NEHRU COLLEGE OF ENGINEERING AND RESEARCH CENTRE
(NAAC Accredited)
(Approved by AICTE, Affiliated to KTU University, Kerala)

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

Course Material

S6:EC304: VLSI

About the Department:

Department of ECE established in 2002 with an intake of 60 students to undergraduate (B.Tech) programme and enhanced to an intake of 120 students from 2006. The department offers two Postgraduates(M.Tech) programmes in "Electronics". "Applied Electronics & Communication System" from 2011 with an intake of 18 students and " VLSI Design" from 2012 with an intake of 18.Highly qualified, experienced and dedicated staff members are the backbone of the Department. The Department always strive hard to satisfy the knowledge thirst of both students and faculties by organizing workshops / technical talks / conferences etc. The faculty members are actively involved in research work and regularly present/ publish their work in various national and international conferences / journals. The ECE Department is proud to host state-of- the art Laboratories in the area of VLSI, Embedded Systems, Microprocessor and Microcontrollers, Circuits, Analog and Digital Communication and Microwave and Optical communication.The ECE department formally inaugurated the ECHOS (The ECE Association) in 2009 and under this banner many extra-academic activities have been conducted such as paper presentation, quiz competition, workshops and seminars. Also the department has two magazines that have been developed on the basis of the creative skills of our imaginative students. There is an Embedded Club that meets on monthly basis to discuss innovative projects and publication based activities. Department is closely associated with INSTITUTE OF ELECTRONICS & TELECOMMUNICATION ENGINEERS (IETE) Palakkad Centre to organize technical events like guest lecture, seminars and conferences.

Vision of the institute:

To mould true citizens who are millennium leaders and catalysts of change through excellence in education.

Mission of the institute:

NCERC is committed to transform itself into a center of excellence in Learning and Research in Engineering and Frontier Technology and to impart quality education to mould technically competent citizens with moral integrity, social commitment and ethical values. We intend to facilitate our students to assimilate the latest technological know-how and to imbibe discipline, culture and spiritually, and to mould them in to technological giants, dedicated research scientists and intellectual leaders of the country who can spread the beams of light and happiness among the poor and the underprivileged.

Vision of the department:

Providing Universal Communicative Electronics Engineers with corporate and social relevance towards sustainable developments through quality education.

Mission of the department:

1) Imparting Quality education by providing excellent teaching, learning environment.

2) Transforming and adopting students in this knowledgeable era, where the electronic gadgets (things) are getting obsolete in short span.

3) To initiate multi-disciplinary activities to students at earliest and apply in their respective fields of interest later.

4) Promoting leading edge Research & Development through collaboration with academia & industry.

Program Educational Objectives (PEOs)

I. To prepare students to excel in postgraduate programmes or to succeed in industry / technical profession through global, rigorous education and prepare the students to practice and innovate recent fields in the specified program/ industry environment.

II. To provide students with a solid foundation in mathematical, Scientific and engineering fundamentals required to solve engineering problems and to have strong practical knowledge required to design and test the system.

III. To train students with good scientific and engineering breadth so as to comprehend, analyze, design, and create novel products and solutions for the real life problems.

IV. To provide student with an academic environment aware of excellence, effective communication skills, leadership, multidisciplinary approach, written ethical codes and the life-long learning needed for a successful professional career.

Program Outcomes (Pos):

1. **Engineering Knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. **Problem Analysis**: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. **Design/development of Solutions**: Design solutions for complex engineering problems and design system components or processes that meet t h e specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. **Conduct Investigations of Complex Problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. **Modern Tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. **The Engineer and Society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. **Environment and Sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. **Individual and Team Work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. **Project Management and Finance**: Demonstrate knowledge and understanding of t h e engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. **Life-long Learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes (PSO):

1. Facility to apply the concepts of Electronics, Communications, Signal processing, VLSI, Control systems etc., in the design and implementation of engineering systems.

2. Facility to solve complex Electronics and communication Engineering problems, using latest hardware and software tools, either independently or in team.

Mapping of PEOs with the Program Outcomes (POs):

The Electronics and Communication Engineering Program outcomes leading to the achievement of the objectives can be summarized in the following Table.

		Program Outcomes										
		a	b	c	d	e	f	g	h	i	j	k
PEOs	1	X	X	X								X
	2	X	X	X	X		X					X
	3		X	X	X	X					X	
	4				X	X	X	X	X	X	X	X

Course Outcome:

After completion of this course the students will be able to

01. Understand and describe material preparation, diffusion, oxidation and ion implantation.
02. Knowing the epitaxial and lithography process in the IC Fabrication with the isolation methods.
03. Describe, analyze, formulate and construct CMOS Inverters with layout design rules
04. Understand and explain MOSFET logic design in the bank end process
05. Demonstrate basics of Different Type of Memory and Sensing Circuits.
06. Demonstrate the function of Adders and Multiplier techniques

CO-PO Mapping

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	1							1	2	
CO2		3			2					1	2	
CO3		3	2		1					1	2	
CO4	2	3								1	2	
CO5		3			2					1	2	
CO6		2	2		3					1	2	

COURSE PLAN:

1 Introduction to VLSI

2 Material Preparation - Production of EGS

3 Crystal Growth - CZ Process and FZ Process

4 Wafer Preparation

5 Thermal Oxidation - Growth Mechanism

6 Dry - Wet Oxidation - Deal & Grove Model

7 Diffusion - Ficks laws - Diffusion with constant surface

8 Ion Implantation - Range Theory - Annealing

9 Epitaxy - VPE & MBE

10 Lithography - Photo lithography - Steps

11 Electron beam lithography - Etching & Metal deposition

12 Component fabrication - Transistor

13 Fabrication - Diode - Resistor - Capacitor

14 N Well CMOS IC fabrication

15 CMOS Invertors - DC Characteristics

16 Switching Chara of CMOS Invertor

17 Power dissipation of CMOS Invertor

18 Layout Design rules - Stick diagram

19 Layout of CMOS Invertor

20 Layout of 2 input NAND & NOR gates

21 Pass transistor Logic

22 Complementary Pass Transistor Logic

23 Transmission gate logic

24 Realization of functions

25 Realization of functions

26 4 x 4 MOS ROM Cell array(OR - NOR - NAND)

27 SRAM Six transistor

28 DRAM - Three transistor & One Transistor Dynamic Memory Cell

29 Sense Amplifiers - Inroduction

30 Differential Voltage Sensing Amplifiers

31 Introduction to PLDs & FPGAs

32 Design of PLAs

33 Adders - Static adder - Carry By pass adder

34 Linear carry select adder

35 Square root carry select adder

36 Multiplier - Array multiplier

COURSE CODE	COURSE NAME	L-T-P-C	YEAR OF INTRODUCTION
EC304	VLSI	3-0-0-3	2016

Prerequisite: EC203 Solid State Devices, EC204 Analog Integrated Circuit.

Course objectives:

- To give the knowledge about IC Fabrication Techniques
- To impart the skill of analysis and design of MOSFET and CMOS logic circuits.

Syllabus:

IC Fabrication Technology**,** CMOS IC Fabrication Sequence, CMOS inverters, Design rules, Static CMOS Design, Dynamic CMOS circuits, Pass transistor, Read Only Memory, Random Access Memory, Sense amplifiers, Adders, multipliers, Testing of VLSI circuits.

Expected outcome:

The students will be able to design and analyse various MOSFET and CMOS logic circuits.

Text Books:

1. John P Uyemura, Introduction to VLSI Circuits and Systems, Wiley India, 2006
2. S.M. SZE, VLSI Technology, 2/e, Indian Edition, McGraw-Hill,2003

References:

1. Jan M.Rabaey, Digital Integrated Circuits- A Design Perspective, Prentice Hall, Second Edition, 2005.
2. Neil H.E. Weste, Kamran Eshraghian, Principles of CMOS VLSI Design- A Systems Perspective, Second Edition. Pearson Publication, 2005
3. Razavi - Design of Analog CMOS Integrated Circuits,1e, McGraw Hill Education India Education, New Delhi, 2003.
4. Sung –Mo Kang & Yusuf Leblebici, CMOS Digital Integrated Circuits- Analysis & Design, McGraw-Hill, Third Ed., 2003.
5. Yuan Taur & Ning, Fundamentals of Modern VLSI Devices, Cambridge University Press, 2008

Course Plan

Module	Course content	Hours	End Sem. Exam Marks
I	**Material Preparation-** Purification, Crystal growth (CZ and FZ process), wafer preparation **Thermal Oxidation-** Growth mechanisms, Dry and Wet oxidation, Deal Grove model.	4	15
	Diffusion- Fick‘s Laws, Diffusion with constant surface concentration and from a constant source, diffusion techniques. **Ion implantation**-Technique, Range Theory, annealing.	3	
II	**Epitaxy :** Vapour phase epitaxy and molecular beam epitaxy **Lithography-** Photo lithographic sequence, Electron Beam Lithography, Etching and metal deposition	4	15
	Methods of isolation Circuit component fabrication: transistor, diodes, resistors, capacitors, N-well CMOS IC Fabrication Sequence	3	
FIRST INTERNAL EXAM			
III	**CMOS inverters-** DC characteristics, switching characteristics, power dissipation	4	15

	Layout Design rules, Stick Diagram and layout of CMOS Inverter, two input NAND and NOR gates	4	
IV	**MOSFET Logic Design** -Pass transistor logic, Complementary pass transistor logic and transmission gate logic , realization of functions	6	**15**
SECOND INTERNAL EXAM			
V	**Read Only Memory**-4x4 MOS ROM Cell Arrays(OR,NOR,NAND) **Random Access Memory** –SRAM-Six transistor CMOS SRAM cell, DRAM –Three transistor and One transistor Dynamic Memory Cell	4	**20**
	Sense amplifiers –Differential Voltage Sensing Amplifiers Introduction to PLDs and FPGAs, Design of PLAs.	3	
VI	**Adders**- Static adder, Carry-By pass adder, Linear Carry-Select adder, Square- root carry- select adder **Multipliers**-Array multiplier	4	**20**
END SEMESTER EXAM			

Question Paper Pattern (End Semester Exam)

Maximum Marks : 100 **Time : 3 hours**

The question paper shall consist of three parts. Part A covers modules I and II, Part B covers modules III and IV, and Part C covers modules V and VI. Each part has three questions uniformly covering the two modules and each question can have maximum four subdivisions. In each part, any two questions are to be answered. Mark patterns are as per the syllabus with 70% for theory and 30% for logical/numerical problems, derivation and proof.

MODULE 1

I. What is the need of SiOi layer in MOS fabrication process?
2. State the laws governing the diffusion process.
3. Compare wet and dry oxidation.
4. Write down the range equation for ion implantation and explain each term in it.
s. What are oxides related capacitance and junction capacitance?
6. Explain float zone process of crystal growth. What are its advantages and disadvantages over Czochralski growth?
7. What are the different processes involved in silicon wafer preparation?
8. Explain Czochralski technique of crystal growth.
9. Explain the slicing and polishing of silicon wafers and the purpose of notch on the wafer.
10. Explain the essential steps in IC fabrication.
i 1. Derive the Deal-Grove model of oxidation.
12. With neat sketch, explain ion implantation process.
13. Draw the limited source and constant source diffusion profiles and distinguish them in terms of relevant modelling equations.
14. Explain the oxidation growth mechanism with a neat diagram.
15. Explain the CZ and FZ process for crystal growth and compare them.
16. What is oxidation induced stacking fault and how it can be eliminated?
17. Write the difference between pre-deposition and drive-in processes.
18. What are the different diffusion mechanisms, explain?
19. Explain the different types of ion stopping mechanisms.
20. What is called annealing and how it differs from rapid thermal **annealing.**
21. **Write** the **equation of** projected range and define each term in it.

MODULE 2

1. Mention the types of resistor fabricated in IC environment.
2. What is lithography?
3. With the help of neat diagrams, explain the steps involved in i) Photolithography ii)X-ray lithography.
4. List the n well IC fabrication sequence.
s. Comment briefly on the characteristics of an exposure tool used for lithographic process.
6. What are the properties of metals used for metallization?
7. What are the undesirable capacitances formed in MOS fabrication?
8. Assuming Gaussian distribution for the ion implantation, find the distance from the surface at which the ion concentration falls to half the peak value, for arrange of 0.1q.The straggle value is 0.02s.
9. Explain how resistors and capacitors are fabricated on an IC chip.
10. What are positive and negative photoreists, explain.
i1. What are the important criterions for hetero epitaxy of material A to be possible on a single crystal substrate of material B?
12. What is boundary layer problem? How can it be minimised in a horizontal epitaxy system?
13. Explain any one method of epitaxial growth, with diagram.
14. With diagrams, explain the steps involved in the fabrication of a n-well CMOS IC.
15. What is electron beam lithography?
16. Explain the different types of etching techniques.
17. Explain CVD system for epitaxial growth.
i8. Explain the molecular beam epitaxy, with neat diagrams.
19. Explain the different types of metal deposition techniques.
20. What is the difference between diffusion and epitaxy7

MODULE 3

1. Compare the features of CMOS and bipolar technologies.
2. Define threshold voltage for a MOS transistor. What are the parameters on which it depends on?
3. PMOS is perfect switches for transmission of logic 1 while NMOS is perfects for transmission of logic 0.Justify the statement.
4. Draw the structure of CMOS p-well inverter.
5. Why NMOS technology is preferred more than PMOS technology?
6. **Draw** the Stick diagram, circuit diagram, of i)NMOS inverter ii)CMOS inverter
7. Realize the equation Y=(AB+CD)' in i)NMOS technology ii)CMOS technology
8. Draw the stick diagram of the function, F=(AB+E+CD)'
9. Define the noise margin for CMOS inverter.
10. Derive expression for switching threshold of a CMOS inverter.
11. Derive graphically, the CMOS inverter characteristics.
12. What is meant by charge leakage in dynamic CMOS logic? Explain a method to prevent it.
13. Explain static power dissipation and short circuit power dissipation with reference to CMOS logic.
14. What are the different regions of operation of MOS transistor? How are they related to Vgs, Vds and Vt for nmos and pmos transistors?
15. Calculate the dynamic power dissipation in a chip operating with Vdd of SV at 100MHz with an internal switched capacitance of 300pF.
16. Explain the CMOS inverter DC characteristics with the region of operation in detail.
17. Draw the circuit diagram and stick diagram for implementing the logic function F=A. (B+C)' in static p well CMOS logic.
18. Draw the stick diagram and layout of 2 input NAND gate.
19. Draw the stick diagram and layout of 2 inputs NOR gate.
20. Derive the expression for dynamic power dissipation in a static CMOS inverter.
21. For a CMOS circuit C_L =15fF/gate, ADD =2•5V Rd tp'35ps.If there are lo' gates/chip, determine the dynamic power dissipation at the maximum frequency of operation.

MODULE 4

1. What is meant by charge leakage in dynamic logic? How it is prevented?
2. DraW the layout of a 4 input AND/NAND gate in complementary pass transistor logic.
3. Draw the circuit schematic of 2 input multiplexer using transmission gates.
4. What is np-domino logic?
5. How is charge leakage and sharing problems rectified in dynamic logic?
6. Write down the basic principle of dynamic CMOS logic with circuit diagram.
7. Draw the circuit diagram of transmission gate XOR logic and explain its operation.
8. Draw the circuit of a 4 input NOR gate in domino logic. What are the advantages of domino logic?
9. Design pass transistor logic for 2 input XNOR gate.
10. Implement a 2x1 mux using transmission gates.
11. Implement a 4x1 mux using transmission gates.
12. Draw a 2 input XNOR using transmission gates.
13. Explain the operation of NMOS and PMOS transistor operation in transmission gates.
14. Explain the function of complementary pass transistor.
i5. Draw 2 input NAND gate in static CMOS.
16. Implement the function F—AB(C+D)+DE using static CMOS technology.
17. Compare pass transistor logic and complementary pass transistor logic.
18. Draw a 2 input XOR using transmission gates.
19. What are the advantages and disadvantages of domino logic?
20. Design a multiplexer using CMOs pass transistor logic and explain its operation with its truth table.
21. Give the different symbols for transmission gates.

MODULE 5

i. With the help of a circuit diagram give proper explanation for the **confi guration of 4x4** NOR-Rom array.

2. Write short note on FPGA.

3. Draw the circuit diagram of s sense amplifier and give explanation on how the sensing of operation is carried out in SRAM cell.

4. Draw a 6T SRAM cell.

3. What is the function of CLB in FPGA?

6. Design an AND-OR PLA with putputs,Fl=m0+m2+m6, P2=m0+m5+m6, F3= m3+m4+m7.

7. Draw and explain the schematic and physical structure of a DRAM cell using a trench capacitor

8. Design an FET programmable ROM with the following data.

Address	0	1	2	3	4	5	6	7
Data	0100	1111	1010	0001	1011	0111	1110	1001

9. What are the functions of sense amplifier?

10. Draw the circuit diagram of a 3 transistor dynamic RAM cell and explain its operation.

11. Show the CMOS implementation of a OR-ROM cell to store 8 words of 8 bits.

12. With cross sectional view, explain the principle of an EPROM cell.

13. Show the CMOS implementation of a NOR ROM cell to store 4 words of 4 bits which are as follows 1011,0110,1001 and 1100.

14. Draw a Master slave MUX based latch pair using transmission gates and inverters and explain its operation.

15. Design a PLA for realizing the following outputs.
f l'Z ,2,6,7, f2 =Z3,5,7,9 , f 3=l,3,8,l2

16. Draw and explain multiplexer based positive edge- triggered register using master slave configuration.

17. What are the advantage and disadvantages of dynamic latches and registers over static implementations?

18. Compare dynamic RAM and static RAM.

19. Using a suitable PLA, design and implement a 4 bit binary to gray code.

20. What is the purpose of sense amplifier? Describe the working of a single ended sense amplifier.

21. Draw a DRAM memory cell.

22. Draw a ROM array to store a set of eight 8 bit data using MOS based ROM. Explain how they are written and read.

23. Design an AND-OR PLA with outputs Fl=ml +m6, F2=m0+m5+m6+in7 F3=m3+m4+m7

MODULE 6

1. lmplement a linear 16 blt carry select adder using 4 bit carry Select adder blocks. How does a square root carry select improve the performance?
2. Discuss the principle of Wallace Tree Multiplier. How is the propagation delay reduced as compared to simple array multlplier?
3. Design a 14 bit square root carry select adder. Calculate the worst case delay.
4. Show the design of a 32 bit carry select adder based on 4 bit or similar ripple carry adder. Find the total gate delay for your design.
5. Explain the operation of carry bypass adder.
6. Draw the logic circuit for a 4 bit carry look ahead adder.
7. De5ign a 4x4 array multiplier.
8. Design an 8 bit carry select adder.
9. With the help of suitable diagrams, explain how delay computation time is reduced in carry look ahead adder.
10. Explain the partial product accumulation in a Wallace tree multiplier.
11. Explain the operation of NxM bit array multiplier.
12. Show the critical path in a linear carry select adder scheme and give the expression for critical path delay involved.
13. Explain the principle of a register based multiplier.
14. How is booth encoding used to speed up multiplication process?
15. How are carry look ahead adder module interconnected to get higher order adders?
16. Explain the working of carry bypass adder.
17. Explain square root carry select adder and derive the expression for time delay and **compare** its performance with other adders.
18. With **diagram** explain the implementation of 4 bit carry look ahead **generator iR dynarn»** CMOS logic.
19. **Compare** the **performance** of linear carry select adder **and square root carry select adds.**
20. **Explain the working of** static adders.

Q) Why silicon is used over Germanium in chip?

A) It is cheaper and SiO_2 is more stable than GeO_2.
High intrinsic impedance.

- SSI - 1 to 100 transistors

→ MSI - 100 to 1000

→ LSI - 1000 to 10,000

→ VLSI - 10,000 to 1 million

→ ULSI - 1M to 10M

$$SiO_2 + C \xrightarrow[1600^\circ C]{\text{High temp.}} SiC$$

$$SiC + SiO_2 \longrightarrow \underset{(liq)}{Si} + SiO + CO$$

It has high impurity (MGS)

∴ It is converted into Electronic grade silicon (EGS)

Impurity level in MGS and EGS

	MGS (ppm)	EGS (ppb)
B	35 - 50	≤ 0.01
Cr	50 - 200	≤ 0.01
Cu	15 - 60	≤ 0.1
Ni	20 - 100	0.1 - 0.5
P	20 - 50	≤ 0.5

(i) material preparation
(ii) Si oxidation
(iii) Impurity Diffusion
(iv) Ion implantation
(v) Lithography
(vi) Etching
(vii) Thin film deposition
(viii) Epitaxial growth
(ix) metalisation

(i) material preparation

It involves 3 steps, they are:

- purification - Natural sand into EGS
- crystal growth - Creation of ignots using Cz or Fz process.
- wafer preparation - Ignot into wafer

Purification

Si occuring naturally in the form of silica and silicates is the most important semiconductor for electronic Industry. when compared to Ge, Si excels for following reasons,

→ Si devices can operate at high temperature.
→ Intrinsic resistivity is higher.
→ SiO_2 is more stable than GeO_2.
→ It is also water soluble.
→ Si is available in low cost

Goal of material preparation is preparing high impurity Si ~~crystal~~ wafer. Si wafer refers to a single crystal of Si with a specific orientation, resistivity and dopant concentration. Starting material of Si wafer manufacture is electronic grade silicon. EGS prepared from MGS (Metallurgical Grade silicon). MGS is synthesized from the ore (sand or quartzite). The ore is reduced to Si by mixing with excess of carbon (coal, charwood chips) and heating in submerged electrode arc furnace. The SiO_2 react with excess carbon to first form silicon carbide (SiC) at high temperature SiC react with SiO_2 to form Si.

$SiC(s) + SiO_2(s) \longrightarrow Si(l) + SiO(g) + CO(g)$

This is MGS and it is 90% pure.

To convert MGS to EGS

1) Pyrolysis Method

SiH_4 + Heat → $Si + 2H_2$

SiH_4 (silane)

EGS can be produced by pyrolysis method in which silane (SiH_4) will be reacted with heat. The reaction takes place at high temperature of 900°C. The main advantage of using silane instead of Trichlorosilane is lower production cost and less production of harmful byproduct.

2) Siemann's process

CVD ⇒ Chemical vapour deposition

SiH_4 + Heat → ← HCl

Si (MGS) → Hydrochlorination → $SiCl_4$

↓

Distillation

↓

$SiHCl_3$ → Vaporization and CVD of $SiHC_3$ ←

↓

Chlorosilane recovery → $SiCl_4$

↓ H_2

Hydrogen and HCl recovery

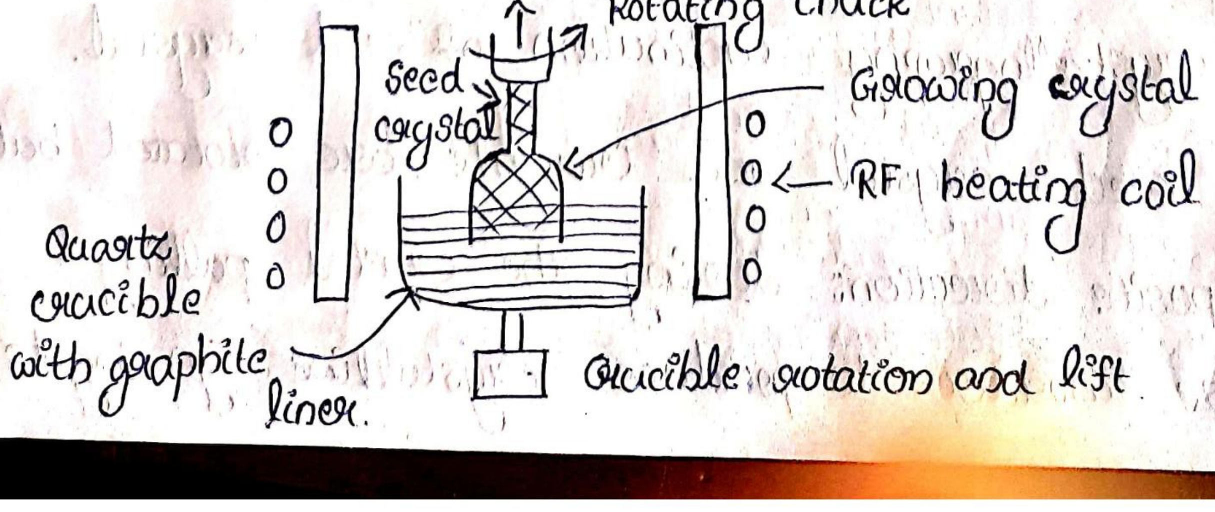
(MGS)

$$2SiHCl_3 + 2H_2 \longrightarrow 2Si_{(EGS)} + 6HCl$$

MGS has to be pulvarised mechanically and reacted with anhydrous hydrogenchloride to form Trichlorosilane. With the help of catalyst the reaction takes place at a nominal temp. of 300°C. The reaction leads products like $SiCl_4$ and chlorides of impurities. At this point purification process occurs. It has to be done by fractional distillation as the products trichlorosilane and unwanted chlorides are liquids at room temp. Purified $SiHCl_3$ is subjected to chemical vapour deposition then it is reduced with hydrogen results electronic grade silicon and recovery of HCl.

crystal growth (creation of ignots)

(i) CZ (czochraolski) process



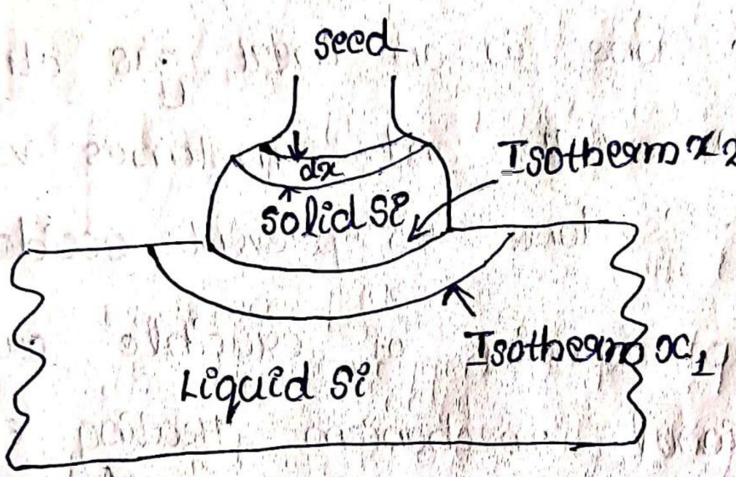
crystal pulling mechanism, ... control system circuitary.

