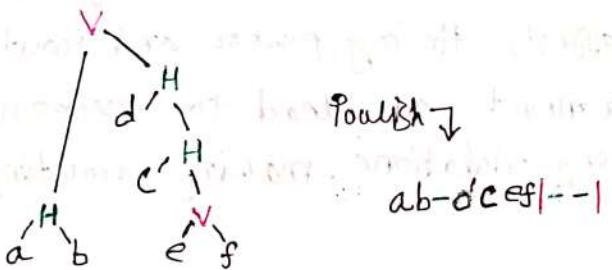
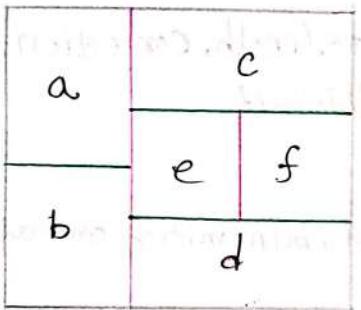
(a) Partitioning

- 田 An example of floorplan representation.

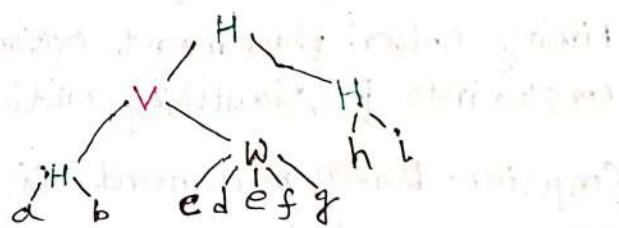
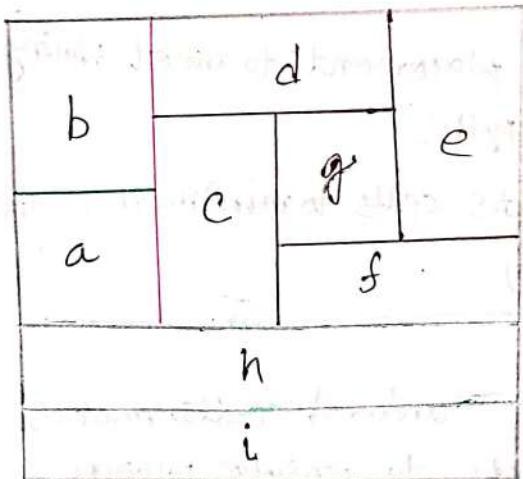
- 田 Horizontal cut '+' vertical cut '\*' (with horizontal and vertical lines)



田



- 田 A hierarchical Floorplan:



## III static Timing Analysis:

- STA Computes signal arrival times and required times without dynamic simulation. It finds critical paths, slack, setup/hold violations by analysing combinational delay + interconnect delay.
- It uses cell timing models and wire delay models.
- STA is crucial during placement/routing to ensure timing constraints are met.
- It guides buffers insertion, gate sizing, path optimization until timing closure.

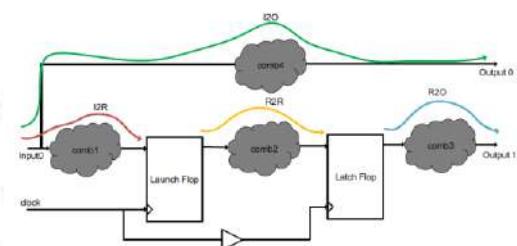


Fig : Static Timing Analysis

## IV Clock and power Routing:

- clock routing ensures the clock signal reaches all flip-flop with low skew and balanced delay.
- It inserts buffers, branch points and balances path lengths (H-shape, spine, grid).
- Power routing distributes VDD and GND across the chip via robust metal stripes to prevent voltage drop.
- Must consider IR drop, current density, electromigration, noise.
- Good clock/power routing is vital for performance, reliability and noise immunity.