EGS is a polycrystalline structure. Cz process is used to convert polycrystalline silicon into single crystal silicon ingot. The material is then heated to a temperature 1500°C ie, slightly inexcess of silicon melting point 1420°C. A small single crystal rod of silicon called seed crystal is then dipped into the silicon melt. The conduction of heat the seed crystal will produce a reduction in temp of the melted incontact with seed crystal to slightly below Si melting point. The Si will freeze onto the end of seed crystal and end of seed crystal is slowly pulledup of the melt it will pullup solidified mass of Si that will be a crystallography continution of seed crystal. Both seed crystal and crucible are rotated but in opposite directions during the crystal pulling process inorder to produce crystalline sgn ingots remains near to the surface of the ... temperature and pulling rate are correctly chosen. The diameter of ingot is controlled by pulling rate and melt temp with inward diameter of about 100-150mm (4-6 inches). The ingot length will generally be the order of 3m. The crystal pulling is done in an inert gas atmosphere usually Ar or He and sometimes vaccum is used. This is done to prevent oxidation.

The furnace consist of crucible, crucible support and rotational mechanism, Heating element and power supply and a chamber. The crucible material should be chemically unreactive with molten Si, also the material should have high melting point, thermal stability and hardness. The material used for crucible is silicon nitride and fused silica. Graphite is used as crucible support as it has high temperature properties.

The crystal pulling mechanism consists of sealed shaft or chain, rotational mechanism and of digrowth process - pull rate and crystal rotation. The pulling mechanism must have minimum vibration and great precision.

Relationship b/w pull rate and crystal diameter



$$V_p \propto \frac{1}{\sqrt{r}}$$

$$L \frac{dm}{dt} + K_L \frac{dT}{dx_1} A_1 = K_S \frac{dT}{dx_2} A_2$$

$\frac{dT}{dx_1} \rightarrow$ Temp gradient of x_1

$\frac{dT}{dx_2} \rightarrow$ Temp gradient of x_2

$L \rightarrow$ Latent heat of fusion

$K_L \rightarrow$ Thermal conductivity

$A_1, A_2 \rightarrow$ cross section area

$x_1, x_2 \rightarrow$ Isotherm.

For CZ process pull rate is inversely to square root of crystal radius. This is based on I order heat balance equ. which represents dominant heat flux changes during freezing process. x_1 is the constant temp surface which is isotherm just inside liquid. During freezing process which occurs b/w these 2 points heat released to allow the molted Si transform to solid state. This heat must be removed from freezing interface. It is a primary process of heat transfer upto the solid ingot. When the cross sectional areas $A_1 = A_2$ middle term of equ. ① is neglected and equ. will be $L \frac{dm}{dt}$, where $\frac{dm}{dt}$ is amount of freezing per unit time.

$$\frac{dm}{dt} = V_p A N$$

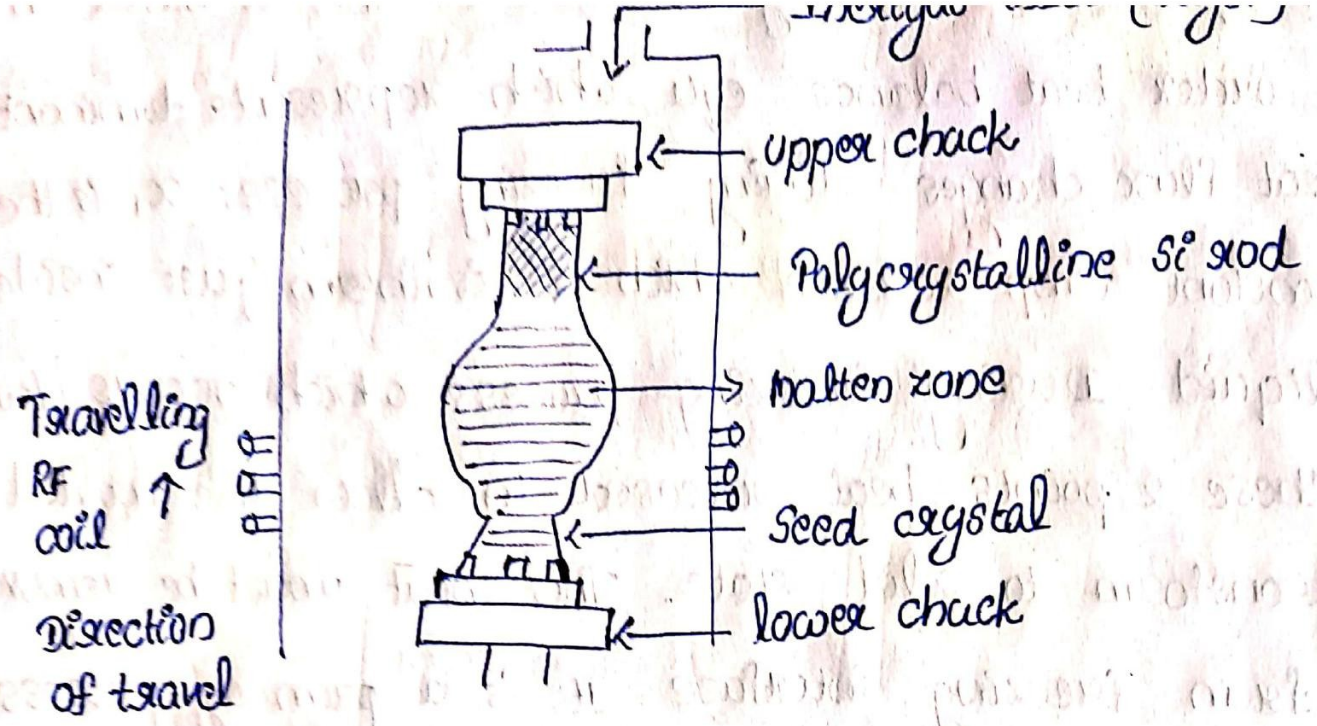
Amount of freezing $\rightarrow \frac{dm}{dt}$

$V_p \rightarrow$ Pulling rate , $A \rightarrow$ Cross sectional area.

$N \rightarrow$ Density of Si

$$(2) \Rightarrow \quad L V_p A N = K_S \frac{dT}{dx_2} A_2$$

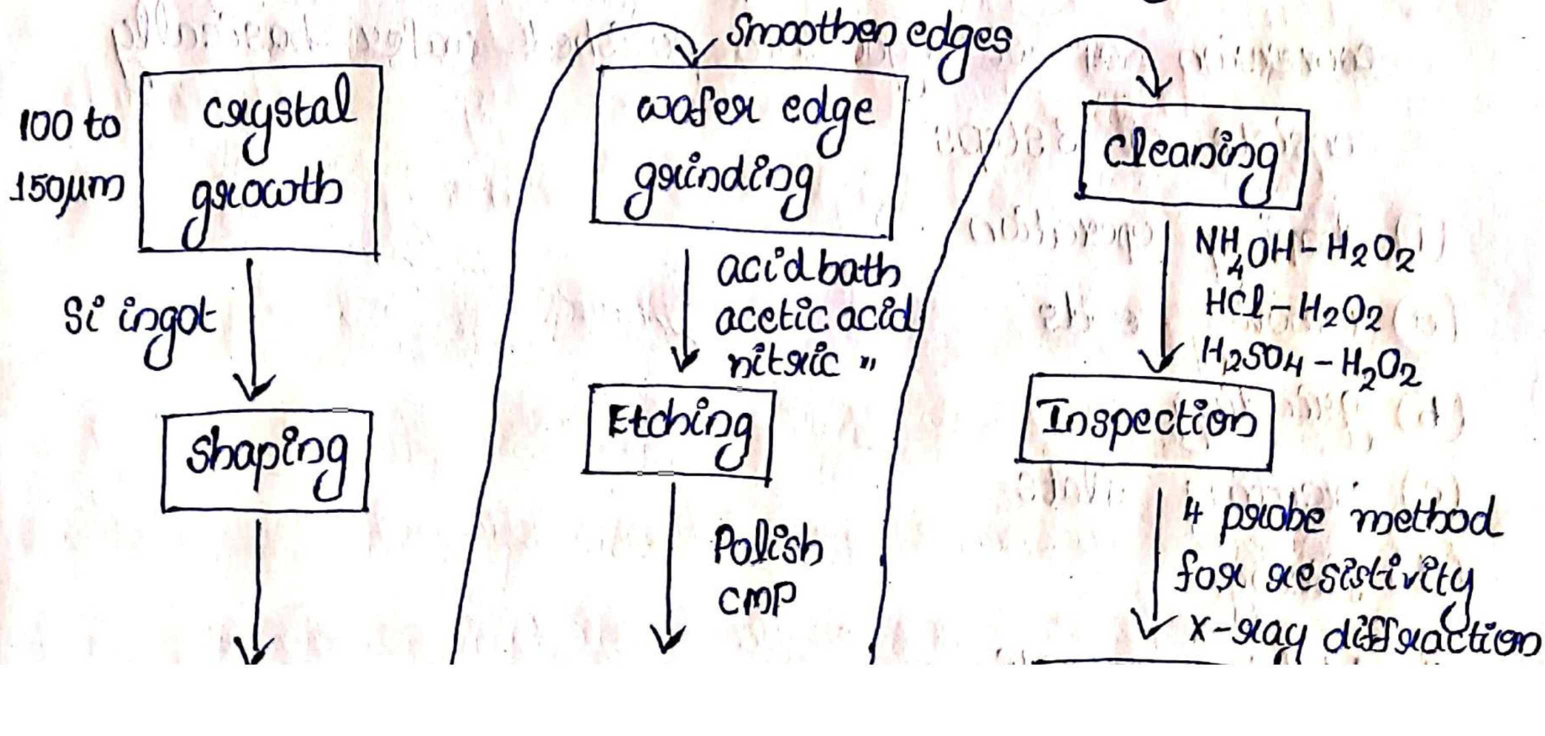
maximum pulling rate, $V_p = K_S \frac{dT}{dx_2} \frac{A_2}{LAN} = \frac{K_S}{LN} \frac{dT}{dx_2}$



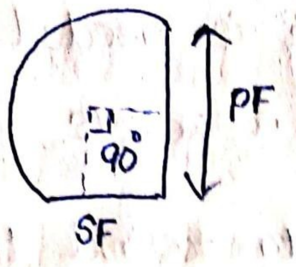
Floatzone process is suited for small wafer production with low oxygen impurity. Here a poly crystalline EGS rod is fused with single crystal seed of desired direction and orientation. This is taken in an inert gas furnace and melted along the length of the rod by an travelling RF coil. The Si will freeze onto the end of crystal seed and as the seed crystal is slowly pulled out of the melt, it will pullup a solidified mass of Si that will be a crystallographic continuation of seed crystal. The movement of RF coil starts from the fused region containing seed and travels up as shown in figure. When the coil moves up the region below solidifies with same crystalline orientation as that of the seed crystal. The furnace is filled with an inertgas like Ar to reduce gaseous impurity and add desired concentr. of dopand. Also since no crucible is needed it can be used to produce oxygen free Si wafers.

Disadvantage - The difficulty is to extend this technique for large wafer since the process produces large no. of dislocations and hence it is used for small scale applications requiring low oxygen content wafers.

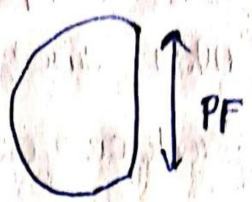
Wafer preparation (Si shaping)



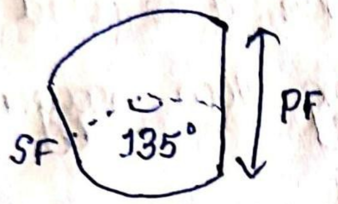
(a) ptype <100>



(b) ptype <111>



(c) ntype <100>



(d) ntype <111>

Wafer preparation

conversion of Si ingot into polished wafer basically consists of 3 steps

(i) shaping operation

(a) removing ends

(b) grinding

(c) creating flats

(d) slicing

(ii) Etching operations

(iii) Polishing operations

Once the crystal ingot is obtained using crystal growth process the extreme top and bottom portions of the ingot are cut using industrial grade tipped saws. And the ingot surface is ground to produce a constant and exact diameter ~~with~~ it is usually 100, 155 or 150 mm. Before processing further ingots are checked for resistivity and orientation using 4 point probe technique (for resistivity) and x-ray diffraction (for orientation) at both ends. After this 1 or more flats are ground along the length of ingot. There are 2 types of flat

→ primary flat - This is a ground related to specific crystal direction. This act as a virtual reference to the orientation of wafer and also as a mechanical locator in the automated processing equipment.

→ secondary flat - This is used for identification of wafer, dopant type and orientation. After making the flats the individual wafers are sliced per required thickness. Inner diameter slicing is most commonly used technique. As a final shaping step edge contouring when a radius is ground on the rings of the wafer.

etched through any damage and contaminated regions. This is usually done in acid bath with a mixture of HCl, nitric acid and acetic acid.

- Polishing - the surfaces are polished firstly by a rough abbrasive polish and followed by a chemical mechanical polishing procedure (CMP). In CMP fine SiO_2 particles suspended in a acqueous sodium hydroxide solution is used. Wafers are typically single sided or double sided polished.

- Chemical cleaning - After polishing wafers are thoroughly cleaned to remove organic contaminants and other metal impurities. Commonly used are aqueous crystal of NH_4OH, H_2O_2, $HCl - H_2O_2$, $H_2SO_4 - H_2O_2$. All of these solutions are efficient in removing mechanic impurities, but $HCl - H_2O_2$ mixture is the best. The $NH_4OH - H_2SO_4$ based mixture will also remove organic contaminants. A typical cleaning sequence would be a $H_2SO_4 - H_2O_2$ clean followed by HF acid with deionized water rinses following

Oxidation refers to the conversion of Si wafers to Si oxide. Purposes are,

Purpose of growing oxide on Si

- To serve as a mask against implant or diffusion of dopant into Si
- To provide surface passivation.
- To isolate one device from another (dielectric isolation as opposed to junction isolation.
- To act as a component in MOS structure.
- To provide electrical isolation of multilevel metallisation crystals.

Growth mechanism

- Si exposed to ambient condition as a native oxide on its surface, around 3nm thickness. But this is too thin for MOS applications hence a thicker oxide needs to be grown. This is done by consuming underlying Si to form Si oxide. This is a grown layer.

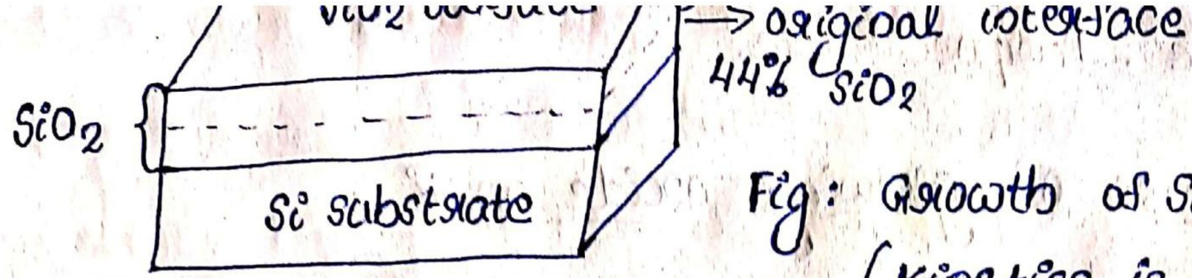


Fig: Growth of SiO_2 (Kinetics is given by deal grove method)

Growth techniques

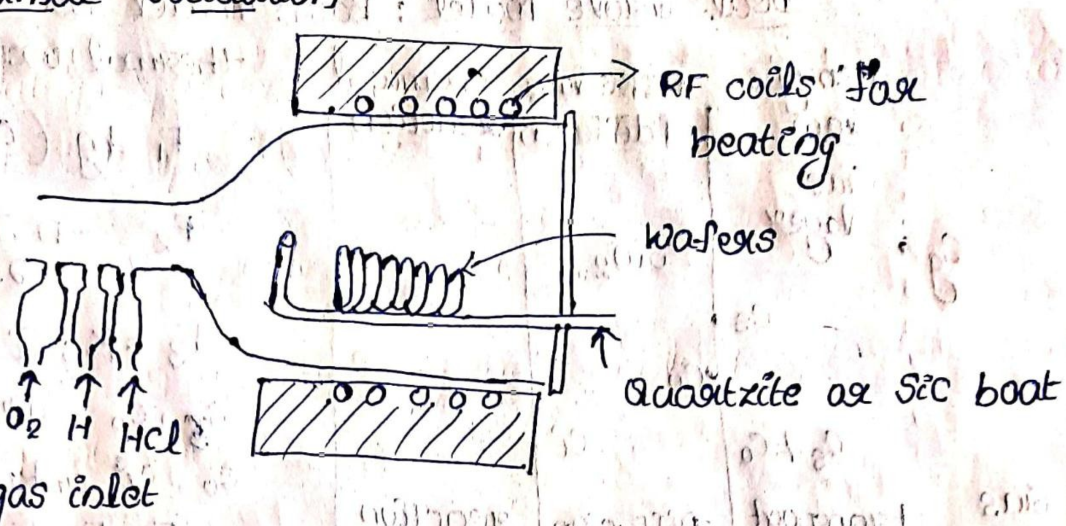
oxidation techniques choosen depend upon the thickness and oxide properties.

(i) Dry oxidation - Thin oxidation layer
(ii) wet oxidation
(iii) High pressure dry or wet oxidation for vapour phase
(iv) CVD
(v) Plasma oxidation

(i) Dry oxidation

$$Si(s) + O_2(g) \rightarrow SiO_2(g)$$

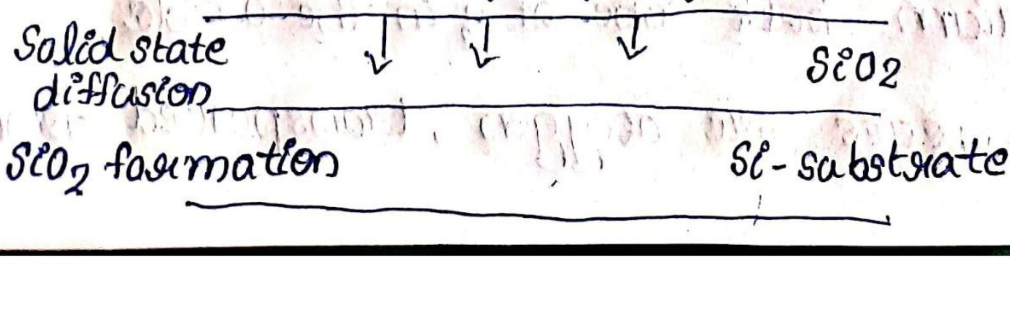
Thermal oxidation



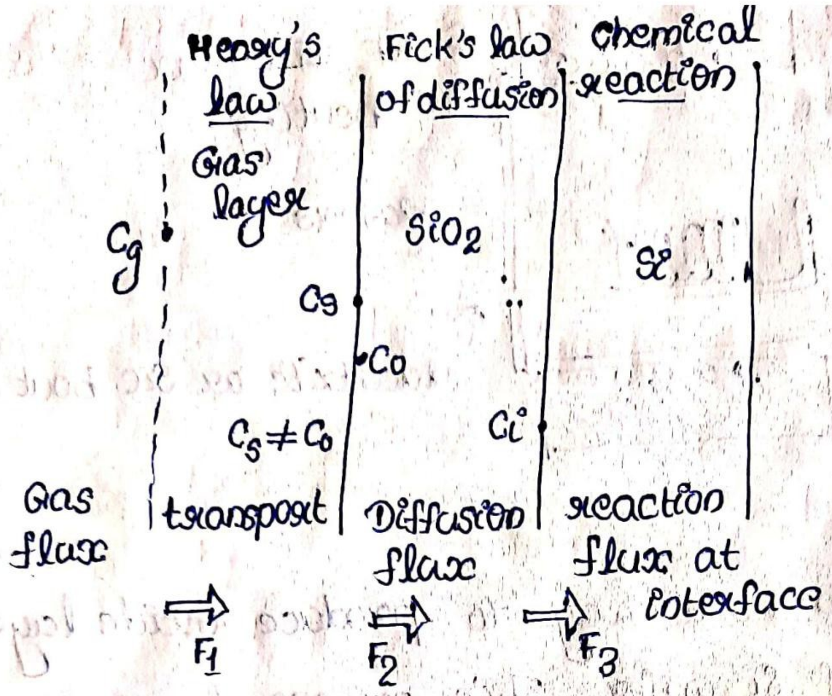
Thermal oxidation is a way to produce oxide layer on the surface of Si wafer. This process forces an oxidising agent to diffuse into wafer at high temp and react with it. Thermal oxidation usually performed at a temp. b/w 800 and 1200°C resulting a high temp. oxide layer. It may either use water or molecular oxygen as the oxidant.

The oxidising ambient may also contain several % of HCl. The chlorine removes melted ions that may occur in oxide layer.

Kinetics of SiO_2 growth



thermal oxidation



$F_1 = h_g (C_G - C_s)$

$F_2 = -D \frac{\partial c}{\partial x}$

• $F_3 = K_s C_i$

C_s, C_o Henry's law

C_o, C_i unknown

$F_1 = F_2 = F_3$

Deal grove model used for predicting oxide thickness for thermal oxides larger than about 300 Å (1 Å = 0.1 nm) ie it describes kinetics of Si oxidation. This model is generally valid for temp. b/w 700 and 1300°C. Therefore 4 O_2 concentration in figure are,

C_g - conc. of O_2 in gas stream far from the wafer

C_s - Conc. of O_2 in the gas at the surface of Si oxide interface.

C_o - Conc. of O_2 in the oxide at outer surface of wafer.

C_i - Conc. of oxidising species at inner surface

C^* - Equilibrium bulk conc. of O_2 in the oxide.

Oxidising species are oxygen, transported from bulk of gas phase to the gas interface with flux F_1 (gas transport flux). Transported across existing oxide towards Si with flux F_2 (diffusion flux). React at the Si, SiO_2 interface with flux F_3 (reaction flux at interface.

- For steady state condition $F_1 = F_2 = F_3$.
- By gas law, $PV = NKT$ or $\frac{P}{KT} = \frac{N}{V} = C^* = C$

If flux F_1 can be linearly approximated by assuming that flux of oxidant from gas phase to gas oxide interface is proportional to the difference of oxidant concentration C_g and C_s. $F_1 = h_g (C_g - C_s)$;

C_g is directly proportional to P_g and C_s is directly proportional to C_s.

$$\frac{P_g}{KT} = \frac{N}{V} = C_g \quad \text{(By Gas law)}$$

$$\frac{P_s}{KT} = C_s$$

$$F_1 = h_g \left[\frac{P_g}{KT} - \frac{P_s}{KT}\right] = \frac{h_g}{KT}\left[P_g - P_s\right]$$

where K is the Boltzman constant.

$P_g \rightarrow$ Partial pressure of O_2 in gas.

$P_s \rightarrow$ " " " gas adjacent to oxide surface

Henry's law states that equilibrium concentration of proportional to partial pressure of that species in the surrounding gas.

Henry's

$C^* \propto P_g$ or $C^* = HP_g$ $\quad$ $H \rightarrow$ constant

IIIly, $C_o = HP_s$

Sub in equ. of flux F_1

$$F_1 = \frac{h_g}{HkT}(C^* - C_o)$$

$h \rightarrow$ Gas phase mass transfer coefficient in solid.

$$h = \frac{h_g}{HkT}$$

$$F_2 \propto \frac{C_o - C_i}{x}$$

$$F_2 = D\left(\frac{C_o - C_i}{x}\right)$$

If the flux F_2 is related to the movement of the oxidising species with the oxidising layer or diffusion of the oxidising species.

Fick's law of diffusion

Diffusion is proportional to concentration gradient. Conc. gradient is defined as the ratio of diffusion in concentration to thickness of oxide.

$$\text{Conc. gradient} = \frac{C_o - C_i}{x}$$

$$F_2 \propto \frac{C_o - C_i}{x}$$

$F_2 = D\left(\frac{C_0 - C_i}{x}\right)$

$D \rightarrow$ Diffusion coefficient

$x \rightarrow$ oxide thickness

Flux F_3 is related to reaction of Si° with oxidising species. $F_3 \propto C_i^\circ$

$F_3 = K_s C_i^\circ$

$K_s \rightarrow$ Rate constant of Si° oxidation

Under steady state condition $F_1 = F_2 = F_3$.

Equating F_3 and F_1

$$K_s C_i^\circ = \frac{hg}{HKT}\left(C^* - C_0\right)$$

$$K_s C_i^\circ = \frac{hgC^*}{HKT} - \frac{hgC_0}{HKT}$$

$$C_0 = \left[1 + \frac{K_s x}{D}\right] C^*$$

Equating F_2 and F_3

$$D\left(\frac{C_0 - C_i^\circ}{x}\right) = K_s C_i^\circ$$

$$DC_0 - C_i D = K_s C_i^\circ x$$

$$DC_0 = C_i\left(K_s x + D\right)$$

$$C_i^\circ = \frac{DC_0}{K_s x + D}$$

$$K_s C_i^\circ = \frac{K_s C_0}{\left(1 + \frac{K_s x}{D}\right)}$$

$$h\left(C^* - C_0\right) = K_s\left[\frac{C_0}{1 + \frac{K_s x}{D}}\right]$$

$$hC^* - hC_0 = K_s\left[\frac{C_0}{1 + \frac{K_s x}{D}}\right]$$

$$hC^* = K_S\left(\frac{C_0}{1+\frac{K_S x}{D}}\right) + hC_0$$

$$C_0 = \frac{hC^*\left[1+\frac{K_S x}{D}\right]}{K_S + h\left[1+\frac{K_S x}{D}\right]}$$

$$C_0 = \frac{C^*\left[1+\frac{K_S x}{D}\right]}{1+\frac{K_S}{h}+\frac{K_S x}{D}}$$

Limiting case 1

$D \to \infty$

$$C_i = \frac{C_0}{1+\frac{k_S x}{D}} = C_0$$

$$C_0 = \frac{C^*(1+0)}{1+\frac{k_S}{h}+0} = \frac{C^*}{1+\frac{k_S}{h}}$$

Limiting case 2

$D \to 0$

$C_i = 0$

$C_0 = C^*$

When diffusibility = 0, $C_i = 0$ and $C_0 = C^*$. Here the oxidation process purely controlled by diffusion. Here oxidation rate depends on supply of oxidant to the surface.

is no difference b/w the $1+\frac{K_s}{h}$ conc. of oxidising species from outer and inner surface. Here the oxidation rate depends upon chemical reaction rate constant (K_s) so it is called reaction controlled oxidation.

To calculate rate of oxide growth we define N_1 as the no. of oxide molecule in a unit volume of oxide layer. Growth rate characteristics can be obtained by dividing the flux at interface by the no. of molecules of oxidants per unit volume of oxide.

$$\frac{dx}{dt} = \frac{F_3}{N_1}$$

$$F_3 = K_s C_i$$

$$\therefore \frac{dx}{dt} = \frac{K_s C_i}{N_1}$$

$$N_1 \frac{dx}{dt} = K_s C_i = \frac{K_s C_0}{\frac{K_s x}{D}+1} = K_s \cdot \frac{1}{\left(\frac{K_s x}{D}+1\right)} \frac{C^*\left[1+\frac{K_s x}{D}\right]}{1+\frac{K_s}{h}+\frac{K_s x}{D}}$$

$$N_1 \frac{dx}{dt} = \frac{K_s C^*}{1+\frac{K_s}{h}+\frac{K_s x}{D}}$$

$$\int\left[1+\frac{K_s}{h}+\frac{K_s x}{D}\right]dx = \int\frac{K_s C^*}{N_1}dt$$

$$x+\frac{K_s x}{h}+\frac{K_s x^2}{2D} = \frac{K_s C^*}{N_1} \times t + P \quad \text{--- (1)}$$

initial conditions to above equ.

initial condition, $t=0$, $x=x_i$

$$x_i + \frac{K_s x_i}{h} + \frac{K_s x_i^2}{2D} = P$$

$$P = \left(1 + \frac{K_s}{h}\right) x_i + \frac{K_s x_i^2}{2D}$$

$$(1) \Rightarrow \quad x + \frac{K_s x}{h} + \frac{K_s x^2}{2D} = \frac{K_s C_i^* \times t}{N_1} + \left[1 + \frac{K_s}{h}\right] x_i + \frac{K_s x_i^2}{2D}$$

$$x^2 + x\left(1 + \frac{K_s}{h}\right)\frac{2D}{K_s} = \left(\frac{K_s C^* t}{N_1} + \left(1 + \frac{K_s}{h}\right) x_i + \frac{K_s x_i^2}{D}\right) \div \frac{K_s}{2D}$$

$$x^2 + x\frac{2D}{K_s} + \frac{x}{h} 2D = \frac{C^* 2Dt}{N_1} + \left[1 + \frac{K_s}{h}\right] x_i \times \frac{1}{\frac{K_s}{2D}} + x_i^2$$

$$x^2 + x\left[\left(1 + \frac{K_s}{h}\right)\frac{2D}{K_s}\right] = \left[\frac{2C^* D}{N_1}\right] t + x_i^2 + \left[\left(1 + \frac{K_s}{h}\right)\frac{2D}{K_s}\right] x_i$$

Let $A = \left(1 + \frac{K_s}{h}\right)\frac{2D}{K_s}$

$B = \frac{2C^* D}{N_1}$

$$\Rightarrow \quad x^2 + Ax = Bt + x_i^2 + Ax_i$$

$$= B\left[t + \left(\frac{x_i^2 + Ax_i}{B}\right)\right]$$

$$\tau = \frac{x_i^2 + Ax_i}{B}$$

$\rightarrow x^2 + Ax = B(t+\tau)$

$$x^2 + Ax - B(t+\tau) = 0$$

$\sqrt{(1+x)} = 1 + \frac{x}{2}$

$$x = \frac{-A \pm \sqrt{A^2 - (4 \times -B(t+\tau))}}{2} = \frac{-A \pm \sqrt{A^2 + 4B(t+\tau)}}{2} \quad \text{--- (2)}$$

Applying limiting conditions then,

(2) This is dealgrove oxidation kinetics equ. which produces growth of SiO_2.

When $t \to 0$

$$x = \frac{-A \pm A\sqrt{1 + \frac{4B(t+\tau)}{A^2}}}{2}$$

$$= \frac{-A \pm \left[A^2 + \frac{2B}{A}(t+\tau)\right]}{2}$$

$$= \frac{-A + A + \frac{2B}{A}(t+\tau)}{2}$$

$$x = \frac{2B}{2A}(t+\tau) = \frac{B}{A}(t+\tau)$$

When $t \to \infty$

$$x = \frac{-A \pm \sqrt{A^2 + 4B(t+\tau)}}{2}$$

$$= \frac{-A \pm A\sqrt{1 + \frac{4B}{A^2}(t+\tau)}}{2}$$

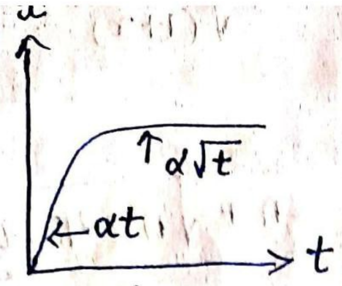
$$= \frac{-A \pm A\sqrt{\frac{4B}{A^2}(t+\tau)}}{2}$$

$$= \frac{-A + A \times \frac{2B}{A}(t+\tau)}{2}$$

$$= \frac{2\sqrt{B}\sqrt{t+\tau}}{2} = \sqrt{Bt}$$

As $t \to 0$, $x = \frac{B}{A}(t+\tau)$, linear

As $t \to \infty$, $x = \sqrt{Bt}$, parabolic



follows a linear rate of growth.
when $t \to \infty$ oxidation follows a parabolic rate of growth.

Relationship b/w oxide growth and oxide thickness

$$x^2 + Ax - B(t+\tau) = 0$$

$$x^2 + Ax = B(t+\tau)$$

diff. w.r. to t on both sides

$$2x\frac{dx}{dt} + A\frac{dx}{dt} = B\frac{dt}{dt} + 0 \qquad [\tau = \text{constant}]$$

$$(2x+A)\frac{dx}{dt} = B$$

$$\frac{dx}{dt} = \frac{B}{A+2x}$$

$\frac{dx}{dt}$ → oxide growth rate

x → oxide thickness

$\frac{dx}{dt}$ slows down with an increase in x.

Impurity segregation

Impurities both intensional and unintentional are introduced into Si ingot intentional dopants are mixed into melt during crystal growth. while

support ambient etc. The ... impurities ...
different solubility in solid and melt. And equili. Segregation coefficient K_0 can be defined as ratio of equili. conc. of impurity in solid to that in liquid.

$$K_0 = \frac{C_S}{C_L}$$ (Conc. of impurity in solid)

Q) Cz process

1) A Si ingot which should contain 10^{16} Boron atoms/cm^3 is to be grown by Cz technique. What conc. of Boron atoms should be in melt to give the required conc. in ingot. If the initial load of Si in the crucible is 60kg. How many grams of Boron (atomic weight = 10.8) should be added. The density of molten Si is 2.53g/cm^3. Given segregation constant K_0 = 0.8.

A) Given,

no. of boron atoms = 10^{16} atoms/cm^3

molten silicon = 60 kg

K_0 = 0.8

Atomic weight = 10.8

Density of Si = 2.53g/cm^3

$$K_0 = \frac{C_S}{C_L}$$

$$C_L = C_S/K_0 = \frac{10^{16}}{0.8} = 1.25\times10^{16}\ cm^{-3}$$

Density of Si

$= \frac{60 \times 10^3}{2.53} = 23.715 \times 10^3 cm^3$

Conc. of boron atoms in the melt = C_L × melt volume

$= 1.25 \times 10^{16} \times 23.715 \times 10^3$

$= 2.96 \times 10^{20}$ Boron atoms

1 Avagadro = 10.8g

No. of grams of Boron = $\frac{2.96 \times 10^{20} \times 10.8}{6.022 \times 10^{23}}$

$= 5.30 \times 10^{-3}$

Thermal oxidation

2) A 1000Å SiO_2 layer grown in dry O_2 at 1000°C followed by 0.5μm thick oxide in wet oxidation at same temp. Determine total time of oxidation.

Given, $x = 0.1 \mu m$, $\tau = 0.37$

$B = 0.12 \mu m^2/hr$

$B/A = 0.07 \mu m/hr$

A) $x^2 + Ax = B(t+\tau)$

$\frac{x^2}{B} + \frac{x}{(B/A)} = t + \tau$

$\frac{(0.1 \times 10^{-6})^2}{0.12 \times 10^{-6}} + \frac{0.1 \times 10^{-6}}{0.07 \times 10^{-6}} = t_1 + 0.37$

$t_1 = 1.87 hrs$

$$\frac{[illegible]}{B} + \frac{[illegible]}{(B/A)} = \frac{[illegible]}{1.27} + \frac{[illegible]}{1.27} = 0.113\,hrs$$

$$\frac{(0.1\times10^{-6})^2}{0.12\times10^{-6}} + \frac{(0.1\times10^{-6})}{0.07\times10^{-6}} = t_2 + 0.113$$

$$t_2 = 1.29\,hrs$$

$$T_{total} = t_1 + t_2 = 1.87 + 1.29 = 2.358$$

Mathematical analysis of diffusion inorder to fully characterise diffusion we need to determine following parameters,

(1) How far dopant atoms go inside the semiconductor atoms

(2) Doping profile

(3) Where the junction is present

(4) Conc. of impurity at surface.

(5) Determine there we need to perform mathematical analysis

Constant source diffusion

Impurity concentration at the surface is is maintained at a constant level throughout the diffusion cycle. Boundary conditions are,

$$C(0,t) = C_s \quad , \quad C(\alpha, t) = 0 \quad , \quad C(x,0) = 0$$

$C(x,0) \rightarrow$ [illegible]

out.

Solving Fick's law,

$$C(x,t) = C_S \, \text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right), \; t > 0$$

erfc → complementary error function

x → distance in cm

D → Diffusivity

t → Time for diffusion in seconds.

$$\text{erfc}(x) = 1 - \text{erf}(x)$$

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-\lambda^2} d\lambda$$

To obtain total amount of impurity, $Q_T = \int_0^{\infty} C(x,t)\,dx$

$$\int_0^{\infty} C(x,t)\,dx = \int^{\infty} C_S \, \text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) dx$$

Let $\left(\frac{x}{2\sqrt{Dt}}\right) = z$

$$dx = dz \cdot 2\sqrt{Dt}$$

$$Q_T = \int_0^{\infty} C_S \, \text{erfc}(z) \, 2\sqrt{Dt} \, dz$$

$$= C_S \, 2\sqrt{Dt} \int_0^{\infty} \text{erfc}(z) \, dz$$

$$= \frac{C_S \, 2\sqrt{Dt}}{\sqrt{\pi}}$$

Depth of diffusion, $c(x,t) = C_S \, \text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$

$c(x,t) = C_B$

$$C_B = C_S \, \text{erfc}\left(\frac{x_j}{2\sqrt{Dt}}\right)$$

Let C_B be the amt of impurity in semiconductor before carrying out the diffusion also called background diffusion. The depth of diffusion is equal to the background conc. of purity.

$c(x,t) = C_B$ for depth of diffusion

$$c(x,t) = C_S \, \text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$$

$$C_B = C_S \, \text{erfc}\left(\frac{x_j}{2\sqrt{Dt}}\right)$$

$$\frac{C_B}{C_S} = \text{erfc}\left(\frac{x_j}{2\sqrt{Dt}}\right)$$

$$x_j = 2\sqrt{Dt} \, \text{erfc}^{-1}(C_B/C_S).$$

In the fabrication of monolithic IC's constant source diffusion is commonly used for isolation and emitter diffusion because it maintains high surface conc. by a continuous introduction of dopant.

Limited source diffusion / Const. dose diffusion / Drive in D

Here it predetermined amt. of impurity is introduced into the crystal, the diffusion takes place in 2steps

(i) pre deposition step ... are deposited on Si wafer during a short time.

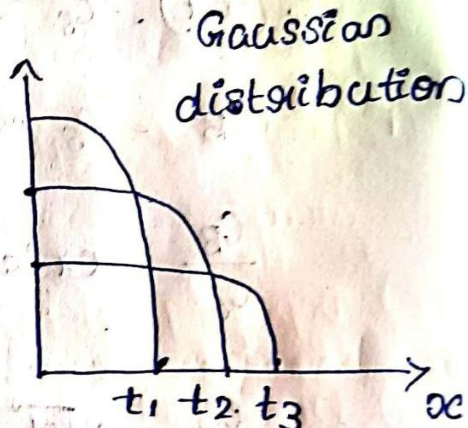
(ii) Drive in step - Impurity source is turned off and amt of impurities already deposited during I step are allowed to diffuse into the Si wafer.

With these type of diffusion the depth of the penetration of impurities during predeposition step is assumed to be negligible as compared to the final depth after completing the drive cycle.

(1) $\int_0^\infty c(x,t)\,dx = Q_T$ constant conc.

(2) $c(x,0) = 0$

(3) $c(\infty,t) = 0$



Solving Fick's law,

$$c(x,t) = \frac{Q_T}{\sqrt{\pi D t}}\, e^{-x^2/4Dt}$$

In this case total amt of impurity atoms is constant and surface conc. decreases as a function of time. A finite quantity of diffusing matter is placed on Si wafer diffusion progresses from this limited source and it is assumed that all dopant is consumed during the process.

$Q_T \rightarrow$ Amt of material placed on surface before diffusion

$D \rightarrow$ Diffusion coefficient

$x \rightarrow$ " distance

$t \rightarrow$ " time

Surface concentration

It can be calculated by substituting 0 in equ.

$$c(x,t) = \frac{Q_T}{\sqrt{\pi Dt}} e^{-x^2/4Dt}$$

surface conc. decreases with increase in time.

$$c(0,t) = \frac{Q_T}{\sqrt{\pi Dt}}$$

$$C_S = \frac{Q_T}{\sqrt{\pi Dt}}$$

Junction depth

It is the point in the semiconductor were impurity conc. is equal to background doping conc. ie, at $c = C_B$ and $x = x_j$

$$C_B = \frac{Q_T}{\sqrt{\pi Dt}} e^{-x_j^2/4Dt}$$

$$e^{-x_j^2/4Dt} = \frac{C_B \sqrt{\pi Dt}}{Q_T}$$

$$\frac{C_B\sqrt{\pi Dt}}{Q_T} = \frac{1}{e^{x_j^2/4Dt}}$$

$$e^{x_j^2/4Dt} = \frac{Q_T}{C_B\sqrt{\pi Dt}}$$

Taking natural log on both sides,

$$\frac{x_j^2}{4Dt} = \ln\left(\frac{Q_T}{C_B\sqrt{\pi Dt}}\right)$$

$$x_j^2 = 4Dt \ln\left(\frac{Q_T}{C_B\sqrt{\pi Dt}}\right)$$

$$x_j = \left(4Dt \ln \frac{Q_T}{C_B\sqrt{\pi Dt}}\right)^{1/2}$$

Diffusion systems

Impurities are diffused from their compound source. The method of impurity delivery to the wafer is determined by nature of impurity source. 2 step diffusion is widely used technique. Type of impurity diffusion whether complimentary error fn distribution or Gaussian distribution is determined by the choice of operating condition.

2 step diffusion consists of pre-deposition step and drive in step. In the former step a constant source

diffusion is carried out for short time at a temp. 1000°C. In later step impurity supply is shut off and existing dopant is allowed to diffuse into the body of semiconductor at a temp. 1200°C in an oxidising atmosphere. Oxide layer which forms on the surface of wafer during this step prevents further impurities from entering, already deposited impurity prevent from diffusing out. Final impurity profile is a fn. of diffusion conditions, i.e. temp, time and diffusion coefficients.

Diffusion furnace

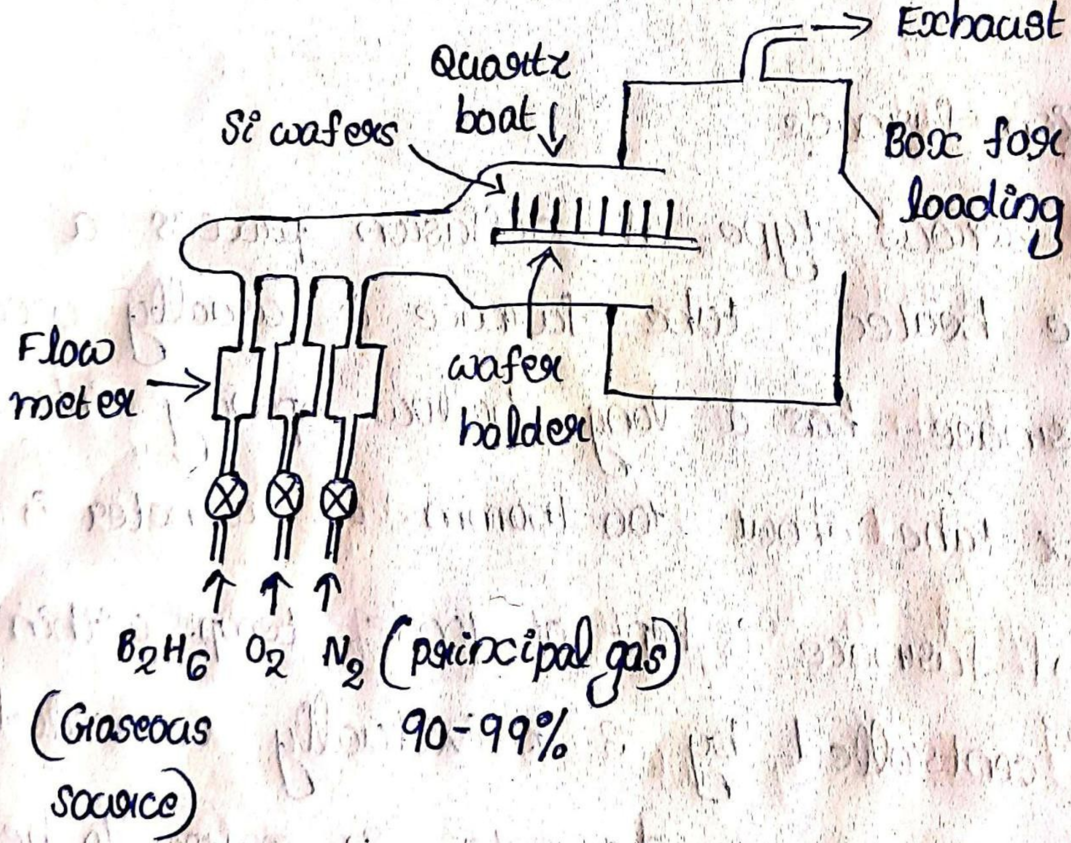
For the various type of diffusion process a resistance heated tube furnace is usually used. Tube furnace has a long hollow opening into which a quartz tube about 100-150mm in diameter is placed. Temp. of furnace is kept at 1000°C, temp. within furnace can be controlled by 3 individually controlled adjascent resistance elements. Si wafers to be processed are stacked up vertically into slots in a quartz boat, and insert into the furnace tube.

Boron

$D = 10^{-16}\ m^2/sec$ at $1150^\circ C$

Solid solubility - 5×10^{26} atoms/m^3

Diffusion of p type impurity, Boron is an exclusive choice as an accepted impurity in Si. It has a moderate D, which is convenient for precisely controlled diffusion. Surface conc. can be widely varied due to its solid solubility.



$$Si + O_2 \rightarrow SiO_2$$

$$2B_2H_6 + 3O_2 \rightarrow B_2O_3 + 6H_2$$

Boron diffusion using diborane source (
is a gaseous source of boron. This can be directly introduced into diffusion furnace. A no. of other gases are metered into furnace. Principal gas flow in the furnace will be N_2 which acts as a relatively inert gas, and is used as a carrier gas to be a dilutant for other more reactive gases. N_2 carried gas will generally makeup 90-99% of total gas flow. A small amt of O_2 and very small amt of gaseous source of Boron will makeup rest of gas flow. The following reactions will be occuring simultaneously at the surface of Si wafers.

This process is a chemical vapour deposition (CVD) of a glassy layer on Si surface, which is a mixture of Si glass and Boron glass is called Boron Silica glass (BSG). BSG glassy layer is a viscous liquid at a diffusion temp. Boron conc. in BSG is [illegible] Si surface will be saturated with Boron at solid solubility limit throughout time of diffusion process as long as BSG remains presented. This is a constant source diffusion. It is often called deposition diffusion.

Mathematical Analysis Of Diffusion

①

In order to fully characterise diffusion we need to determine the following parameters

- how far the dopent atoms go inside the semiconductor atoms.
- doping profile
- where the junction is present
- Conc. of impurity at the surface.

To determine these parameters we need to perform mathematical analysis.

1. Constant surface conc / diffusion from unlimited source / Predeposition.

Here the impurity conc at the semiconductor surface is maintained at a constant level through out diffusion cycle.

The boundary conditions are

$C(0,t) = C_s$. (C_s is the surface concentration)

$C(\infty, t) = 0$ (conc. of impurity at $x=0$. from surface)

$C(x,0) = 0$ (conc. of impurity before diffusion is carried out).

$C(x,0)=0$ ② $C(0,t)=C_s$ $C(\infty,t)$

Solving Fick's law, the boundary condition,

$$C(x,t) = C_s \ erfc\left(\frac{x}{2\sqrt{Dt}}\right) \quad t>0.$$

erfc - complimentary error function

x - distance in cm.

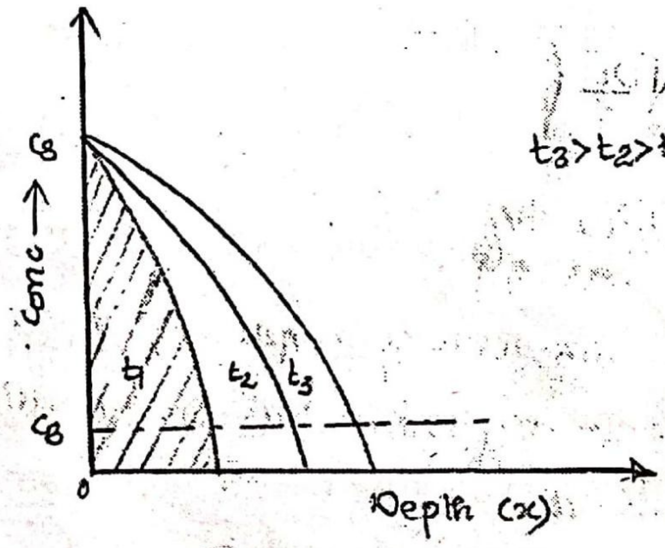
D - Diffusivity constant

t - time for diffusion in sec.

$$\frac{C(x,t)-C_0}{C_s-C_0}$$

$$erf(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-\eta^2}$$

$$erfc(x) = 1 - erf(x)$$



$t_3 > t_2 > t_1$

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x exp(-\eta^2)$$

$$\int_0^\infty erfc(x)\, dx = \frac{1}{\sqrt{\pi}}$$

The plot shows that irrespective of the duration during which diffusion is carried out all the 3 curves starts at C_s . ie surface conc is always constant

To obtain Total Amount of Impurity

To calculate the total amount of impurity which has been introduced into the semiconductor during diffusion we calculate the area under the curve

$$Q_T = \int_0^\infty C(x,t)\,dx$$

$$= \int_0^\infty C_S \,\mathrm{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)dx$$

Let $\frac{x}{2\sqrt{Dt}} = z$.

$$\frac{dx}{2\sqrt{Dt}} = dz$$

$$dx = 2\sqrt{Dt}\,dz$$

$$Q_T = \int_0^\infty C_S \,\mathrm{erfc}(z)\, 2\sqrt{Dt}\,dz$$

$$= C_S\, 2\sqrt{Dt} \boxed{\int_0^\infty \mathrm{erfc}(z)\,dz} \quad \frac{1}{\sqrt{\pi}}$$

$$Q_T = \frac{C_S\, 2\sqrt{Dt}}{\sqrt{\pi}}$$

$$\boxed{Q_T = 2C_S\sqrt{\frac{Dt}{\pi}}}$$

→ Depth of Diffusion

Let C_B be the amount of impurity in the semiconductor before carrying out diffusion this is also called as background concentration.

The depth of diffusion or junction occurs where the conc of impurity due to diffusion is equal to the background conc. Let this depth be x_j then at $x = x_j$.

$C = C_B$

$$\therefore \quad C_B = C_S \,\mathrm{erfc}\left(\frac{x_j}{2\sqrt{Dt}}\right)$$

$$\frac{C_B}{C_S} = \operatorname{erfc}\left(\frac{x_j}{2\sqrt{Dt}}\right)$$

$$\boxed{x_j = \operatorname{erfc}^{-1}\left(\frac{C_B}{C_S}\right) \cdot 2\sqrt{Dt}}$$

2 Constant ~~Source~~ Dose diffusion / Constant total impurity / Drive in diffusion

In this case the total amount of impurity atoms is constant and the surface conc. decreases as a function of time. Here a finite quantity Q_T of diffusing matter is placed on a silicon wafer. Diffusion progresses from this limited source and it is assumed that all the dopent is consumed during the process.

The boundary condition are:

i) $\int_0^\infty C(x,t)\,dx = Q_T$ = constant

ii) $C(x,0)=0$

iii) $C(\infty,t)=0$.

Solving Fick's law by this boundary conditions

$$\boxed{C(x,t) = \frac{Q_T}{\sqrt{\pi D t}}\, e^{-x^2/4Dt}} \quad \text{— (1)}$$

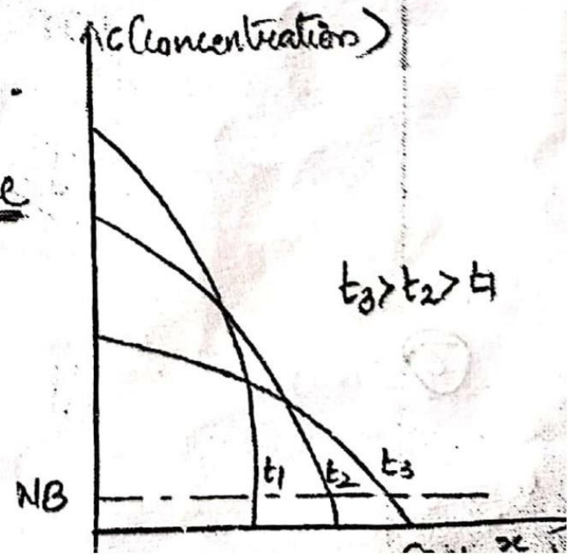
Q_T - amount of material placed on the surface before diffusion

D - Diffusibility.

x - diffusing ion distance

t - diffusion time in sec.