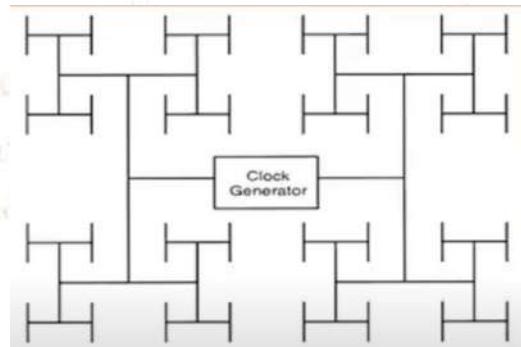
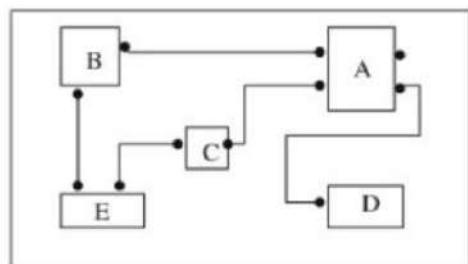


Fig : Clock Routing



(d) Detailed Routing

## D Placement

Placement is the process of assigning exact physical locations to standard cells, macros and blocks within the chip's core area after floorplanning.

Its aim is to minimize area, wirelength, delay, congestion, and power while meeting design constraints.

### Importance:

- Ensures efficient interconnection of modules.
- Directly affects timing, power and routability.
- Poor placement can lead to excessive wirelength, congestion, and timing violations, making routing difficult.

### Types:

- (i) Global placement: provides rough placement, minimizing overall wirelength and congestion.
- (ii) Detailed placement: refines placement by adjusting cells locally while ensuring no overlap.
- (iii) Timing-Driven placement: optimizes placement to meet timing constraints by prioritizing critical paths.
- (iv) Congestion-Driven placement: distributes cells to minimize routing congestion and improve routability.

## ► Routing:

Routing is the process of connecting standard cells, macros, and I/O pins with metal interconnects to ensure proper signal, clock and power flow across the chip. It ensures correct connectivity while meeting timing, power and area constraints.

### Types:

#### (i) Grid Routing:

- Early simple approach

- The layout area is divided into a grid, and wires must follow the grid lines.
- Ensures predictable routes but is less flexible and may waste area.

### (ii) Global Routing:

- Divides chip into coarse regions (routing bins)
- Decides approximate routing paths through these regions without assigning exact tracks.
- Focuses on congestion estimation, minimizing wire length and routing feasibility.

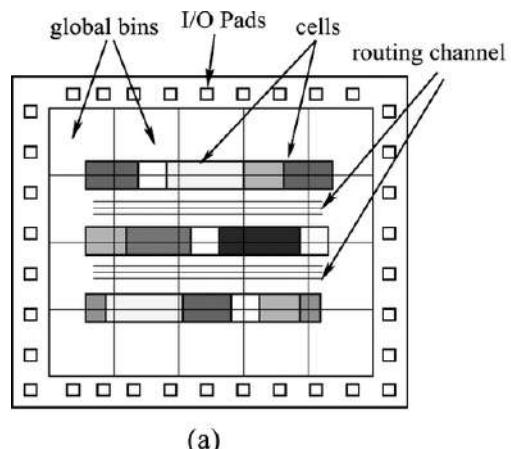
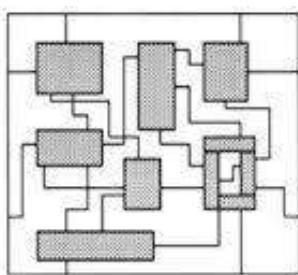
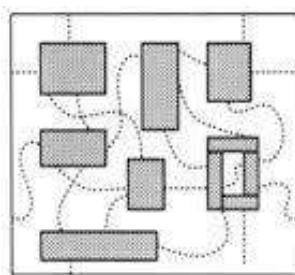


Fig : Grid Routing



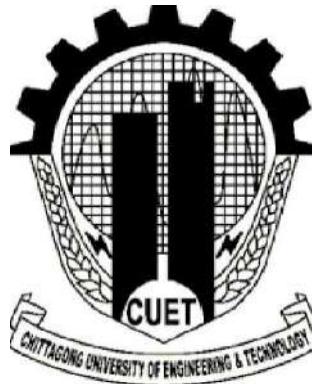
### (iii) Detailed Routing:

- Performs exact wiring on metal tracks and assign vias.
- Produces final mask-level layout, ensuring no overlaps or rule violations.
- Must pass DRC/LVS checks before fabrication.