Eq (1) is Gaussian diffusion Profile



Surface Concentration

It can be calculated by substituting x=0 in eq(1)

$$c(0,t) = \frac{Q_T}{\sqrt{\pi D t}} = C_s$$

Junction Depth

It is the point in semiconductor where the impuri[ty] is equal to the background dopping conc.

i.e at $C = C_B$, $x = x_j$

$$\therefore c(x_j, t) = \frac{Q_T}{\sqrt{\pi D t}} e^{-x_j^2/4Dt} = C_B$$

$$\left(\frac{C_B\sqrt{\pi D t}}{Q_T}\right) = e^{\frac{-x_j^2}{4Dt}}$$

$$\frac{C_B\sqrt{\pi D t}}{Q_T} = \frac{1}{e^{x_j^2/4DT}}$$

$$e^{\frac{x_j^2}{4DT}} = \frac{C_B\sqrt{\pi DT}}{Q_T} \frac{Q_T}{C_B\sqrt{\pi DT}}$$

$$e^{\frac{x_j^2}{4DT}} = \frac{C_B\sqrt{\pi DT}}{Q_T} \frac{Q_T}{C_B\sqrt{\pi DT}}$$

$$\frac{x_j^2}{4DT} = \ln\frac{Q_T}{C_B\sqrt{\pi DT}}$$

$$\boxed{x_j = \sqrt{4DT \ln\frac{Q_T}{C_B\sqrt{\pi DT}}}}$$

mixture
Borosilica Of silica & Boron glass, is called glass.

glassy layer
The BSG is a viscous liquid at diff. temperatures.

The Boron conc. in BSG is such that the Si surface will be saturated w/ Boron at a solid solubility limit throughout the time of diff. process as long as BSG remains present. This is a const. source Diff.

It is often called deposition diff.

~~Diffusion of p type impurity~~

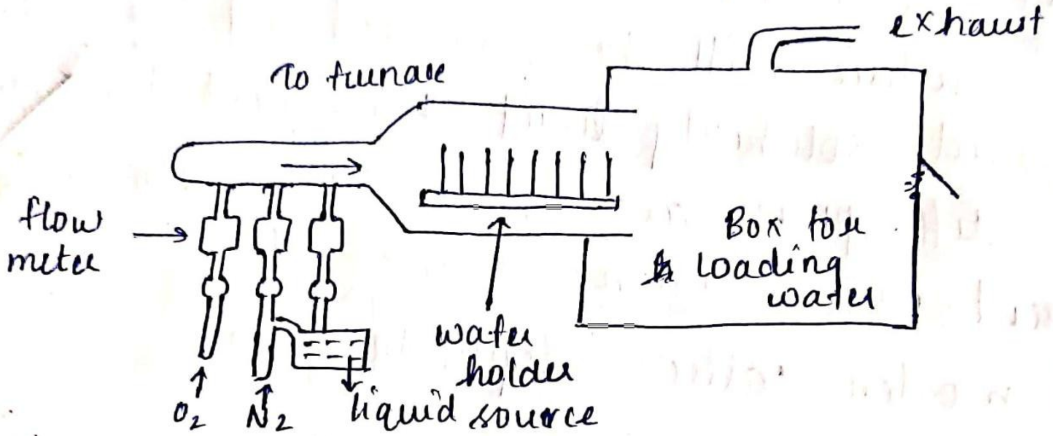
This diff. step is referred as predeposition step in which the dopant atom deposit into the surface regions of Si wafer

The Borosilica glass is preferable because it protect the Si atoms from pitting or evaporating and act as getter for undesirable impurities in the Si.

The predeposition step is followed by a second diff process in which the external dopand source (BSG) is removed such that no additional dopants enter the Si. During this diff. process, the dopants are already in the Si, move further and are thus redistributed.

The in. depth incs and at the same time, the surface conc. decreases. This type of diff. is called driving or limited source diff. (Gaussian distribution)

Boron Diff. using ~~Bro~~ Boron tri Bromide souce:



This is a liquid souce of Boron. In this case, a controlled flow of carrier gas is bubbled through ~~Boron~~ triBromide soln. which with Oxygen again produces ~~Bron~~ Boron trioxide (Borosilica Glass) at the surface of the wafers as per the above eqn.

$$4BBr_2 + 3O_2 \longrightarrow B_2O_3 \text{ (Boron glass)} + 2Br_2$$

$$Si + O_2 \longrightarrow SiO_2 \text{ (glass)}$$

Diffusion of N-type Impurity:

For Phosphorous diff. such compounds as phosphine (PH_3) and Phosphorous Oxychloride ($POCl_3$) can be used. In this case of diff. using $POCl_3$, the reactions occuring at the Si wafer surfaces will be

Ex

$$Si + O_2 \longrightarrow SiO_2 \text{ (Silica glass)}$$

$$4POCl_3 \xrightarrow{+3O_2} 2P_2O_5 + 6Cl_2$$
(Phosp. glass)

This will result in the prodn. of glassy layer on the Si wafer that is a mixture of Phosp. glass & Si glass called Phosphoro Silica glass (PSG) which is a viscous liquid at diff. temp.

The mobility of P atom in this glassy layer and this P conc. is such that, the P conc. at Si surface will be maintained at solid solubility limit throughout the time of diff. process.

The rest of the process for P diffusion is similar to Boron diff. i.e., after predeposition step, ~~dirving~~ driving ~~deposit~~ diffusion is carried out.

Other common N-type dopants are Antimony and Arsenic. These dopants have low diff. const. Therefore, they are useful materials for earlier diff. stages such as N^+ buried layers. Antimony is sometimes preferred because it's less toxic but arsenic has a higher solid solubility limit and can provide bigger surface concentration of dopants.

Properties of Diffusion:

1) solid solubility & dopant ↑, SC ↑
2) Diffusion temp.
3) Diffusion time.
4) Surface cleanliness and defects in Si crystal.

1) This is the factor which decides the no. of atoms per unit volume can be added for a specific diffusion, that means, number of atoms per unit vol. must be less than solid solubility limit.

2) Diff. temp → ↑ the temp. gives more thermal energy and thus higher velocities to the diffused impurities. The diff. coeff. critically depends upon temp. Therefore, the temp. profile of diff. furnace must have higher tolerance of temp. variation over its entire area.

3) Diff. Time: Increases diff. time or diff. coeff. D have similar effect on jn. depth. For gaussian distr., the net conc. will decrease due to impurity compensation and can approach 0 with increase in diff. time.

ASS

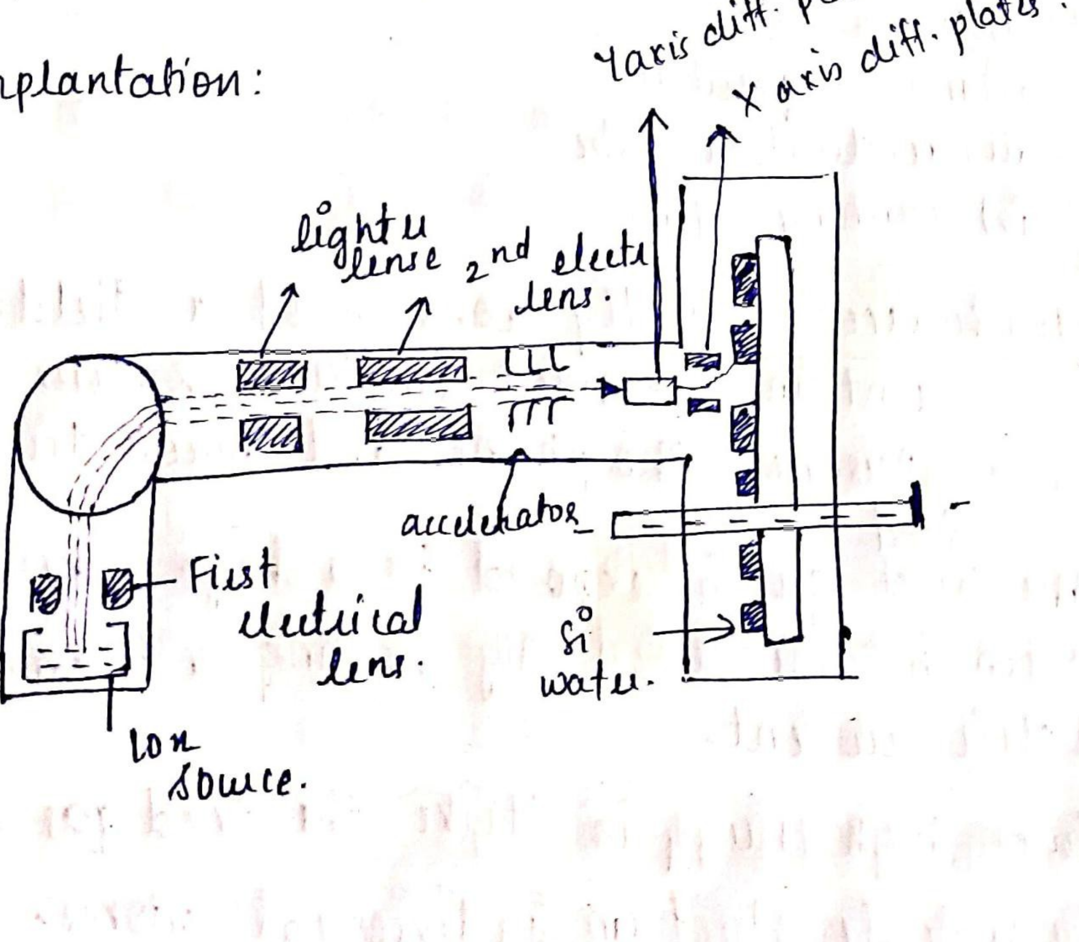
For constant source diff, net impurity conc. on the diffused side of pn jn. shows a steady state increase with time.

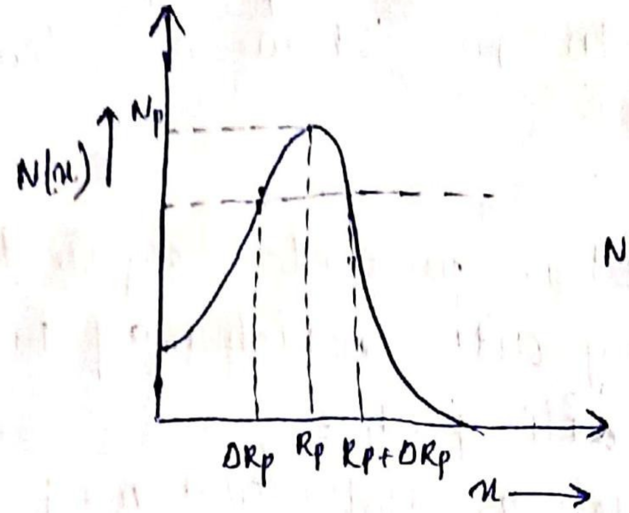
4) The Si surface must be prevented against contaminants during diff. which may interfere with uniformity of diff. profile.
The crystal defects such as dislocation may produce localize impurity conc. This results in the degradation of jn. chara.

MODULE: 2.

Assgt: Method of isolation ckt component Fabrication!

Iron Implantation:





$$N(x) = N_p \cdot \exp\left(\frac{(x - R_p)^2}{2\Delta R_p^2}\right)$$

Ion Implantation. Impurity profile.

Ion implantation s/m consists of:

i) Ion source
ii) mass seperator
iii) acceleration tube
iv) scanning s/m.

<u>Ion source</u>: usually consists of a field gas that contains desired implant species like Arsane, Phosphane and Boron tri fluoride (BF_3)

The ~~field~~ feed gas is exposed to a high energy e⁻s which is produced by heating a filament using electric current.

These high energy e⁻s strike the feed gas & break the molecules to form individual atoms.

<u>Mass seperator</u>;

Ions extracted from source contain diff. ion species &

travel at a relatively ↑ speed into mass. sep. chamber. Chamber consists of an analyzed magnet shaped at 90°. Mg. field of analyzer magnet causes the ion species to be deflected into an arc.

If the mg. field strength is B, ion charge is Q, then the ion moves in a circle of radius given by

$$r = \frac{1}{B} \frac{\sqrt{2 V m}}{q}.$$

Vm is the source volt. into mass of ion.

The mg. field B can be adjusted such that only the desired ion will pass through the slit and others will be rejected.

Acceleration tube:

To achieve additional ion acceleration, the ions which come out of m.s are accelerated in an $\vec{E}$ inside the the acc. column.

Increasing the acc. voltage will increase the ion implantation or jn. depth into the wafer.

Scanning slm:

Electrostatic scanning deflects ion beam across a stationary wafer by applying specific controlled volt to a set of x y deflection planes.

Ion Implantation Profile:
(graph drawn before)

Range Theory:

Ion stopping:

As each implanted ion enters the target, it undergoes a series of collision with target atoms until it finally comes to rest at some depth.

This depth is called projected range expressed in μm.

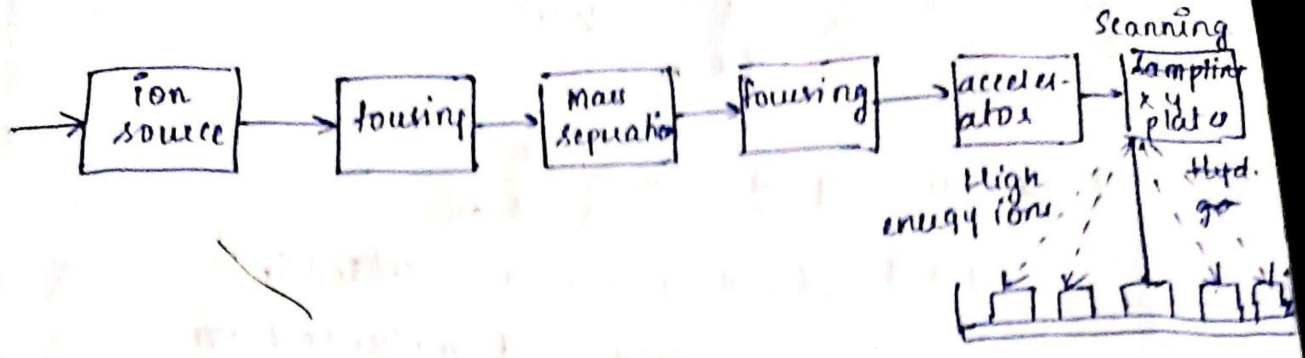
Projected range depending on impurity used and the implantation energy.

There are 2 basic ion stopping mechanism by which an energetic ion are brought to rest.

1) Nucleus stopping
2) E^- stopping.

Nucleus stopping:

The +ve ion scatters as it encounters +vely charged nucleus. The nucleus stopping is elastic so the energy lost by incoming ion is transferred to the target atom. Thus it dislocate the target nuclei from their original state.



- Ion implantation profile is a Gaussian distribution.
- Implantation of ions expressed as a fn of dist. x from the surface

$$N(x) = N_p \cdot \exp\left(\frac{-(x-R_p)^2}{2\Delta R_p^2}\right)$$

where N_p is the peak conc. of implanted ions.

when $x = R_p$.

$N(x) = N(R_p) = N_p$.

$$N_p = \frac{Q}{\sqrt{2\pi}\,R_p} = \frac{0.4\,Q}{\Delta R_p}$$

Q is the implantation ~~toss~~ dose given by

$$Q = \frac{1}{A}\int \frac{I}{q}\,dt$$

, where A = area of implantation.

I = ~~condn~~ Implantation beam current

ΔR_p is straggle deviation in projected range.

LSS Theory of Ion Implantation:

~~E~~ Electron stopping: It is caused by the interaction of ion with a cloud of es surrounding the target atoms.

Electronic stopping is given by:

$$S_e = \left(\frac{dE}{dx}\right)_e = K\sqrt{E}$$

Nucleus stopping $S_n = \left(\frac{dE}{dx}\right)_n$

Total stopping power = $N(S_e + S_n)$

N = no. of implanted ions

$$S_T = N\left(\left(\frac{dE}{dx}\right)_e + \left(\frac{dE}{dx}\right)_n\right)$$

Implantation dosage Q:

It is the no. of implanted ions per unit surface area.

$$Q = \frac{1}{A}\int \frac{I}{q}\, dt$$

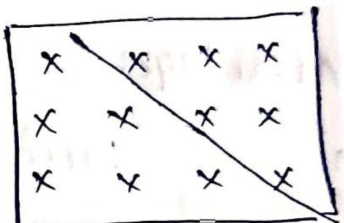
Annealing:

After the ions have been implanted; ~~they are~~ They are lodged principally in intestitial position, in the Si crystal structure and surface region into which the ion implantation has taken place will be heavily damaged by the impact of high energy ions. The ~~disc array~~ disarray of Si atoms in the surface region is ~~often~~ open to the extend that this region is no longer crystalline in structure but rather amorphous. To restore this surface region back to well-ordered crystalline state and to allow Implanted ions to go into the substitutional sides in the crystalline struct. the wafer must be subjected to annealing process.

The annealing process usually involves heating of wafer at a temp. of 1000°c. for about 30 mins.

Most commonly used annealing techniques are laser beam & e beam annealing.

In such annealing techniques only the surface region of the wafer is heated and recrystallized.



Due to ~~channeling~~ regularity of the single crystal structure, impurity atoms moving in certain dirn can find an open corridor or channel b/w crystal atoms. The ions are moving down through this channel by extended oscillations that result more ~~dep~~ deeper penetration into the target atoms.

$$N_B = N_p \exp\left(-\frac{(x-R_p)^2}{2\Delta R_p^2}\right)$$

$$\frac{N_B}{N_p} = \exp\left(-\frac{(x_j - R_p)^2}{2\Delta R^2 p}\right)$$

$$\ln \frac{N_B}{N_p} = -\frac{(x_j - R_p)^2}{2\Delta R^2 p}$$

$$\ln \frac{N_P}{N_B} = \frac{(x_j - R_p)^2}{2\Delta R^2 p}$$

$$(x_j - R_p)^2 = 2\Delta R_p^2 \times \ln\left(\frac{N_P}{N_B}\right)$$

$$(x_j - R_p)^2 = 2\Delta R_p^2 \times \ln\left(\frac{N_P}{N_B}\right)$$

$$x_j - R_p = \sqrt{2} R_p \left(\ln\left(\frac{N_P}{N_B}\right)\right)^{1/2}$$

$$x_j = R_p + \sqrt{2} R_p \left[\ln\left(\frac{N_P}{N_B}\right)\right]^{1/2}$$

Advantages:

1) Precise control of dopant conc.

2) control of dopant penetration depth.

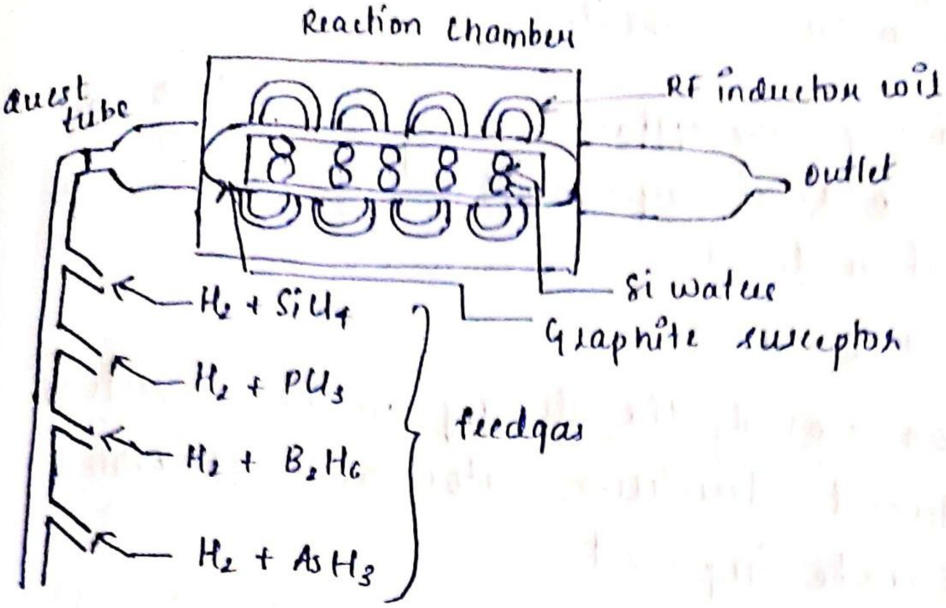
3) Dopant uniformity

4) Low temp. processing.

Disadvantages:

1) Ion implant equipement is highly expensive.

2) High energy ions may damage crystaline structure. In some case it can't be ~~tomplet~~ fully repaired through repairing process.



Epitaxy means arranged upon.
In Epitaxy, a mono crystalline film is formed on the top of the monocrystaline surface.
Thus Epitaxy is a crystalline growth process in which foundation layer acts as a seed crystal
The Epitaxy layer formed on substrate may be n or p-doped.
For p-type doping diBorane and for n-type doping, phosphine are used with the scheme of $SiCl_4$ Hyd. gas.
The Epitaxial growth of pure Si can be represented as

$$SiCl_4 + 2H_2 \underset{1250^\circ C}{\rightleftharpoons} Si + 4HCl$$

Mainly 2 types of Epitaxy :

1) Homo Epitaxy: when an epitaxial layer & substrate on which epitaxy layer formed are of same

material. It is homo epitaxy.
eg: Si grown on Si substrate.

2) Hetero Epitaxy: The epitaxial layer & substrate on which epitaxial layer is formed is not of identical material.

Advantages OF Epitaxy:

1) Designer can control the doping in the structure.
2) using epitaxial structures, performance of RAM & cmos ICs can be improved.

~~3) 2 main epi~~

2 main epitaxial process:
i) Vapour phase epitaxy
ii) Molecular beam epitaxy

In chemical vapour deposition, the film is formed on the surface of the substrate by thermal decomposition and by the reaction of various gaseous compounds. As in CVD, the epitaxial layer is formed from gaseous vapour phase, hence it is called vapour phase epitaxy.

VPE

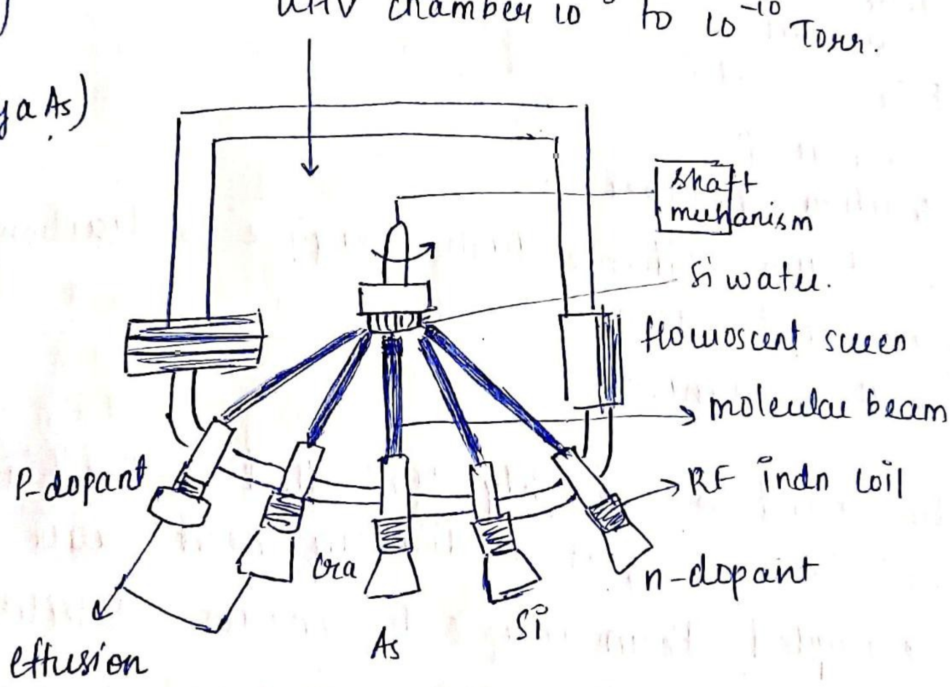
SiH_4
$SiHCl_3$
SiH_2Cl_2
$SiCl_4$

} Si substrate formation.

AsH_3
PH_3
B_2H_6

} dopant.

MBE (GaAs)



(MBE)
Molecular Beam Epitaxy based on evaporation.
It is a process of ~~forming~~ formation of atomic layer by atomic layer crystal growth based on the reaction of molecular or atomic beams with a heated crystalline substrate performed in a ~~ult~~ ultra high vacuum env. (10^{-8} – 10^{-10} Torr.)

In MBE, the film is evaporated & deposited one layer at a time. In this process, no chemical rxns are considered.

The term molecular beam refers to a unidirectional dynamatic flow of atoms or molecules having no collisions among them.

Main parts:

- Effusion cell:

~~The main p~~

- Reaction chamber
- read gun (reflective high energy e^- retraction s/m)
- fluorescent screen.
- Shaft mechanism

Effusion cell is a highly controlled and efficient deposition source for MBE, it radiately heat ~~ttti~~ crucible w/ close coupled thermocouple to ensure stability.

[Effusion - process in which gas escape from a container through a diameter considerably smaller than mean free path of molecules.]

The growth rate of MBE is typically 0.01 to 0.3 μm/min.

MBE is carried out under temp. ranging b/w 600°c to 900°c. which is comparitively low temp.

As this process is very expensive, it is extensively used in special applications such as GaAs technology, Si on insulator & Si on ~~sulphide~~ sapphire technology

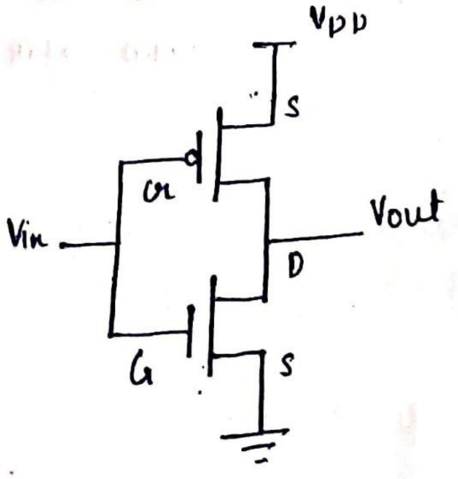
Adv. of MBE over ~~CVD~~ CVD

- Low temp. process, ∴ advantageous for VLSI technology
- Auto doping & auto diff. (diffusion) are minimized
- It is used for generating complicated doping profiles ~~at~~ as it regulates amount of dopant.
- As MBE is based on evaporation of Si, hence no chemical rxns involved in it.

Disadvantages:

- Very difficult to maintain a very low pressure of 10^{-8} to 10^{-10} Torr ~~due to~~ for the formation perfect & pure crystalline structure.
- Very Expensive
- growth rate of MBE is 0.01 to 0.3μm/min which is very small compared to CVD. (1μm/min)

CMOS Inverter:

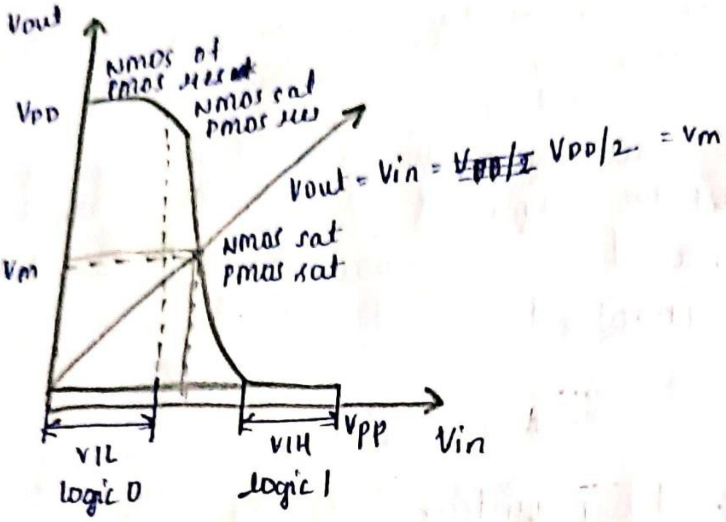


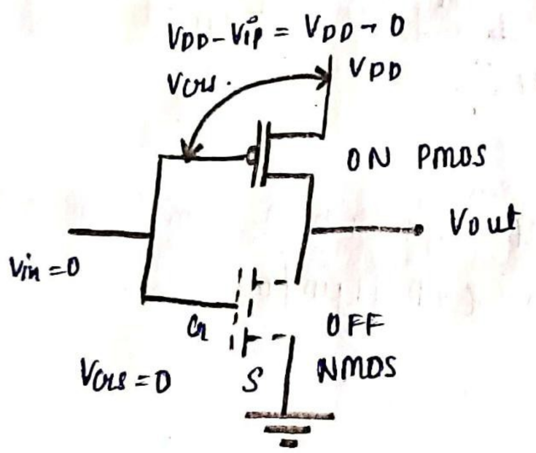
Transistor Switch Model

	condn	State of MOS
NMOS	$V_{gs} < V_{in}$	OFF
NMOS	$V_{gs} > V_{in}$	ON
PMOS	$V_{gs} < V_{tp}$	OFF
PMOS	$V_{gs} > V_{tp}$	ON

NMOS, I_D, D → S

PMOS, I_D, S → D





Low Input Voltage

When $V_{in} = 0$

$V_{GSn} = V_{in} = 0$

$V_{GSn} < V_{in}$

NMOS = off

$V_{GSP} = V_S - V_G$

$= V_{DD} - V_G$

$= V_{DD} - V_{IN}$

$= V_{DD} \rightarrow$ PMOS ON.

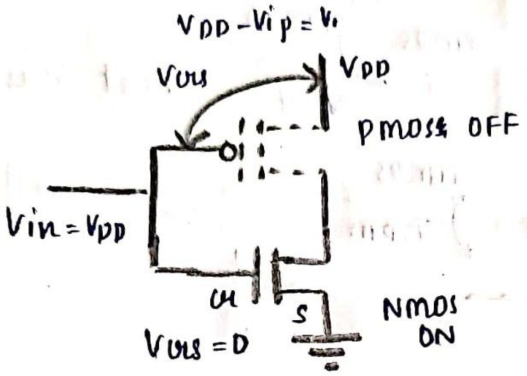
Logic swing at the o/p is $V_{OH} - V_{OL}$, V_{OH} is the o/p high volt. of the ckt i.e., V_{DD}.

V_{OL} is the o/p low volt. of the ckt, i.e., 0 volt.

Logic swing = $V_{DD} - 0 = V_{DD}$,

this is eq. to the value of full power supply.

This is called full rail o/p.



When $V_{in} = V_{DD}$

$V_{GSn} = V_{in} = V_{DD}$

$V_{GSP} = V_S - V_G$

$= V_{DD} - V_{DD}$

$= 0$

PMOS OFF

Logic High Input

V_{OL} = low value of o/p = 0V

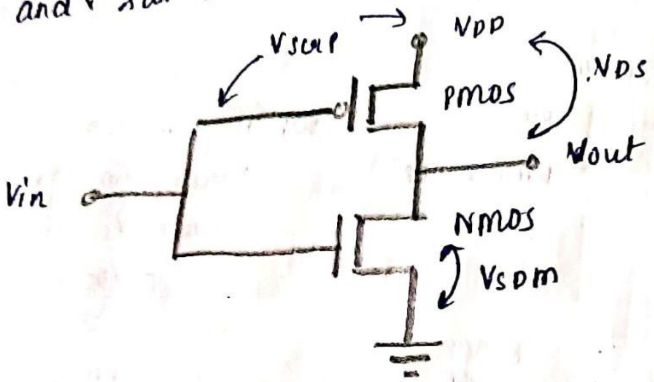
V_{OH} = high value of o/p = V_{DD}

V_{IL} = 0V = i/p ↓ value

V_{IH} = Input high value = V_{DD}

~~Drain current eqn of mos Transistor.~~

Saturation of NMOS transis. happens at $V_{sat} = V_{GS} - V_{Tn}$
and sat. of PMOS transis. happens at $V_{sat} = V_{GS} - V_{Tp}$.



$V_{sat} = V_{GS} - V_{Tn}$.

$V_{SGP} = V_{DD} - V_{in}$ $\quad$ $V_{GSN} = V_{in}$

$V_{SDP} = V_{DD} - V_{out}$ $\quad$ $V_{DSN} = V_{out}$

$V_{sat} = V_{GS} - V_{TD}$

At Threshold voltage:

(V_m)

PMOS sat, NMOS sat

$I_{DP} = I_{DM}$ — Drain current eqn of MOS transis.

$$I_{Dn} = \frac{\beta_n}{2} (V_{GSN} - V_{Tn})^2$$

$$I_{DP} = \frac{\beta_P}{2} (V_{SGP} - V_{TP})^2$$

General:

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

Saturation of NMOS transistor happens at

$V_{sat} = V_{GS} - V_{Tn}$

and sat. of PMOS trans- happens at $V_{sat} = V_{GS} - V_{TP}$

at sat: $V_{in} = V_{DD}/2 \rightarrow \frac{V_{DD}}{2} = V_m$

$V_{SGP} = V_{DD} - V_{in}$

$V_{SGP} = V_{DD} - V_m$

$V_{SGn} = V_{in} = V_m$

$$I_{Dn} = \frac{\beta_n}{2} (V_m - V_{Tn})^2$$

$$I_{DP} = \frac{\beta_P}{2} (V_{DD} - V_m - V_{TP})^2$$

$$\beta_n = k_n \left(\frac{W}{L}\right)_n$$

$$\beta_P = k_P \left(\frac{W}{L}\right)_P$$

~~$I_{Dn} = I_{DP}$~~

~~$I_{Dn} = I_{DP}$~~

~~$\frac{\beta_n}{2} (V_m - V_{Tn})^2$~~

$I_{Dn} = I_{Dp}$

$\frac{\beta_n}{2}(V_m - V_{tN})^2 = \frac{\beta_p}{2}(V_{DD} - V_m - |V_{tp}|)^2$

$\frac{\beta_n}{\beta_p}(V_m - V_t)^2 = (V_{DD} - V_m - |V_{tp}|)^2$

Taking sq. root on both sides:

$\sqrt{\frac{\beta_n}{\beta_p}}(V_m - V_{tn}) = (V_{DD} - V_m - |V_{tp}|^2)$

$\sqrt{\beta_n/\beta_p}\, V_m - \sqrt{\beta_n/\beta_p}\, V_{tn} = (V_{DD} - V_m - |V_{tp}|)$

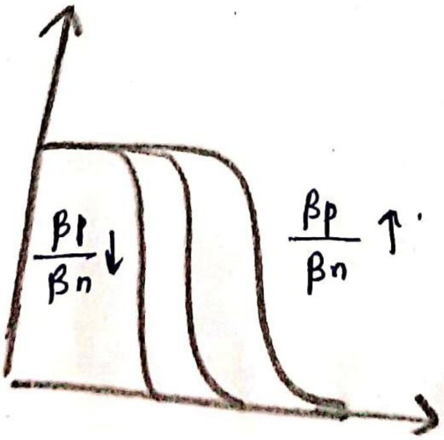
$\sqrt{\beta_m/\beta_p}\, V_m + V_m = V_{DD} - |V_{tp}| + \sqrt{\beta_m/\beta_p}\,(|V_{tn}|)$

$$V_m = \frac{\sqrt{\beta_n/\beta_p}\, V_{tn} + V_{DD} - |V_{tp}|}{1 + \sqrt{\beta_n/\beta_p}}$$

$k_n = \mu_n$

$\beta_p/\beta_n < 1$, low skewed

$\beta_p/\beta_n > 1$, high skewed.



(β_p/β_n) For high skewed inverted chara, charact. shifted to right end.

For low " " " " left end

Q. Calculate switching threshold of CMOS inverter, with the following parameter

$V_{DD} = 5\text{ V}$

$V_{tn} = 0.6\text{ V}$

$V_{tp} = -0.6\text{ V}$

$\beta_n =$ ~~16 μA~~ $50\ \mu A/V^2$

$\beta_p =$ ~~80 μA~~ $35\ \mu A/V^2$

$$V_m = \frac{\sqrt{50/35} \times 0.6 + 5 - |0.6|}{1 + \sqrt{50/35}}$$

$$V_m = 2.33\text{ V}.$$

※ $\sqrt{\frac{\beta_p}{\beta_n}} > 1$ – high skewed

$\sqrt{\frac{\beta_p}{\beta_n}} < 1$ – low skewed.

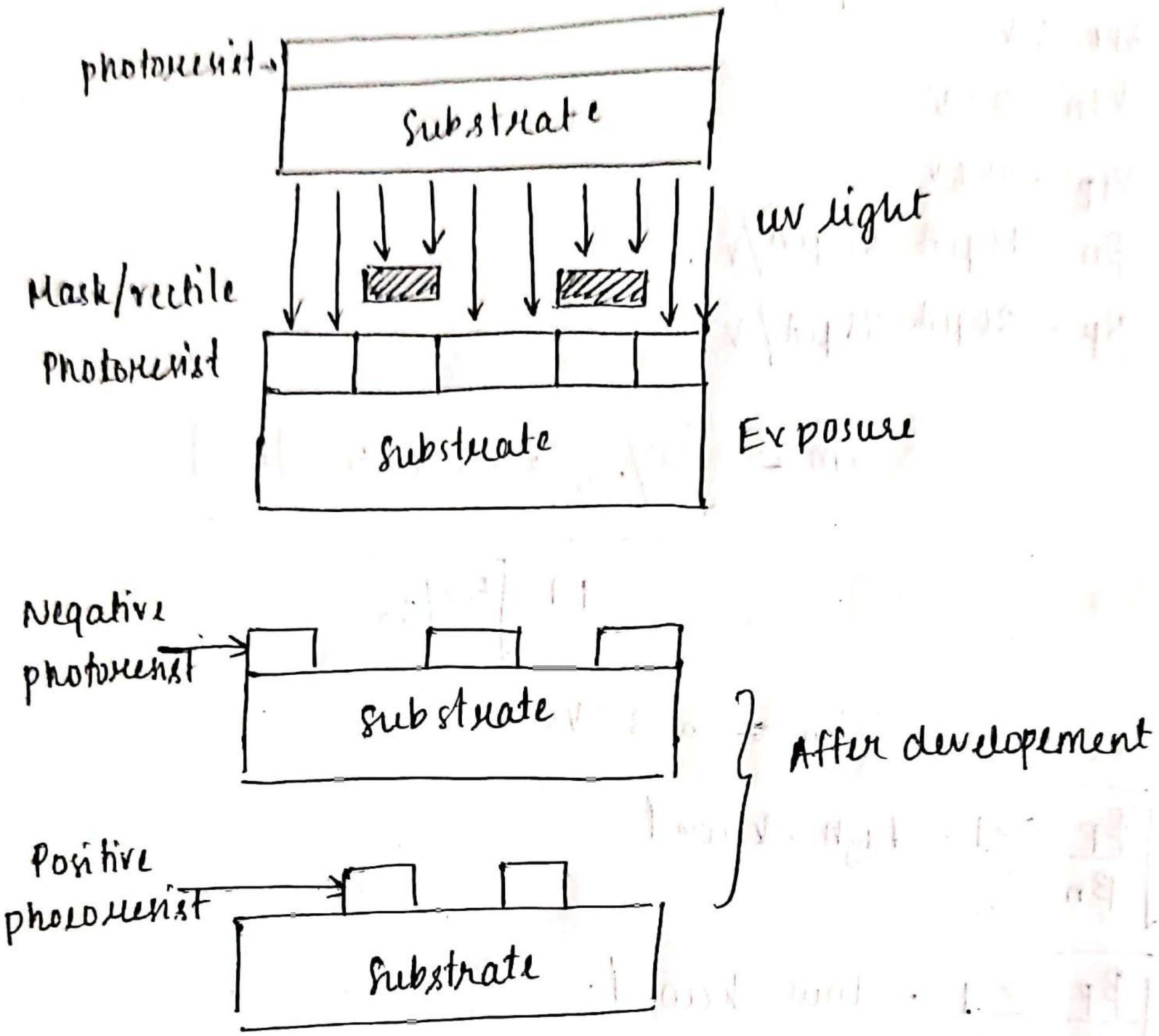
$= 0.83 < 1 \rightarrow$ low skewed.

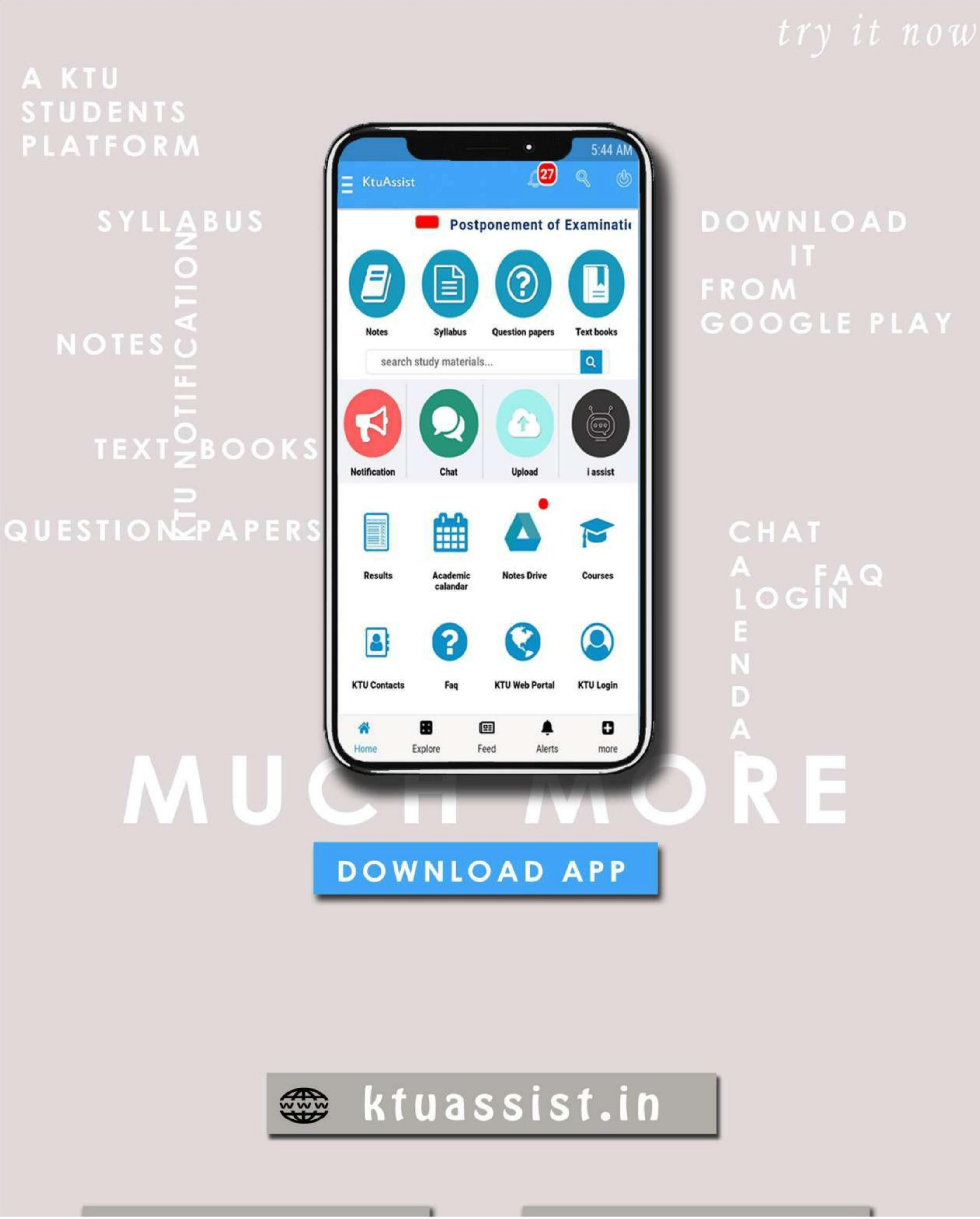
For symm. inverter $\beta_n = \beta_p$.

PHOTOLITHOGRAPHY:

Negative photoresist
- Becomes insoluble after exposure
- When developed, the unexposed parts dissolved
- Cheaper

Positive Photoresist
- Becomes soluble after exposure
- When developed, the exposed parts dissolved
- Better resolution


photoresist
Substrate
uv light
Mask/rectile
Photoresist
Substrate
Exposure
Negative photoresist
substrate
After developement
Positive photoresist
Substrate


try it now
A KTU STUDENTS PLATFORM
SYLLABUS
NOTES
TEXT BOOKS
QUESTION PAPERS
KTU NOTIFICATION
DOWNLOAD IT FROM GOOGLE PLAY
CHAT
FAQ
LOGIN
CALENDAR
5:44 AM
KtuAssist
27
Postponement of Examinati
Notes
Syllabus
Question papers
Text books
search study materials...
Notification
Chat
Upload
i assist
Results
Academic calandar
Notes Drive
Courses
KTU Contacts
Faq
KTU Web Portal
KTU Login
Home
Explore
Feed
Alerts
more
MUCH MORE
DOWNLOAD APP
ktuassist.in

Learning Outcomes,

Student will be able to explain the Static, dynamic, Short circuit power consumption and sources of power consumption

CONTENTS

INTRODUCTION

- I. Static power consumption : due to leakage current drawn from the supply

  Leakage currents

  Subthreshold current

  tunnelling current

  reverse biased diode leakage currents

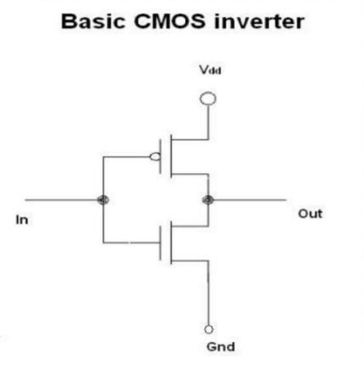
- II. Dynamic power consumption

  Charging and discharging (switching) power dissipation

  Short circuit power dissipation

CMOS inverter mode for static power consumption

In static condition CMOS is either in ON or OFF , they are not switching from one to other
V or I =0 in these two states, power dissipated is “0”



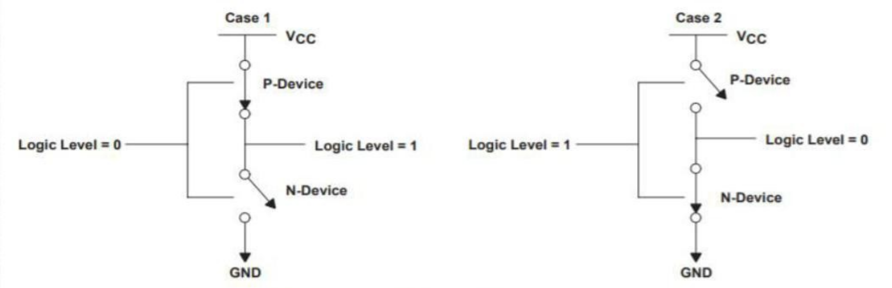


Figure 1. CMOS Inverter Mode for Static Power Consumption

STATIC POWER CONSUMPTION

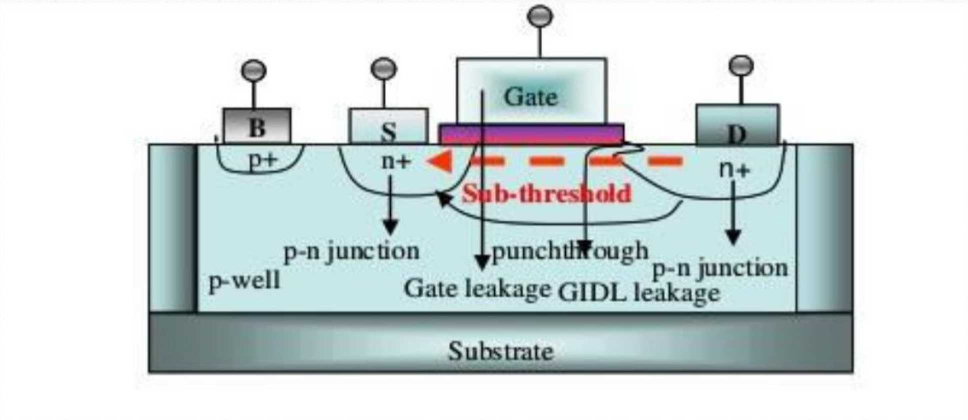
- Static power is defined as the power consumption due to constant current from Vdd to Ground in the absence of the switching activity.
- Static Power dissipation not related to clock frequency or switching activity.
- It occurs when PMOS and NMOS transistors are operating in quiescent mode
- Reasons for static power consumptions

  1.reduction in transistors channel length

  2.reduced gate oxide thickness

  3.reverse biased diode leakage currents

Sources of Leakage current



These are the 6 short channel mechanisms causes static power dissipation

1.Reverse bias pn junction leakage

2.Subthreshold leakage

3.Gate oxide leakage(tunnelling)

4.Gate induced Drain leakage

5.Channel punch through tunnelling

6.Gate current due to hot carrier injection

1,2,4,5 are off state leakage mechanisms

3- tunnelling occurs both in off and on states

6- can occur in off state , but more typically due to high electric field during active mode of operation.

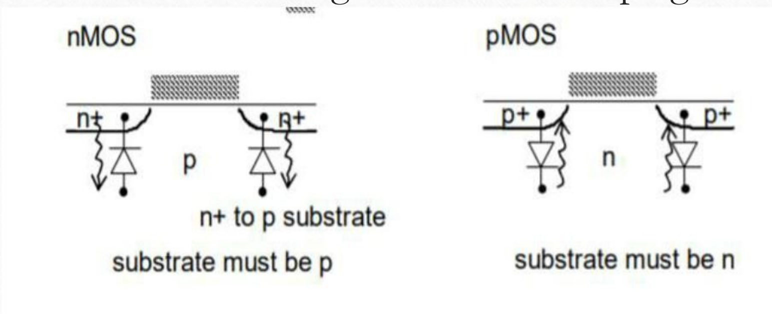
- **1.PN JUNCTION REVERSE BIASED LEAKAGE CURRENT**

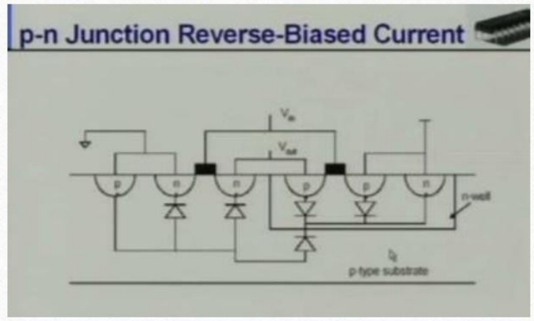
➢ Source and drain junctions are normally reverse-biased , so they will leak current

---

➢ increase with scaling since doping levels are very high

➢ breakdown voltage decreases as doping increases





- **2. SUBTHRESHOLD LEAKAGE CURRENT**

- Subthreshold or weak inversion conduction current between source and drain
- It occurs when the gate voltage is below the threshold voltage,Vth.

$$I_{sub} = I_s \cdot e^{\frac{q(V_{GS} - V_T - V_{offset})}{nKT}} \left(1 - e^{\frac{-qV_{DS}}{KT}}\right) \qquad P_{static} \approx I_{sub} V_{DD}$$

- Subthreshold leakage is the most important contributor to static power in CMOS
- Note that it is primarily a function of VT
- Higher VT, exponentially less current!

- **3.GATE OXIDE LEAKAGE (OXIDE TUNNELLING)**

- tox has been scaling with each technology generation
- We have reached the point where tox is so small the direct tunneling occurs (tox < 2nm)
- Gate leakage = f(tox, VG)
- The downscaling of the gate oxide thickness increases the field across the oxide.
- Results in electron tunnelling from gate to substrate or from substrate to gate.
- NMOS leakage is 3-10X PMOS leakage (electrons vs. holes).
- Below 20 Angstrom, the leakage increases by 10X for every 2Angstrom in gate thickness reduction

- **4.GATE INDUCED DRAIN LEAKAGE (GIDL)**

➢ Drain-to-substrate leakage due to band-to-band tunneling current in very high field depletion region in gate-drain overlap region

---

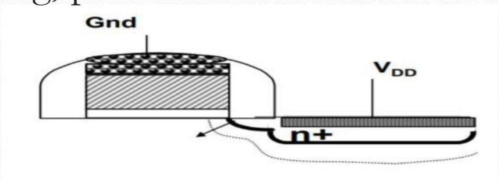
➢ When gate biased to cause an accumulation layer to form the silicon surface, silicon surface under the gate has same potential as the p substrate.