**TAKE IT EASY  
PLEASE**



# Chittagong University of Engineering & Technology



Department of Electronics and Telecommunication Engineering

Assignment On

## The Clock Design in VLSI

Course No : ETE 403  
Course Title : VLSI Technology  
Date of Experiment : 06/10/2025  
Date of Submission : 12/10/2025

R E M A R K S

SUBMITTED BY

Name : Showkot Hosen  
Roll no : 2008010  
Level : L-4  
Term : T-I

## ■ Clocking in Digital Systems.

Clocking is a fundamental concept for synchronizing digital systems, especially sequential circuits, acting as a harmonizer for operations within a chip.

Most contemporary digital systems are synchronous, meaning a clock signal controls all internal operations, applied from an external source and synchronizing storage elements like flip-flops.

Faster clocks generally lead to faster operations, but various factors like delay, clock frequency, and power consumption must be considered.

Asynchronous systems, though potentially faster, are significantly harder to design and verify.

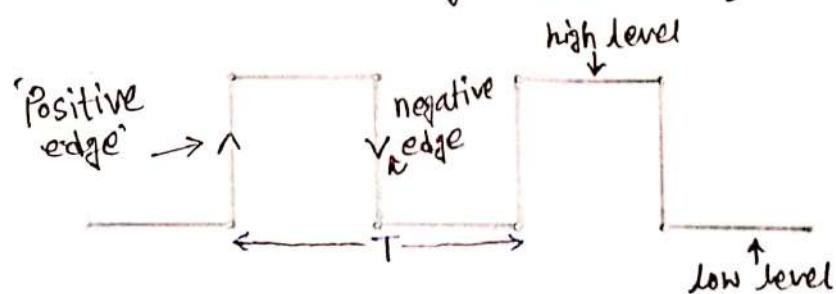


Fig.1. Key clocking parameters

### ■ Clock Parameters:

An ideal clock is periodic, fluctuating between high and low states over a defined period  $T$ .

Edge triggered flip-flops are activated on either the positive (0 to 1) or negative (1 to 0) edge of the clock signal, commonly used in most circuits today.

Latches differ from flip-flops as their storage is controlled by the clock's level rather than its edge; they remain open and store input as long as the clock is high.

## ■ Pipelining in Digital Systems:

- Breaks computation into smaller stages separated by registers.

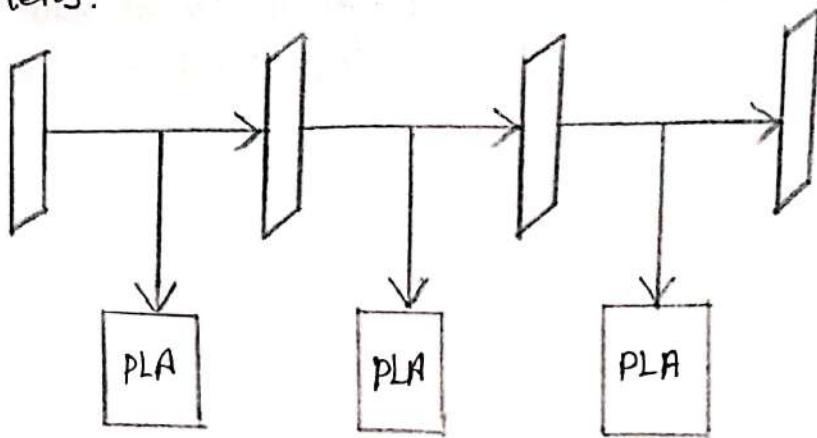


Fig 2. Pipelining.

- Multiple data items processed simultaneously  $\rightarrow$  higher throughput
- A K stage pipeline gives  $\sim K \times$  Speedup (ignoring overhead)

## CPU performance factors:

$$\bullet \text{Execution time} = IC \times CPI \times CCT$$

• Improve performance by reducing  
Instruction count, CPI or clock cycle Time.

Where,

IC = Instruction count.

CPI = clocks per Instruction.

CCT = clock cycle time

## Timing constraints:

- Setup time ( $t_{\text{setup}}$ ): Data stable before clock edge.
- Hold time ( $t_{\text{hold}}$ ): Data stable after clock edge.