➢ This accumulated holes acts like the P region more heavily doped than the substrate.

➢ Thus the depletion layer narrowed causes field crowding, peak field increases.

➢ Results in Band –Band tunneling

➢ Thinner oxides, lightly-doped drains and high VDD



- 5.**CHANNEL PUNCH THROUGH TUNELLING**

- In short channel devices, due to the drain source proximity, depletion regions at drain- substrate, substrate to source extend in to the channel
- Depletion boundaries decreases with channel length reduction.
- Reverse bias increases with the increase in Vds, results pushing the boundaries from junction
- Combination of channel length and reverse bias causes the depletion region to merge, this is called punch through
- The drain voltage beyond the punch through , lowers potential barrier for majority carriers.
- These increases the subthreshold leakage current.

- V_{PT} is proportional to $N_B(L-W_j)$
- L is the channel length, Wj is the Junction width, V_{PT} is the punch through voltage., NB doping concentration at the bulk

- **6. GATE CURRENT DUE TO HOT CARRIER INJECTION**

- Due to high electrical field near Si/SiO2 interface, electrons or holes can gain sufficient energy to cross the potential barrier
- These carriers enter in to the oxide layer
- This is known as hot carrier injection.
- These hot carriers create leakage current instead of flowing through the channel region.
- Affects the subthreshold value, changes the switching characterestics.

- **Static power dissipation is given by,**

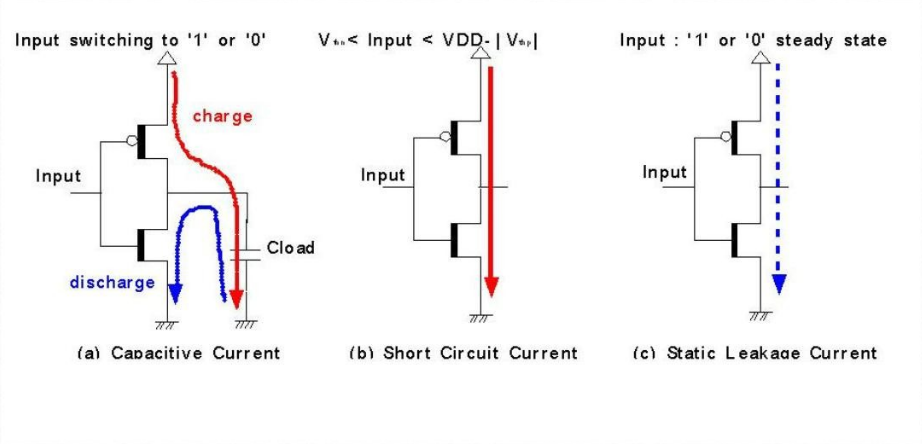
- $P(stat) = V_{DD} \cdot I_{DD(stat)}$
- $I_{DD}(stat)$ is the sum of all leakage currents contributes static power consumption
- $I_{DD}(stat)$=subthreshold current+reverse biased diode leakage current+tunnelling current+other leakage currents

DYNAMIC POWER CONSUMPTION

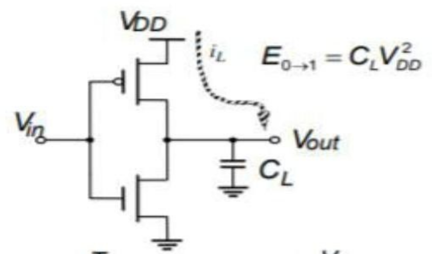
- Dynamic power is the energy consumed during switching, ie; on logic transitions
- Consists of two components – Switching power and internal power

- **Switching power** results from the charging and discharging of external capacitive load on the output
- **Internal power (transient power)** results from the short circuit current flows through the NMOS and PMOS stack during logic transition.
- Internal power is consumed during the short period of time when the input is at intermediate voltage level, during which both transistors can be conducting
- Short circuit current between supply rails is also called crowbar current

DYNAMIC & STATIC POWER CONSUMPTION



DYNAMIC POWER CONSUMPTION



$$E_{0\to1} = \int_0^T P_{DD}(t)dt = V_{DD}\int_0^T i_{DD}(t)dt = V_{DD}\int_0^{V_{DD}} C_L dv_{out} = C_L V_{DD}^2$$

$$E_C = \int_0^T P_C(t)dt = \int_0^T v_{out} i_L(t)dt = \int_0^{V_{DD}} C_L v_{out} dv_{out} = \frac{1}{2} C_L V_{DD}^2$$

CMOS POWER DISSIPATION

$P_{dyn}=\alpha.f.C_L.V_D$ [illegible] ㄷㅅ

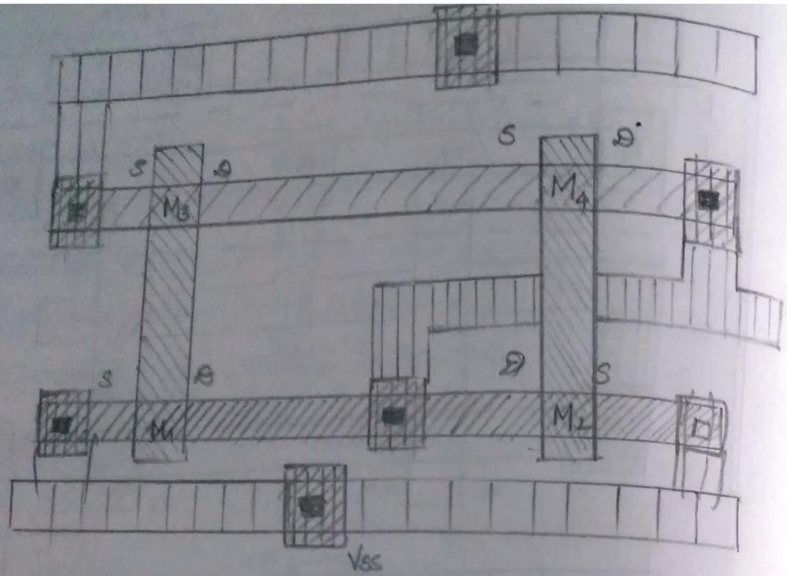
- Dynamic power dissipation is a factor of switching frequency **f** of gate
- Activity factor of inverter **α**
- **$P_{stat}= V_{DD}.I_{DDsat}$**

Power dissipation in CMOS

- Power dissipation in CMOS is the sum of dynamic and static power cosumption components
- P total = P(dynamic)+P (static)
- Dynamic power consumption depends upon clock frequency(f) , capacitive load (C) , supply voltage(Vdd)
- Static power consumption depends upon the supply voltage and leakage currents
- Power consumption of the CMOS is proportional to the square of the supply voltage.

REFERENCE

- Principle of CMOS VLSI Design , Neil Weste, Pearson Education
- http://www.inf.ufrgs.br/logics/docman/book_emicro_butzen.pdf



IV

MODULE 4

MOSFET Logic Design

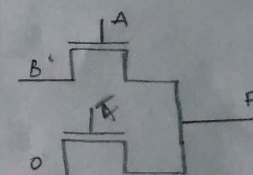
Pass Transistor Logic

Boolean Functions can be realised by using transistor as a switch. This is called pass transistor logic. Pass transistor is nothing but the NMOS or a PMOS transistor. It is called pass transistor because when the gate of NMOS transistor is high it behaves as a on switch and passes the signal between the source and drain. when the gate voltage is low it behaves as a open switch.

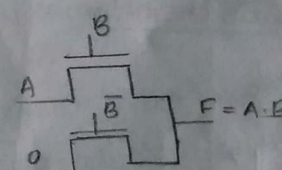
AND Gate

$F = A \cdot B$

$= A \cdot B + \bar{A} \cdot 0$



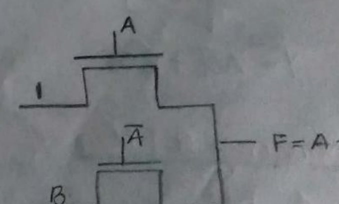
$F = B \cdot A + \bar{B} \cdot 0$



OR Gate

$F = A + B$ $\quad 1+B=1$

$= A \cdot 1 + \bar{A} \cdot B$

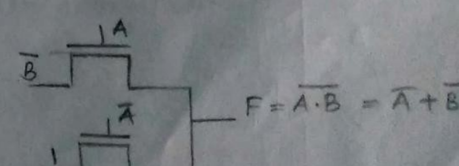


NAND Gate

$F = \overline{A \cdot B}$

$= \bar{A} + \bar{B}$

$F = A \cdot \bar{B} + \bar{A} \cdot 1$

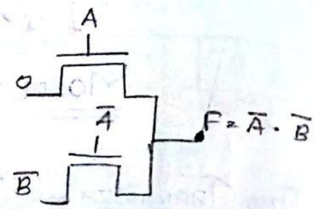


NOR Gate

$F = \overline{A+B}$

$= \bar{A} \cdot \bar{B}$

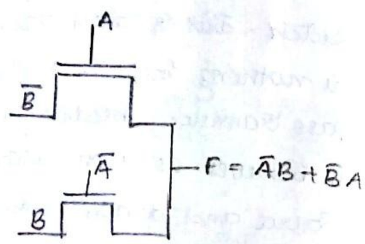
$F = 0 + \bar{B}$



XOR Gate

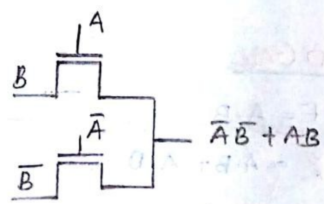
$F = A \oplus B$

$= \bar{A}B + \bar{B}A$

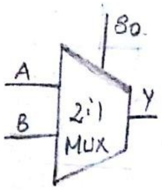


XNOR Gate

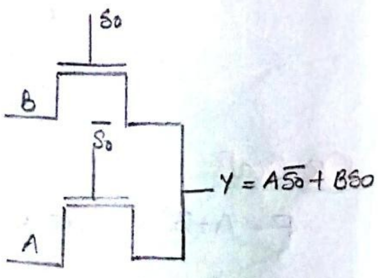
$F = \bar{A}\bar{B} + AB$.



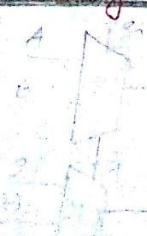
2:1 MUX using Pass transistor



$Y = A\bar{S_0} + BS_0$

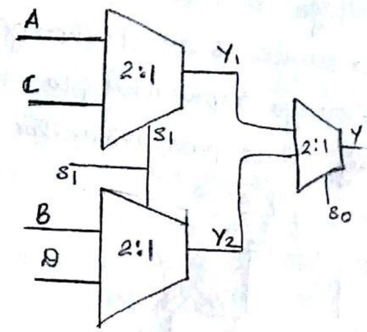


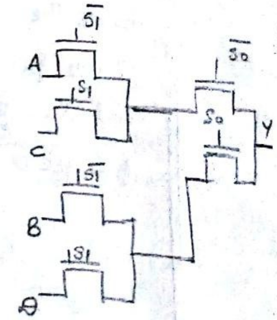
4:1 Mux using 2:1 Mux



$Y = A\bar{S_0}\bar{S_1} + BS_0\bar{S_1} + C\bar{S_0}S_1 + DS_0S_1$

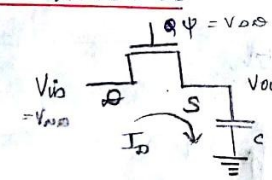
$= \bar{S_0}(A\bar{S_1} + CS_1) + S_0(B\bar{S_1} + DS_1)$





Pass characteristics

NMOS Transistor



PMos source – higher
NMos source – low

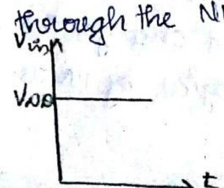
Nmos transmitter of 1 with degradation
0 without degradation

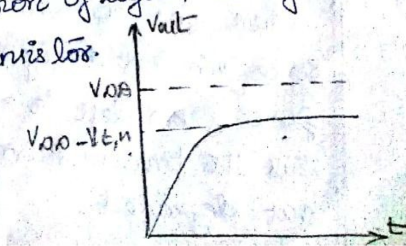
Assume that the load capacitor is fully discharged ie, $V_{out} = 0$.

Logic 1 is applied

When ϕ = logic 1 (Vdd), V_{in} = logic 1 = Vdd/V_{DD}

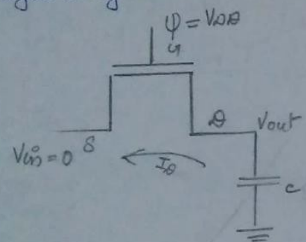
The pass transistor conducts and begins to charge the capacitor towards Vdd, as the o/p voltage approaches $V_{DD} - V_{t,n}$, the NMos transistor turns off because $V_{gs} < V_{t,n}$ the load capacitor will remain charged at ($V_{dd} - V_{t,n}$) this implies that the transmission of logic 1 is degraded as it passes through the NMOS transistor.

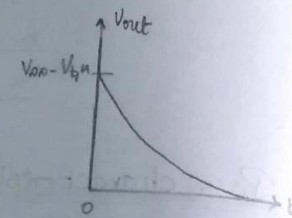




Logic 0 is applied.

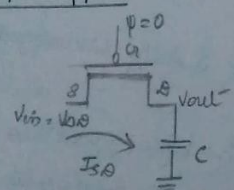
When V_{in} = logic 0 & $\varphi = 1$ the pass transistor conducts. Since the voltage at Vout is greater than V_{in} the current flows from right to left, discharging the load capacitor and the o/p voltage approaches zero. Thus the transmission of logic 0 by NMOS pass transistor is not degraded.

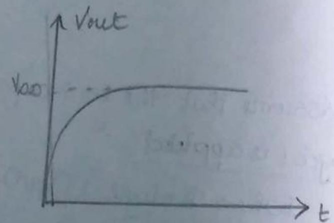


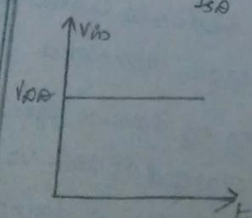


PMOS Transistor as a Switch

Logic 1 is applied

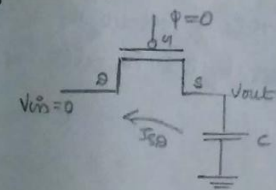


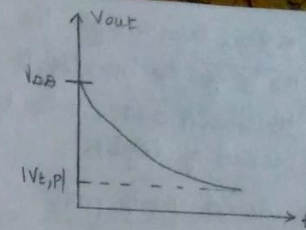




When $V_{in} = V_{DD}$ and $\varphi = 0$, PMOS transistor conducts. Since V_{in} is at a higher potential than Vout the current flows from left to right and charges the capacitor to V_{DD}. Thus the transmission of logic 1 by PMOS transistor is not degraded.

Logic 0 is applied.

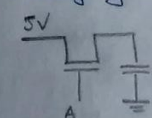




When V_{in} = 0V and φ is = 0 the PMOS transistor turns on. Since the potential at Vout is greater than V_{in} current flows from right to left and the capacitor discharges. As the capacitor discharges the source voltage (Vout) decreases when V_{gs} is less than $|V_{t,P}|$ the PMOS transistor turns off, thus the o/p voltage is equal to $|V_{t,P}|$. ∴ the transmission of logic zero by PMOS transistor is degraded.

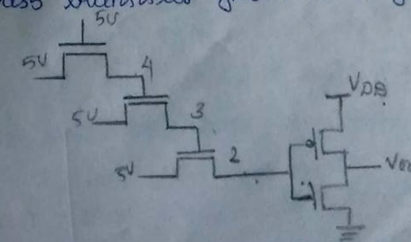
Rules for Designing Pass Transistor Logic (PTL)

1. In designing PTL care must be taken to ensure the existance of both charging and discharging path.



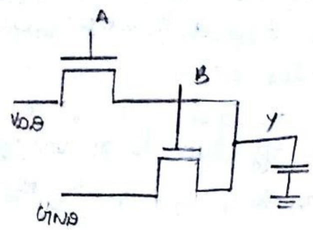
When A=1, the transistor is on and the capacitor charges. When A=0, the transistor is off ideally the o/p node must be at zero volt, but since there is no discharge path the o/p is not pulled low.

2. In designing PTL one must avoid driving the control i/p of a pass transistor from the o/p of another pass transistor.



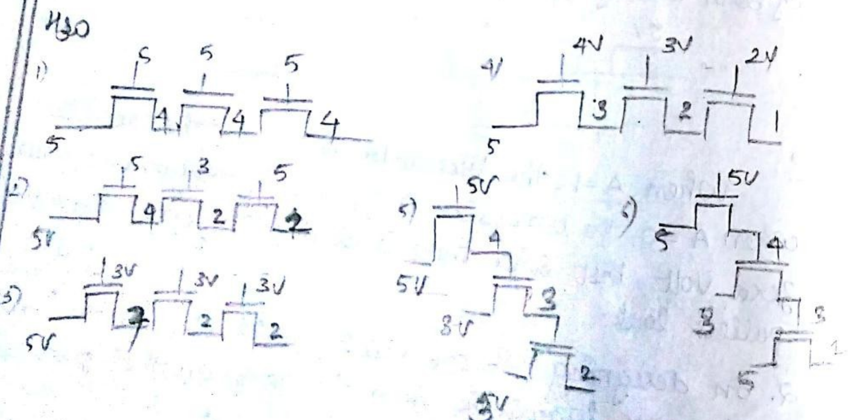
Here eventhough the applied i/p is 5V the Voltage at the i/p of the inverter is 2V this is below the switching threshold and will be treated by the inverter as logic 0 instead of logic 1.

3. Avoid sneak path



A sneak path is created when two pass transistor are both on at the same time and one is connected to VDD while the other is connected to ground. Here the o/p attains some intermediate value between VDD & GND

H.W



2/3/18

Advantages.

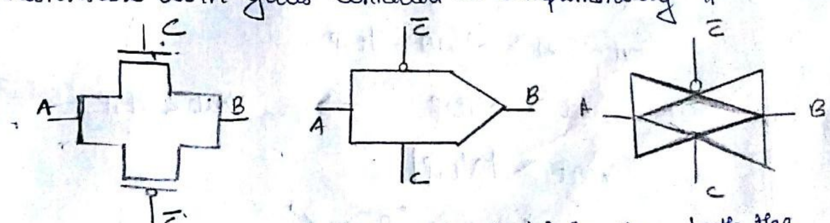
1. Pass transistor logic realisations are ratio less. ie there's no need to have W/L ratio. hence min geometry transistors can be used.
2. More low area to due to smaller no: of transistors. To implement a logic function in pass transistor logic compared to static CMOS logic.
3. Low power dissipation

Disadvantage

1. Not able to transfer the full logic levels properly. ie NMOS transistor has poor transmission of 1. PMOS transistor has poor transmission of 0.
2. Posibility sneak path
3. Higher delay in long chain of pass transistors

Transmission Gates (TG)

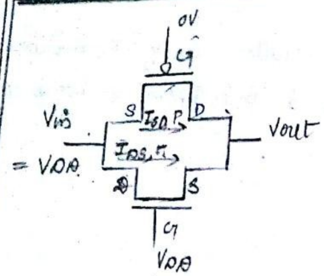
These are used to avoid the weak logic voltages of a single pass transistor. It is a parallel combination of PMOS & NMOS transistors with gates connected to complimentary i/p.



If the control signal c is at logic high then both the transistors are turned on and provide a low resistance current path between nodes A & B. If the control signal is low then both the transistor will be off and the path between node A & B will be open circuited.

~~Pass Transistor~~

Characteristics of Tr



The i/p node A is Connect to a logic high (V_{DD}). The Control signal C=1 . ∴ both the transistors are turned on

For PMOS Transistor

$V_{DS,P} = V_{out} - V_{in}$ Or / $V_{out} - V_{DD}$

$V_{GS,P} = 0 - V_{DD}$

Since gate to source voltage is $-V_{DD}$ PMOS T^r is always on. It can either be in linear or saturation region

when $V_{DS,P} > V_{GS,P} - V_{t,P}$ → 'saturation'

$V_{out} - V_{DD} > -V_{DD} - V_{t,P}$

$V_{out} > -V_{t,P}$ ⇒ $V_{out} < |V_{t,P}|$ (saturation)

$V_{out} > |V_{t,P}|$

When $V_{out} \geq |V_{tp}|$ → 'linear'

For NMOS Transistor

$V_{DS,n} = V_{DD} - V_{out}$

$V_{GS,n} = V_{DD} - V_{out}$

The NMOS transistor is turned on when

$V_{GS,n} > V_{t,n}$

$V_{DD} - V_{out} > V_{t,n}$

$V_{out} < V_{DD} - V_{t,n}$ → ON

Thus the NMOS ~~Tr~~ T^r will be turned off when

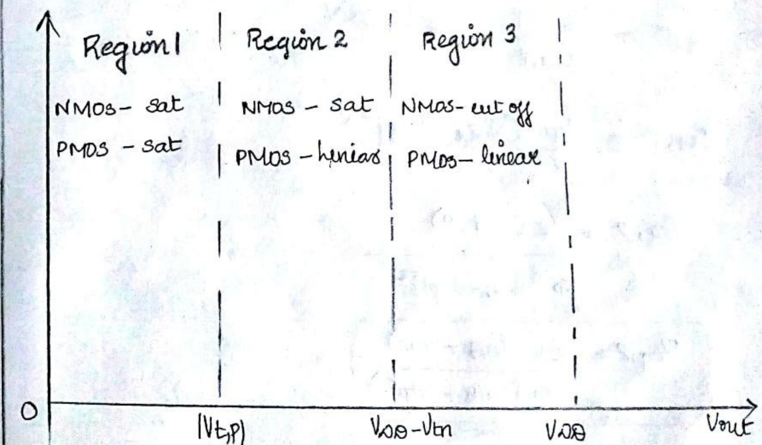
$V_{out} > V_{DD} - V_{t,n}$.

→ NMOS in saturation → $V_{ds,n} > V_{gs,n} - V_{t,n}$

here, $V_{ds,n} = V_{gs,n}$

∴ $V_{ds,n} > V_{gs,n} - V_{t,n}$

hence the NMOS T^r is in the saturation region.



$I_D = I_{D,P}, I_{D,n}$

$$Req,n = \frac{V_{DD} - V_{out}}{I_{DS,n}} = \frac{V_{ds}}{I_{ds}}$$

$$Req,P = \frac{V_{DD} - V_{out}}{I_{SD,P}} = \frac{V_{ds}}{I_{sd}}$$

Total equalent resistance of Tr = $Req_P \,||\, Req_n$.

$$= \frac{V_{ds}}{I_{ds}} \,||\, \frac{V_{ds}}{I_{sd}}$$

3/18

Region I

$$R_{eq,n} = \frac{V_{DD} - V_{out}}{I_{DS,n}}$$

$$I_{D,n} = \frac{k_n}{2}(V_{GS,n} - V_{t,n})^2 \quad (\because \text{NMOS satu})$$

$$= \frac{k_n}{2}(V_{DD} - V_{out} - V_{t,n})^2$$

$$\boxed{R_{eq,n} = \frac{2(V_{DD} - V_{out})}{k_n(V_{DD} - V_{out} - V_{t,n})^2}}$$

$$R_{eq,P} = \frac{V_{sd,P}}{I_{sd,P}}$$

$$I_{sd,P} = \frac{k_p}{2}(V_{SG} - |V_{tp}|)^2$$

$$R_{eq,P} = \frac{2(V_{sd,P})}{k_p(V_{DD} - |V_{tp}|)^2}$$

$$\boxed{R_{eq,P} = \frac{2(V_{DD} - V_{out})}{k_p(V_{DD} - |V_{tp}|)^2}}$$

Region II

$$R_{eq,n} = \frac{V_{DD} - V_{out}}{I_{DS,n}}$$

$$I_{D,n} = \frac{k_n}{2}(V_{GS,n} - V_{t,n})^2 \quad (\because \text{NMOS sat})$$

$$= \frac{k_n}{2}(V_{DD} - V_{out} - V_{t,n})^2$$

$$R_{eq,n} = \frac{2(V_{DD} - V_{out})}{k_n(V_{DD} - V_{out} - V_{t,n})^2}$$

$$I_{D,n} = \frac{k_n}{2}(V_{DD} - V_{out} - V_{t,n})^2$$

$$I_{SD,P} = \frac{k_P}{2}\left(2(V_{SG} - V_{t,P})V_{sd} - V_{sd}^2\right)$$

$$= \frac{k_P}{2}\left(2(V_{DD} - |V_{t,P}|)(V_{DD} - V_{out}) - (V_{DD} - V_{out})^2\right)$$

$$R_{eq,n} = \frac{2(V_{DD} - V_{out})}{k_n(V_{DD} - V_{out} - V_{tn})^2}$$

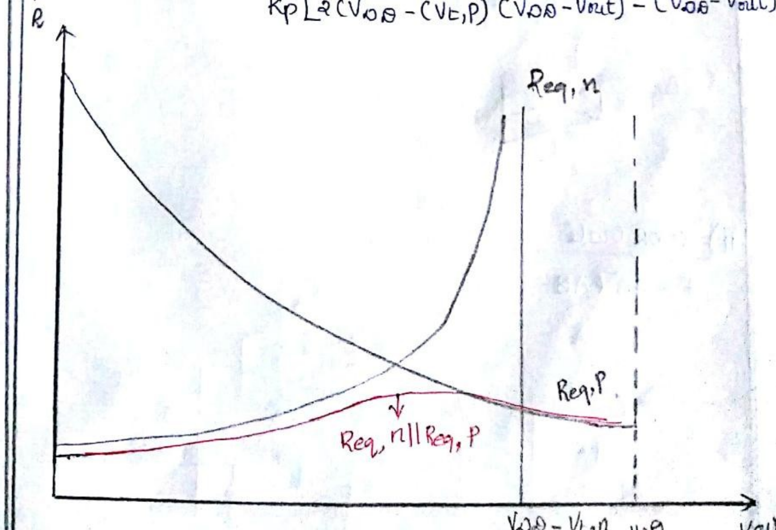
$$R_{eq,P} = \frac{2(V_{DD} - V_{out})}{k_p\left[2(V_{DD} - |V_{t,P}|)(V_{DD} - V_{out}) - (V_{DD} - V_{out})^2\right]}$$

Region III

NMOS - cutoff

PMOS - linear

$$R_{eq,P} = \frac{2(V_{DD} - V_{out})}{k_p\left[2(V_{DD} - |V_{t,P}|)(V_{DD} - V_{out}) - (V_{DD} - V_{out})^2\right]}$$



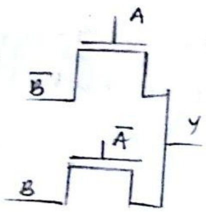
It can be seen that the total equallent resistance of the Trans-mission gate remains relatively constant - ie its value is independent of the o/p Voltage Vo while the individual equivalent resistance of both NMOS & PMOS t^r are strongly dependent on V_{out}

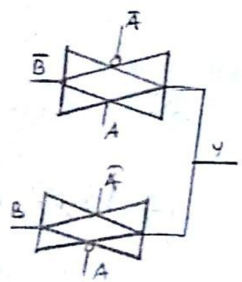
Implementation Of Logic Function Using Transmission Gate

i) XOR Gate

$F = A \oplus B$

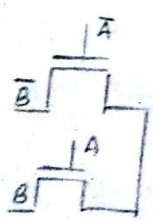
$= \bar{A}B + \bar{B}A$

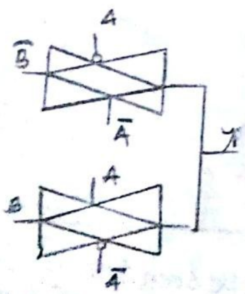




ii) XNOR Gate

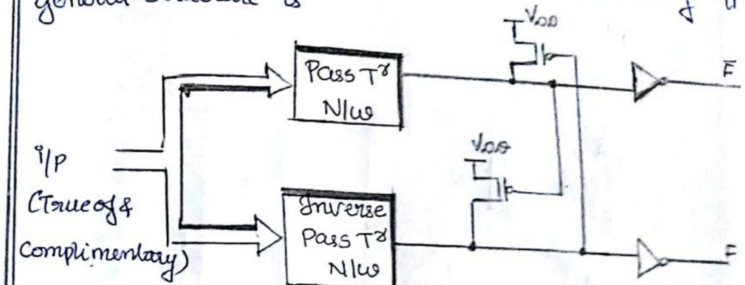
$F = \bar{A}\bar{B} + AB$





Complimentary Pass Transistor Logic

For high performance design CPTL or differential Pass T^r logic (DPTL) is used. It accepts both true and complimentary i/p and produces both true and complimentary o/p. general structure is

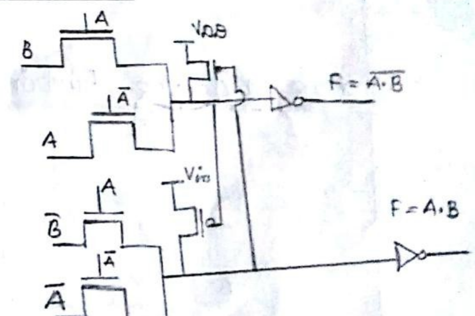


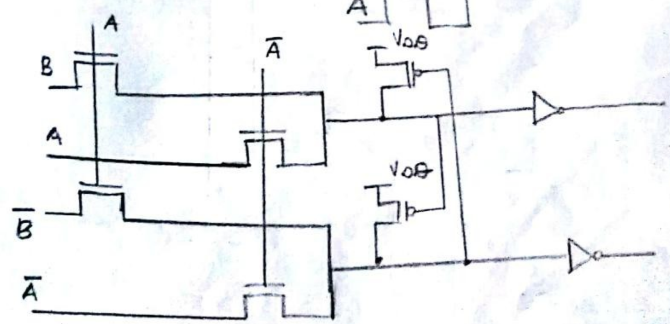
Here the inverting buffer and the weak PMOS T^r provide level restoration.