## Pipeline timing equations:

- $T \geq t_{\text{logic\_max}} + t_{\text{setup}} + t_{\text{skew\_max}} + t_{\text{written\_max}}$
- Ensures correct data capture across stages.

### ④ clock skew

Difference in clock arrival times at different flip-flops due to unequal path delays, causes timing violations. (setup / hold).

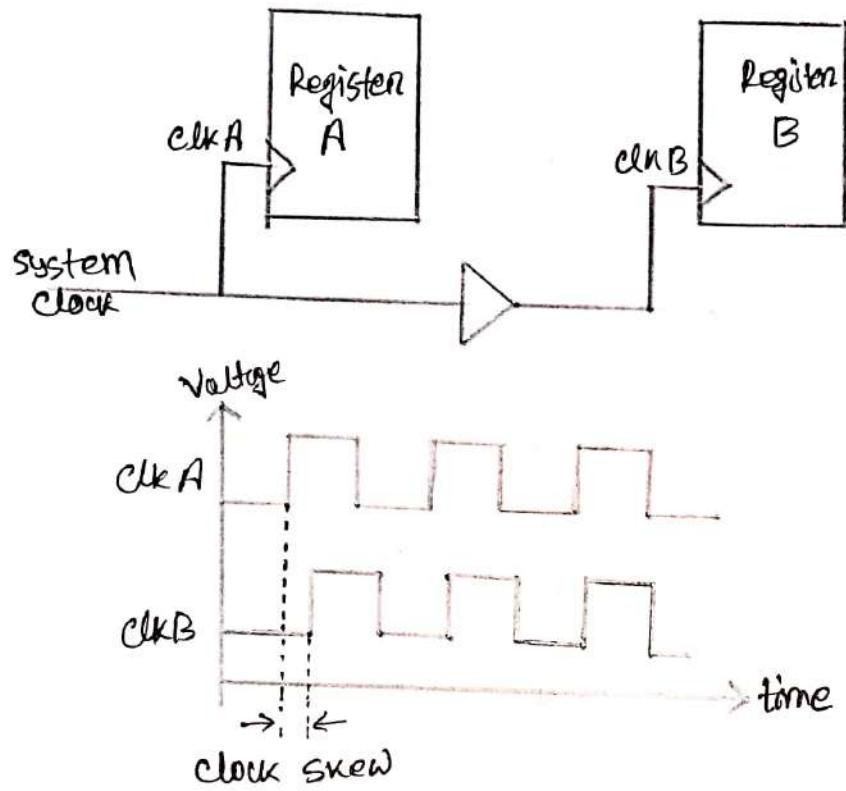


Fig3: clock skew in Digital circuits.

### ⑤ Clock jitter:

Variations in arrival time successive clock edges at the same flip-flop, leading to irregular clock period.

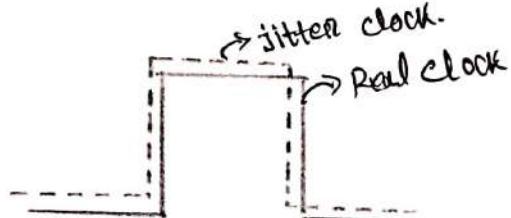


Fig4: clock jitter.

## ■ Key timing parameters

- Clock width ( $t_w$ ): Minimum high duration for flip-flop edge detection.
- Setup ( $t_{sv}$ ): Data stable before clock edge.
- Hold ( $t_H$ ): Data stable after clock edge.
- Propagation delay ( $t_p$ ): Time for output change after clock edge.

## ■ Impact of clock skew on cycle time

- Positive skew ( $c_2$  delayed): Helps timing, reduces cycle time requirement.  
 $T_{clk} \geq T_{ff\_prop} + T_{logic} + T_{setup} - T_{skew} \rightarrow$  higher max frequency.
- Negative skew ( $c_1$  delayed): Hurts timing, increase cycle time:  
 $T_{clk} \geq T_{ff\_prop} + T_{logic} + T_{setup} + T_{skew} \rightarrow$  lowers max frequency.

## Skew tolerance examples:

- Direct FF-FF: To avoid hold violation,  $T_{skew} \leq T_{ff\_prop} - T_{hold}$ .
- With mux: Skew tolerance becomes  $T_{skew} \leq T_{ff\_prop} + T_{mux\ setup} - T_{hold}$
- Equation guide safe skew limits for timing closure.