CPTL For AND & NAND Function:

$F = A \cdot B$ for CPTL no '0'

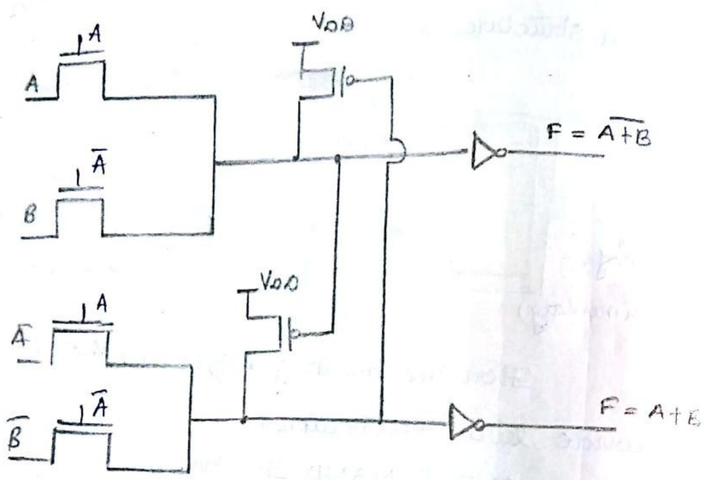
$= A \cdot B + \bar{A} \cdot A$



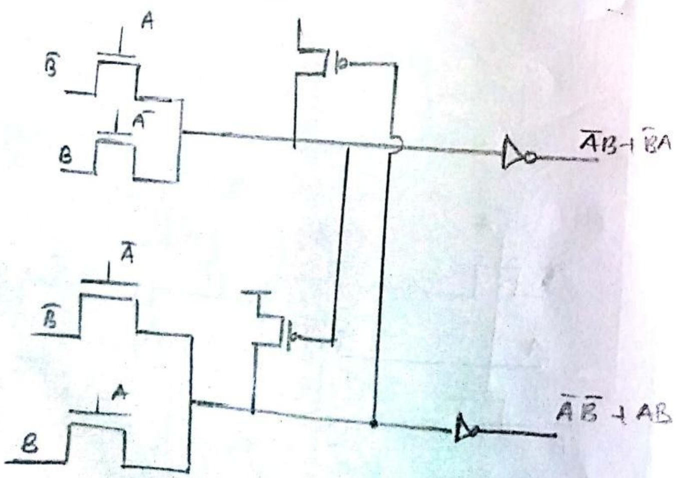


OR & NOR Function

F = A + B



XOR and XNOR Function



4 i/p NAND & AND Gate

CPL gates possess Some Interesting Property

1) Since the circuits are differential, Complimentary i/p data i/p and o/p are always available thus some Complex gates such as XOR and adders can be realised efficiently with smaller no: of transistors

2) CPL belongs to the class of static gates because the o/p defining nodes are always Connected to either V_{DD} or ground through a low resistance path.

3) The design is Very simple Complex gates can be built by cascading standard PT module.

MODULE-5

ROM - READ ONLY MEMORY

Rom are used to store constants, control informations and program instructions in digital system. They provide a fixed, binary output for every binary input. Programs for processors with fixed applications such as washing machines, calculators and game machines, once developed and debugged, need only reading. Fixing the contents at manufacturing time leads to small and fast implementation.

ROM CELLS

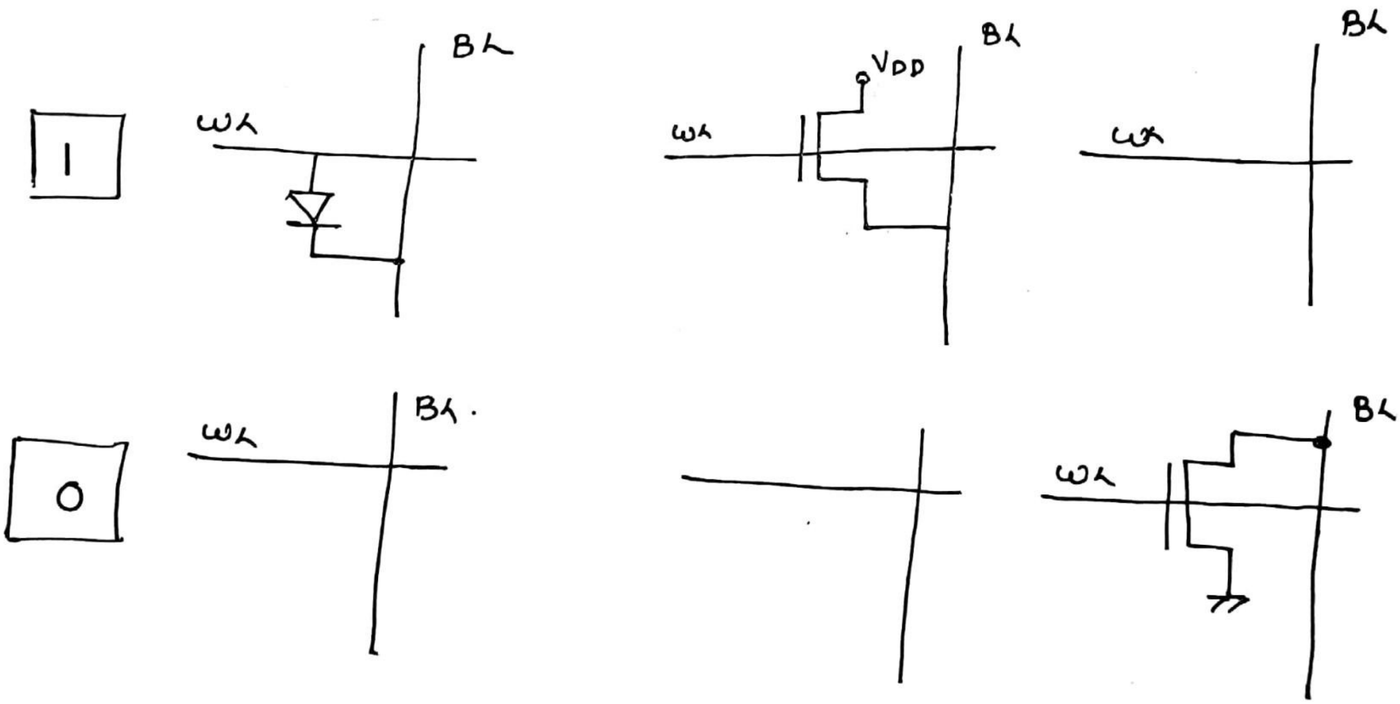


fig 1: Diode ROM. Fig 2: MOS ROM 1. fig3: MOS ROM 2.

The cell should be designed so that a 0 or 1 is presented to the bit line upon activation of its word line.

Diode Rom [fig 1]

1) Assume that BL is connected to ground through a resistor. This is exactly what happens in 0 cell.

→ When WL is high, the diode is enabled and it result in a '1' on the bit line.

Disadv

1) It doesn't isolate the bit line from WL.
2) The current which is required to charge the bit-line capacitance, is provided through the WL.

∴ It is used in small memories.

Note Presence of diode b/w WL and BL – Cell storing '1'
Absence of diode b/w WL and BL – Cell storing '0'

MOS ROM 1 [fig 2]

→ Here diode is replaced by a nMOS transistor, whose drain is connected to the supply voltage.

→ Adv → The output current is provided by MOS transistor in the cell.

→ The operation is identical to that of diode cell.

MOS ROM 2 [fig 3]

→ BL is connected to supply voltage - the output must equal '1'.

→ Absence of transistor b/w WL and BL means a '1' is stored.

→ The '0' cell is realized by providing an mos device between BL and ground.

The two different types of implementation of ROM array are:

1) NOR based ROM array
2) NAND based ROM array.

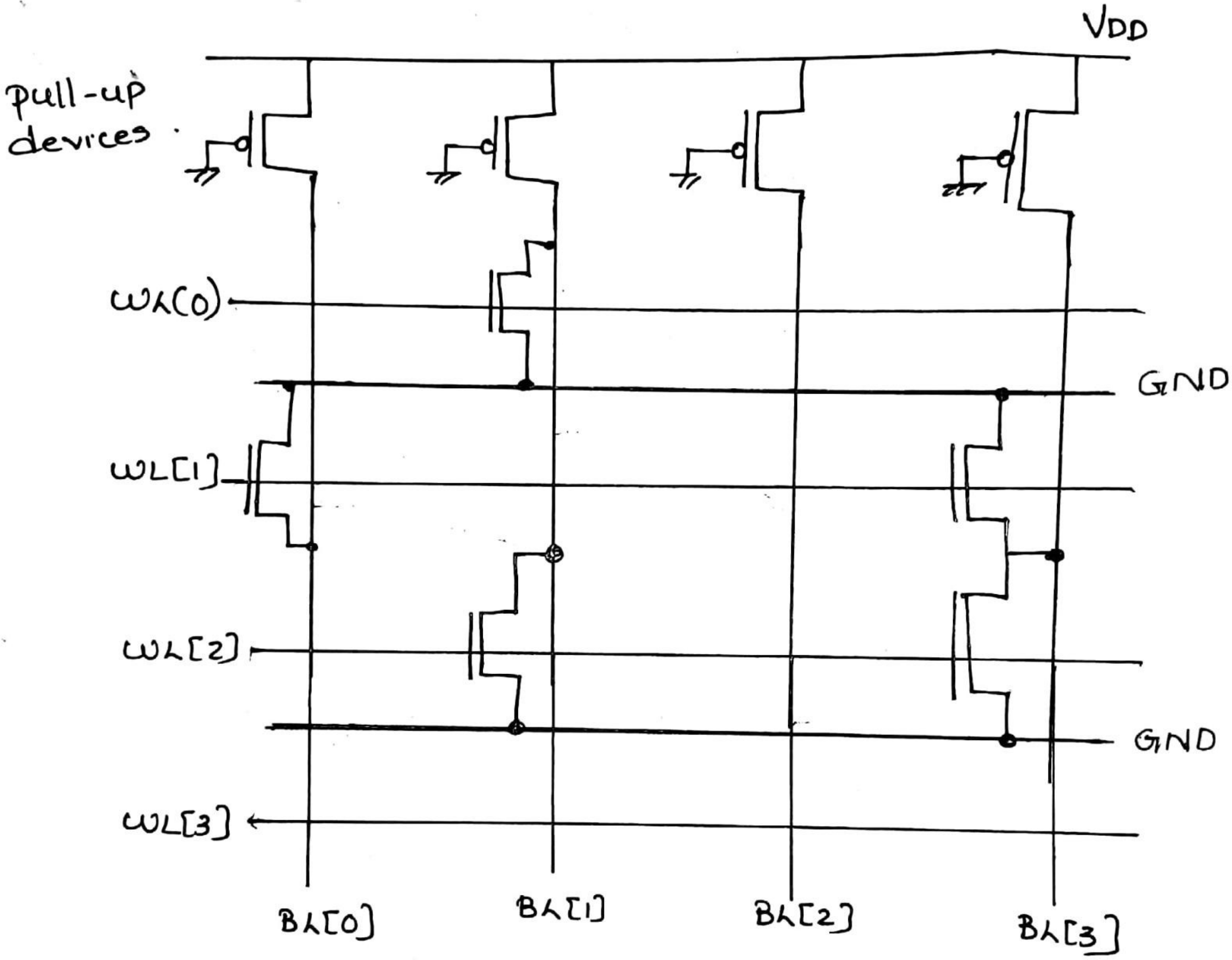
1) NOR based ROM array

Consider 4×4 memory array. Here each column consists of a pseudo-nMOS NOR gate driven by some of the row signals, i.e the word line.

Only one WL is activated at a time by raising its voltage to VDD, while all other rows are held at low voltage level. If an active transistor exists at the crosspoint of a column and the selected row, the column voltage is pulled down to logic low level by that transistor.

If no active transistor exists at the cross point, the column voltage is pulled HIGH by pMOS load device.

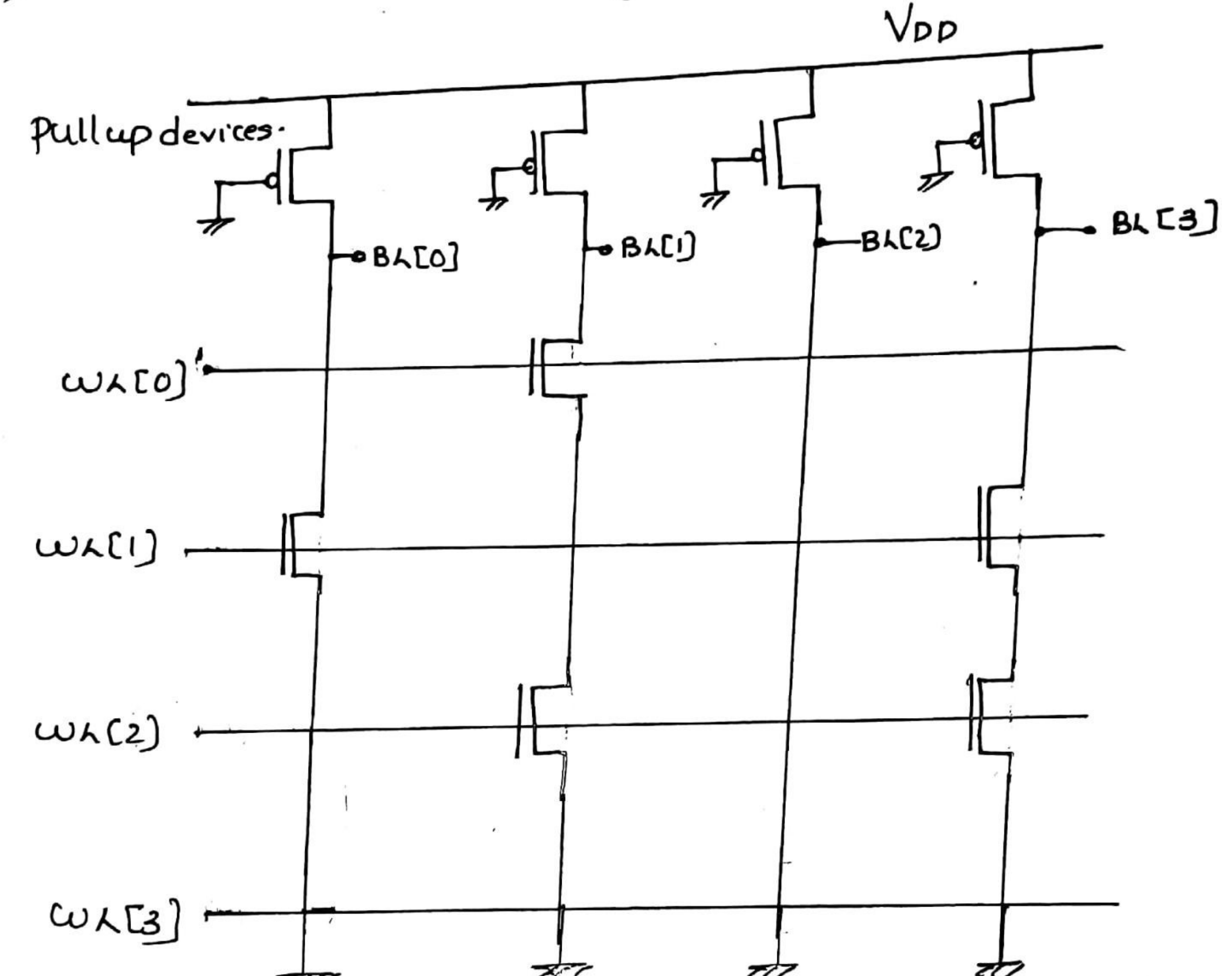
Thus a logic '1' bit is stored as the absence of an active transistor while a logic '0' bit is stored as the presence of an active transistor at the cross point.



Truth Table

WL[0]	WL[1]	WL[2]	WL[3]	BL[0]	BL[1]	BL[2]	BL[3]
1	0	0	0	1	0	1	1
0	1	0	0	0	1	1	0
0	0	1	0	1	0	1	0
0	0	0	1	1	1	1	1

2) NAND Based ROM Array



In normal operation, all WL are logic HIGH voltage level except for the selected line, which is pulled down to logic 0 level. If a transistor exists at the crosspoint of a column and the selected row, that transistor is turned off and the column voltage is pulled HIGH by load device.

If no transistor exist at that cross point, the column voltage is pulled LOW by

A logic '1' bit is stored by the presence of transistor, while '0' is stored at the absense of transistors.

Truth Table

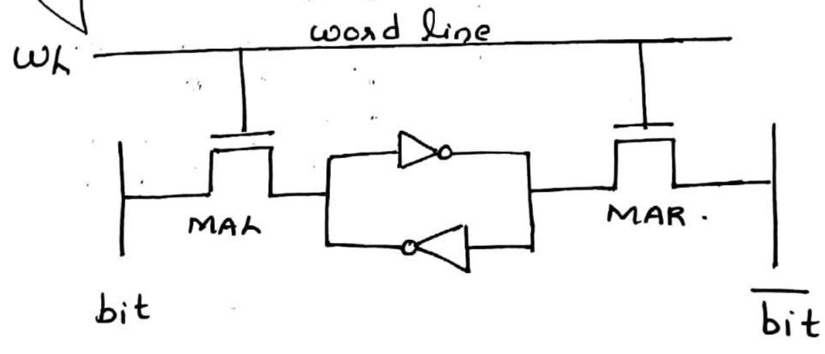
WL[0]	WL[1]	WL[2]	WL[3]	BL[0]	BL[1]	BL[2]	BL[3]
0	1	1	1	0	1	0	0
1	0	1	1	1	0	0	1
1	1	0	1	0	1	0	1
1	1	1	0	0	0	0	0

Static Ram (SRAM)

The acronym RAM stands for Random Access Memory and implies a memory array that allows access to any bit (or group of bits) as needed.

* Memory with both read and write capabilities
* SRAM cells use a simple bistable circuit to hold a data bit. An SRAM cell can hold the stored data bit so long as the power is applied to the circuit.

SRAM have 3 operational modes. When the cell is in a hold state, the value of the bit is stored in the cell for future usage. During a write operation, a logic 0 or 1 is fed to the cell for storage. The value of the stored bit is transmitted to the outside world during a read operation.



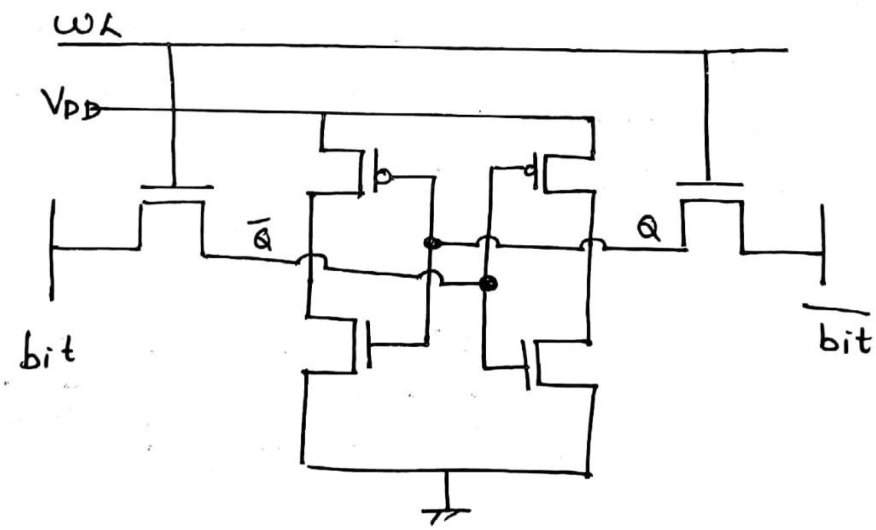
• General SRAM Cell.

A pair of cross-coupled inverters provides the storage, while the two access transistors MAL and MAR provide read and write operations.

The access transistors are controlled by the word line signal WL that defines the operational modes.

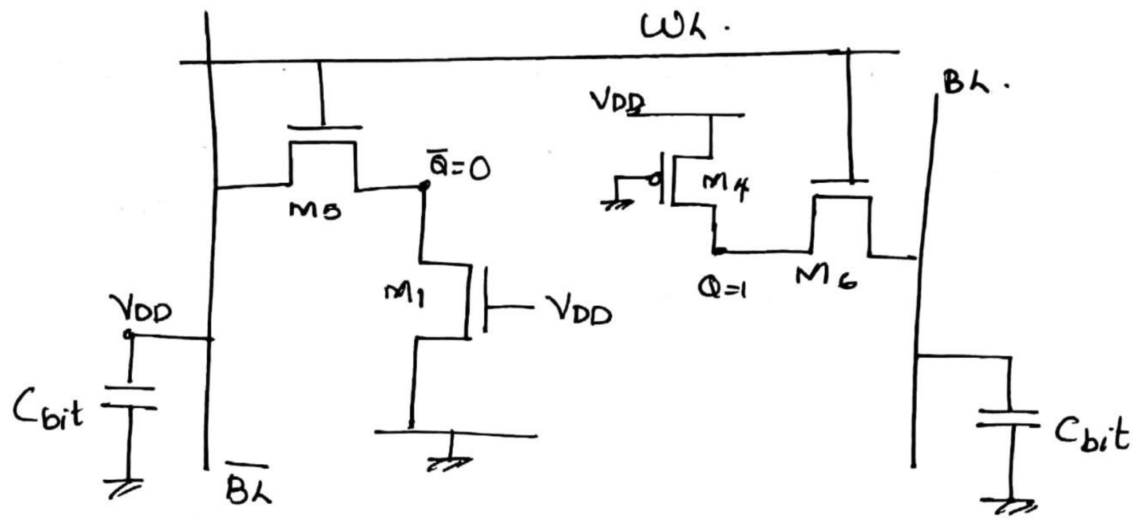
When WL=0, both access FETs are OFF and the cell is isolated. This defines the hold condition. To perform a read or write operation, the word line is brought to a value of WL=1. This turns ON the access transistors connecting the dual-rail data lines bit and $\overline{\text{bit}}$ to the outside circuitry.

A write operation is performed by placing voltages on the bit and $\overline{\text{bit}}$ lines which then act as inputs. For a read operation, the bit and $\overline{\text{bit}}$ lines acts as outputs.



• 6T cell CMOS SRAM.

CMOS SRAM READ OPERATION



Assume that a '1' is stored at Q. We further assume that both bit lines are precharged to VDD before the read operation is initiated. The read cycle is started by asserting the word line, enabling both pass transistors M5 and M6 after the initial word line delay.

During the correct read operation, the values stored in Q and $\overline{Q}$ are transferred to the bit lines by leaving BL at its precharge value and by discharging $\overline{BL}$ through M1-M5.