## ④ Clock Loading Control.

Direct gating of the clock is unsafe, can cause incorrect pulses  $\rightarrow$  extra clock edges  $\rightarrow$  wrong operation.

Alternative exists, such as :

- Restricting load activation to only when the clock is low
- Using an OR gate with an inverted load signal.

However, manipulating the clock directly is discouraged due to risk of clock skew and unpredictable glitches.

## ⑤ Clock design metrics.

- Power Requirement: Clock can consume  $\sim 40\%$  of chip power.
- Skew: Violation in clock arrival times at flip-flops.
- Jitter: Short-term period variations.
- Flexibility: Ease of adding/modifying loads.
- Reliability: Avoid failures
- Trade off depends on application (power vs. performance)

## ⑥ Skew minimization:

- Balance buffers and wiring.
- Symmetric design.
- Minimize process/temperature variation effects.
- Use PLLs or feedback for correction.
- Apply timing-aware synthesis.

## ■ Designing a clock for VLSI circuit:

■ An ideal clock should distribute signals with minimum skew and jitter, while consuming low power and maintaining reliability across process, voltage and temperature variations.

The first step in clock design is to define the frequency, duty cycle and voltage levels of the clock signal.

Once the specifications are fixed, a clock distribution network or clock tree is implemented. This network routes the clock signal from a single source (oscillator or PLL) to all sequence elements. To ensure uniform arrival times, buffers and inverters are added at strategic locations. This method is called clock tree synthesis (CTS).

Since clock lines switch every cycle, they consume a large amount of dynamic power (up to 40% of chip power). To reduce this, techniques like clock gating are used, where unused portions are disabled temporarily from the chip. Additionally phase locked loops (PLL) and delay locked loops (DLL) are often employed to reduce jitter and automatically correct skew.

In practice, designers must check for timing constraints such as setup and hold times, propagation delays and skew tolerance. The final design is verified through static timing analysis (STA) to ensure reliable performance.

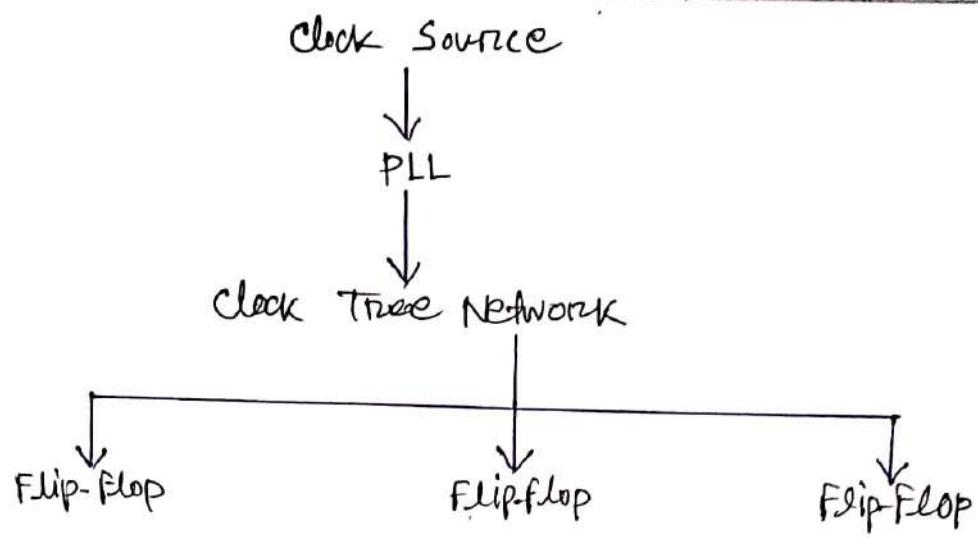


Fig:5: Clock design for a VLSI design.

thus clock design in VLSI is a trade-off between power, performance and reliability, and its optimization directly impacts the maximum operating frequency of the chip.

“ We will show them our signs in the universe and in their own selves, until it becomes manifest to them that this ( the Quran) is the truth.”

[Surah Fussilat, Verse 41 : 53](#)

**The end**