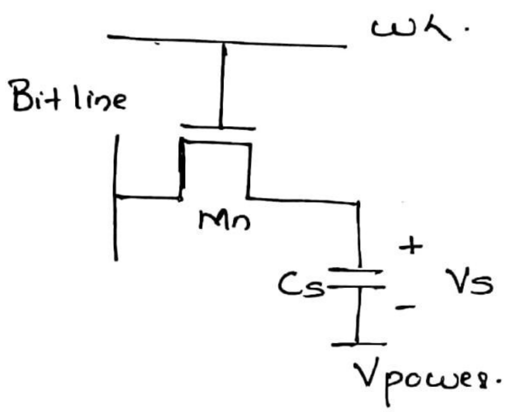
WRITE OPERATION

Assume that a '1' is stored in the cell (Q=1). A '0' is written in the cell by setting $\overline{BL}$ to 1 and BL to '0'. This causes the flipflop to change state. It is assumed that the gates of M_1 and M_4 are at VDD and GND respectively.

DYNAMIC RAM (DRAM)

DRAM are smaller than SRAM cells, which leads to high density storage arrays. DRAMs are slower than SRAM and require more peripheral circuitry.

1T DRAM [1 Transistor DRAM].



It consist of a nFET Mn and a storage capacitor Cs. The cell is controlled by the word line signal WL and BL provides I/O path to the cell. The bottom of capacitor is connected to one of the power supply rails, and is denoted as Vpower in the figure, either VDD or Vss can be used.

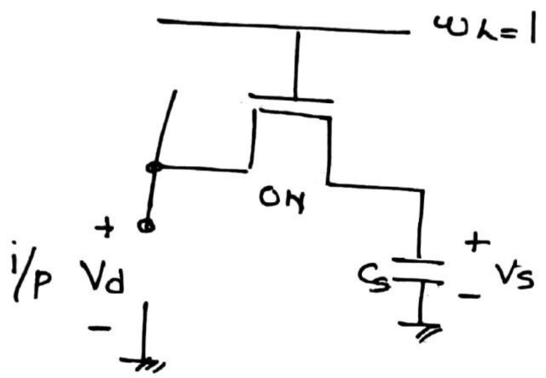
A voltage Vs across the capacitor corresponds to a stored charge Qs of.

$$Q_s = C_s \cdot V_s.$$

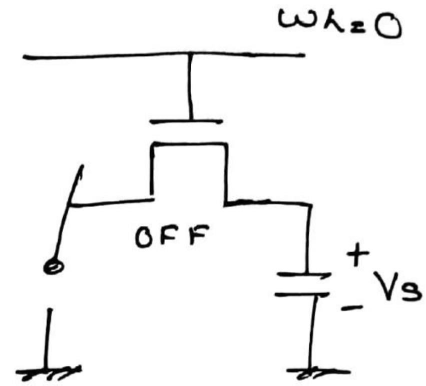
When $V_s = 0$, $Q_s = 0$ => charge state is logic 0

When $V_s > 0$, $Q_s > 0$ => charge state is logic 1.

Write Operation



• Write Operation.



• Hold Operation.

$V_{power} = V_{ss} = 0V$, and apply $WL = 1$, turn on nmos transistor and allows access to the storage capacitor. The input voltage V_d controls the current to / from C_s.

When $V_d = 0$ => results in $V_s = 0V$ across capacitor. [logic 0] ∴ $Q_s = 0$.

When $V_d = V_{DD}$ => results in $V_s = V_{DD} - V_{th}$, which [logic 1] give a charge of $Q_{max} = C_s(V_{DD} - V_{th})$.

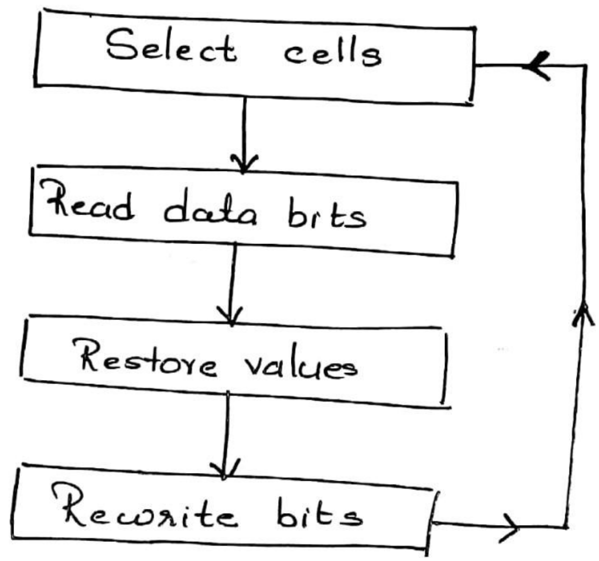
Hold State

It is achieved by turning OFF the transistor by providing $WL = 0$.

Hold time, th, is defined as the longest period of time that the cell can maintain a voltage [logic 1]. The hold time is also called as retention time.

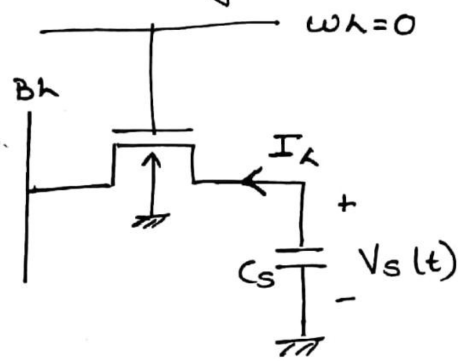
To overcome the charge leakage problem, DRAM array employs a refresh operation, where the data is periodically read from every cell, amplified and then rewritten.

The refresh frequency, $f_{refresh} = \frac{1}{2t_h}$



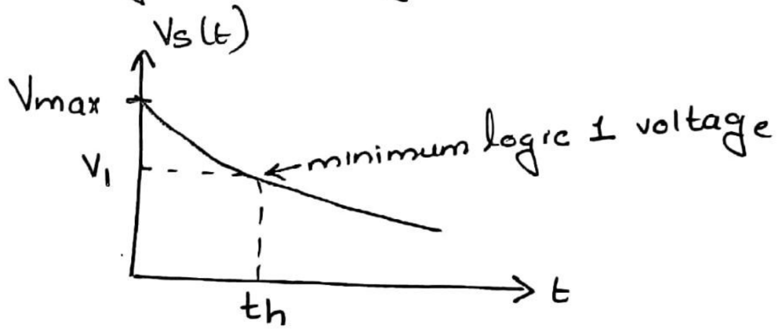
CHARGE LEAKAGE IN DRAM

A logic 1 voltage $V_S = V_{max}$ on the storage capacitor provides the electromotive force for the leakage current I_L, flowing away from C_S.



$$I_L = -\left(\frac{dQ}{dt}\right) = -C_S\left(\frac{dV_S}{dt}\right)$$

The initial voltage $V_S = V_{max}$ gives the voltage decay. The minimum logic '1' voltage is V_1.



To find t_h, assume I_L is a constant, then

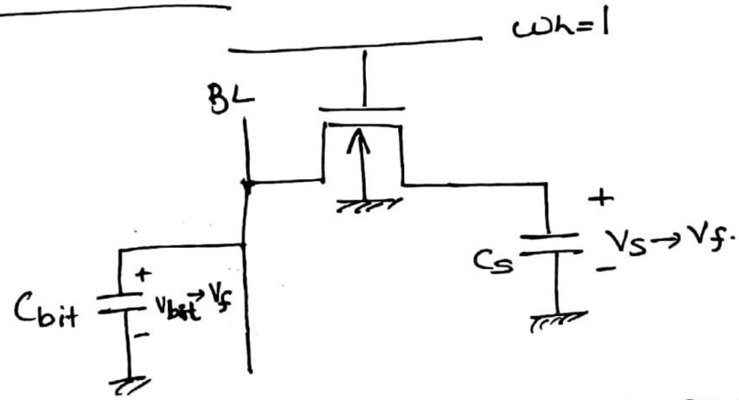
$$I_L = -C_S \cdot \left(\frac{\Delta V_S}{\Delta t}\right)$$

hold time, $t_h = |\Delta t| = \left(\frac{C_S}{I_L}\right) \cdot (\Delta V_S)$

The hold time may be increased by using a large capacitance and minimizing the leakage current.

Memory units must be able to hold data as long as power is applied. To overcome the charge leakage problem, DRAM array employ a refresh operation.

Read Operation



The voltage Vs on the capacitor at the read time provides the voltage to move charge from Cs to the Cbit (bit line capacitance), which set up a charge sharing situation.

Cbit → line capacitance and other parasitic capacitance.

Initial charge on capacitor, $Q_S = C_S \cdot V_S$, where

$V_S = 0 \Rightarrow$ logic 0

$V_S > 0 \Rightarrow$ logic 1

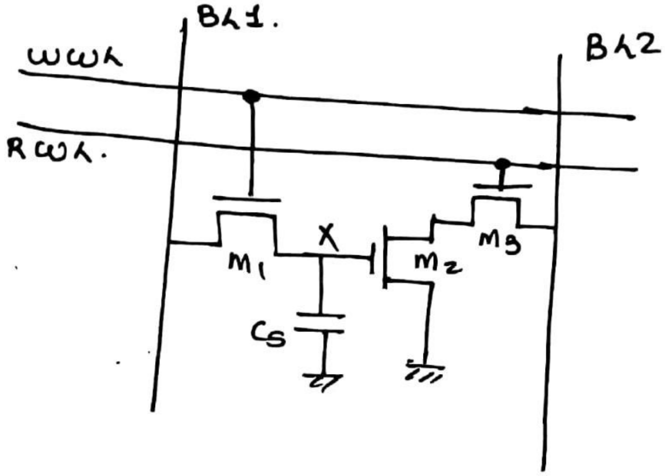
Current flow from Cs to Cbit continues until the voltages are equal to the final voltage $V_f = V_{bit} = V_S$.

The charge is redistributed according to,

$$Q_S = C_S \cdot V_f + C_{bit} V_f = V_f (C_S + C_{bit})$$

$$\therefore V_f = \frac{Q_S}{C_S + C_{bit}} = \frac{C_S \cdot V_S}{(C_S + C_{bit})}$$

3T DRAM [3 Transistor DRAM].

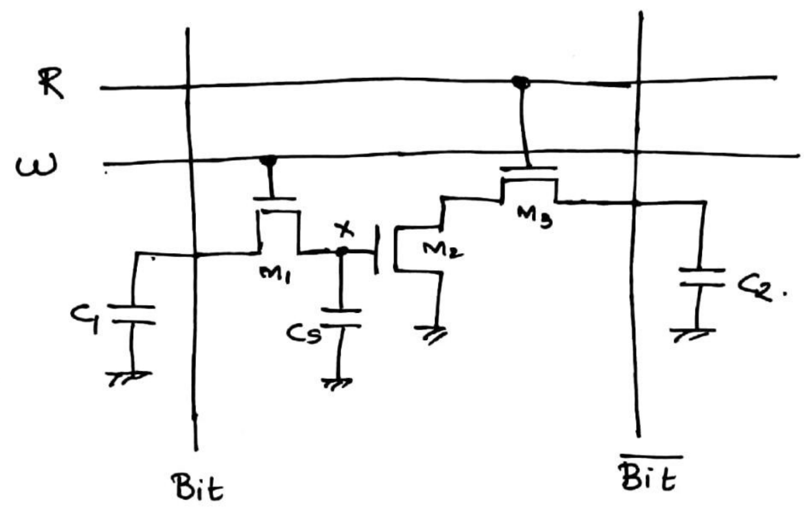


The cell is written by placing appropriate data value on BL1 and WWL = 1. The data is retained as charge on C_s once WWL is lowered. When reading the cell, RWL = 1. The transistor M_2 [storage transistor] is either ON or OFF depending upon stored value. The BL2 is either clamped to VDD with the help of load device or precharged to either VDD or $V_{DD} - V_t$. The precharge approach is generally used.

BL2 = 0 => when '1' is stored

BL2 = 1 => when '0' is stored.

3T DRAM [3 Transistor DRAM]



Bit and $\overline{bit}$ which can act as input lines or output lines. Read (R) and Write (ω) lines, they are just like word line. C_1 and C_2 are the precharge capacitor. Here the values are written into and read from X. Cs is the store capacitor.

Write Operation

1) ω = 1, R = 0
2) When ω = 1, then M_1 is ON, R = 0, so M_2 and M_3 are OFF
3) Bit line act as input and C_1 is in precharge condition.
4) Bit = 1, $X = V_{DD} - V_{tn}$, After V_{tn}, M_1 will be ON.
5) Bit = 0, X = 0
6) If bit = 1, then $\overline{bit} = 0$ and bit = 0, then $\overline{bit} = 1$

After write operation, give ω = 0.

Read Operation

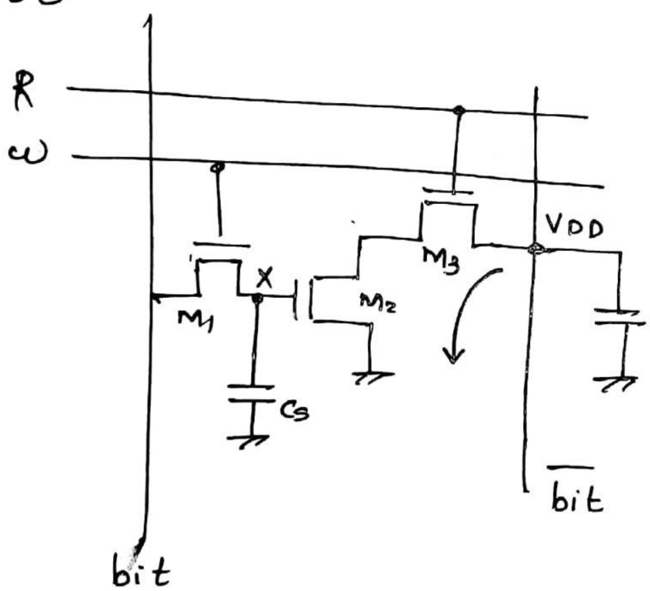
1) R = 1 , w = 0

2) M3 is ON , M_1 is OFF

3) C_2 is precharged .

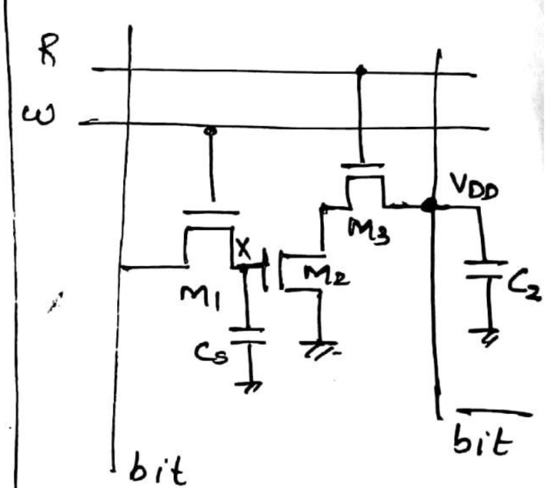
4) When $X = V_{DD} - V_{th}$, then M_2 is ON , M_3 is also ON . The capacitor C_2 will discharge and $\overline{bit} = 0$, sense amplifier output is 1.

5) When X = 0 , then M_2 is OFF , C_2 has no way to discharge and $\overline{bit} = 1$. The output of sense amplifier is 0.



→ R = 1, w = 0 , $X = V_{DD} - V_{tn}$.

C_2 discharge., $\overline{bit} = 0$



→ R = 1 , w = 0 , X = 0

C_2 has no way to discharge , $\overline{bit} = 1$.

PROBLEMS

1. The storage capacitor in a DRAM has a value of $C_S = 55fF$. The circuits restricts the capacitor voltage to a value of $V_{max} = 3.5V$. When the access transistor is OFF, the leakage current off of the cell is 75nA.

a) How many electrons can be stored on C_S?

b) How many fundamental charge units 'q' leave the cell in 1 second due to leakage current?

c) Calculate the time needed to reduce the number of stored charges to 100.

Given $C_S = 55fF$

$V_{max} = 3.5V$

$I_L = 75nA$.

a) The maxm charge that can be stored on the capacitor,

$$Q_{max} = C_S \cdot V_{max}$$

$$= 55\times10^{-15} \times 3.5$$

$$192.5\times10^{-15} = 1.925\times10^{-15}\ C$$

$$\text{No: of charges, } N = \frac{Q_{max}}{q} = \frac{1.925\times10^{-5}}{1.602\times10^{-19}}$$

$$= 1.20.2\times10^{4}$$

b) Leakage current = 75nA.

$$I_L = \frac{\Delta Q}{\Delta t}\ C/s$$

$\therefore$ Rate of $\bar{e}s$ leakage in $1s$ $= \dfrac{75 \times 10^{-9}}{1.602 \times 10^{-19}} =$

$$\gamma_N = 46.81 \times 10^{10} \text{ electrons/sec.}$$

c) Time required for the stored charge to be reduced to 100,

$$\Delta t = \frac{N - 100}{\gamma_N} = \frac{120.2 \times 10^{4} - 100}{46.81 \times 10^{10}}$$

$$= 2.567 \times 10^{-6} s$$

$$= 2.567 \mu s$$

2. A DRAM cell has a storage capacitance of $C_S = 45 fF$. It is used in a system where $V_{DD} = 3.3V$ and $V_{tn} = 0.55V$. The bit line capacitance is $C_{bit} = 250 fF$.

a) Find the maximum amount of charge that can be stored on C_S.

b) Suppose that the voltage on the capacitor is charged a level of V_{max}, The word line controlling the access FET is dropped to a value $WL = 0$, at time $t = 0$. The leakage current is estimated to be $50nA$. To detect a logic 1 state, the voltage on the bit line must be atleast $1.5V$. Find the hold time.

Given,

$C_S = 45fF = 45 \times 10^{-15} F$

$V_{DD} = 3.3V$

$V_{tn} = 0.55V$.

$C_{bit} = 250fF = 250 \times 10^{-15} F$.

a) The maximum stored charge, $Q_{max} = C_S \cdot V_{max}$.

$= C_S \, (V_{DD} - V_{tn})$

$= 45 \times 10^{-15} \times (3.3 - 0.55)$

$= \underline{1.24 \times 10^{-13} C}$

b) $V_f = 1.5V$

$$V_f = \left(\frac{C_S}{C_S + C_{bit}}\right) \cdot V_S.$$

$$1.5 = \frac{45 \times 10^{-15}}{(45 + 250) \times 10^{-15}} \cdot V_S = \left(\frac{45}{295}\right) \cdot V_S.$$

$V_S \times 0.1525 = 1.5$

$\therefore V_S = \underline{9.83V}$.

$V_S \rightarrow$ value is too high, the cell will not work.

Let us assume $V_f = 0.2V$,

$V_S \times 0.1525 = 0.2$

$\underline{V_S = 1.31V}$, which is possible.

Hold time, $t_h = \left(\frac{Q_s}{I_L}\right) \cdot V_s$

$$= \frac{45 \times 10^{-15}}{50 \times 10^{-9}} \times 1.31$$

$$= 1.17 \mu s$$

SENSE AMPLIFIERS

They perform the following functions:

1) Amplification - In certain memory structures such as the 1T DRAM, amplification is required for proper functionality.

2) Delay reduction - The amplifier compensates for the restricted fan-out driving capability of memory cell by accelerating the bit line transition.

3) Power reduction - Reducing the s/l swing on the bit lines can eliminate a substantial part of power dissipation.

4) Signal restoration - Because the read & refresh functions are intrinsically linked in 1T DRAMs, it is necessary to drive the bit lines to the full signal range after sensing.

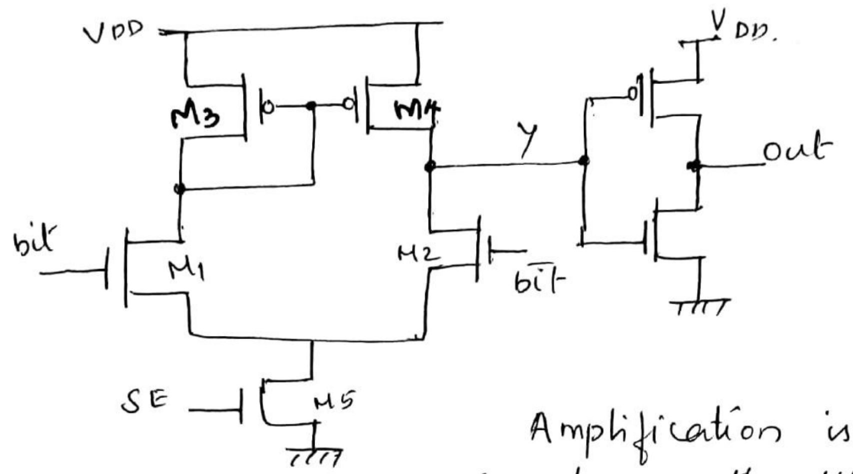
- Sense amplifiers are analog ckts by nature.

DIFFERENTIAL VOLTAGE SENSING AMPLIFIERS

A differential amplifier takes small-signal differential i/ps, and amplifies them to a large signal single ended o/p. It is generally known that a differential approach presents numerous advantages over its single ended counterpart — one of the most important being the common-mode rejection. Such an amplifier rejects noise that is equally injected to both i/ps.

The effectiveness of a differential amplifier is characterized by its ability to reject the common noise and amplify the true difference b/w the signals. The signals common to both inputs are suppressed at the output of the amplifier by a ratio called the common mode rejection ratio (CMRR). Spikes on the power supply are suppressed by a ratio called power-supply rejection ratio (PSRR).

Differential sensing is therefore considered the technique of choice. Unfortunately, the differential approach is only directly applicable to SRAM memories, since these are only the memory cells that offer a true differential o/p.



Amplification is accomplished with a single stage, based on the current mirroring concept. The i/p s/ls (bit & bit) are heavily loaded and driven by the SRAM memory cell. The i/ps are fed to the differential i/p devices (M_1 & M_2), and transistors M_3 & M_4 act as an active mirror load. The amplifier is conditioned by the sense amplifier enable s/l, SE.

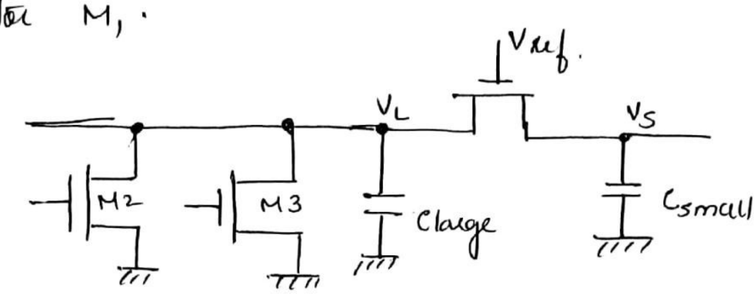
Initially, the i/ps are precharged and equalized to a common value, while SE is low disabling the sensing ckt. Once the read operation is initiated, one of the, bit lines drops. SE is enabled when a sufficient differential s/l has been established, and the amplifier evaluates. The gain of differential-to-single ended amplr is given by

$$A_{sense} = -g_{m1}(r_{o2} \parallel r_{o4})$$

SINGE-ENDED SENSING

While differential sensing is by far the preferred approach, memory cells used in ROMs, E(E)PROMs are DRAMs are inherently single ended. Since the bit lines are typically precharged, an assymmetrically biased inverter is used. An interesting variant, called the charge redistribution amplifier is often used in small mly structures.

The basic idea is to exploit the imbalance b/w a large capacitance C_{large} and a much smaller component C_{small}. The 2 capacitors are isolated by the pass transistor M_1.



charge redistribution amplr.

The initial voltages on nodes L & S (V_{L0} & V_{S0}) are precharged to $V_{ref} - V_{Tn}$ and V_{DD} by connecting nodes S to the supply voltage. Because of the voltage drop over M_1, V_L only precharges to $V_{ref} - V_{Tn}$. When one of the pull-down devices turn ON, node L with its large capacitance slowly discharges. As long as $V_L \geq V_{ref} - V_{Tn}$ transistor M_1 turns ON. A charge redistribution is initiated, and nodes L & S equalize. The resulting s/l can be fed into an inverter with a switching threshold larger than $V_{ref} - V_{Tn}$ to produce a rail-to-rail swing.

RELIABILITY AND TESTING OF VLSI CIRCUITS

- VLSI testing deals with techniques that are used to determine if a die behaves properly after the fabrication sequence is completed.
- Reliability is concerned with projecting the lifetime of a component once it is placed into operation.

General Concepts

For wafer testing - a test probe head allows electrical contact to the I/O points of a die. Several sets of stimuli are applied to i/ps

PROGRAMMABLE LOGIC DEVICES.

A programmable logic devices is an IC that is user configurable and is capable of implementing logic functions. It is an Large Scale Integration (LSI chip) chip that contain a 'regular' structure and allows the designer to customize it for any specific application, i.e it is programmed by the user to perform a function required for the application.

A PLD contains a large no: of gates, flip-flops and registers that are interconnected on the chip. PLD can be reprogrammed in a few seconds and hence give more flexibility for designing.

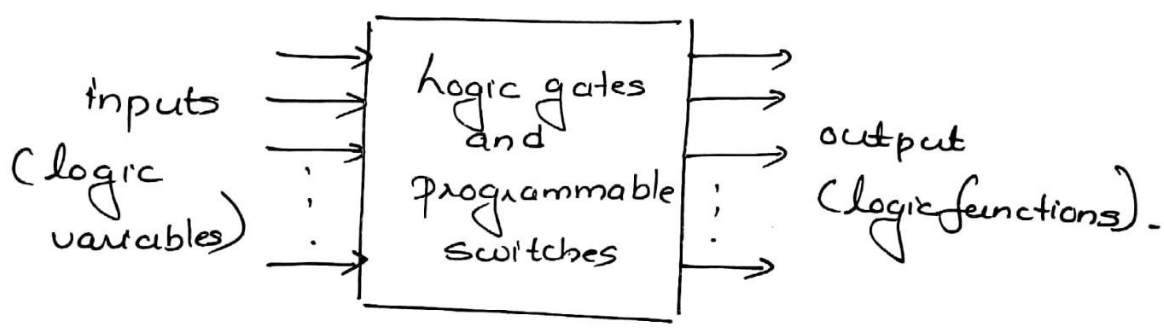
Advantages.

1. Less power requirement.
2. Less space requirement.
3. High design security
4. Easy design modification
5. High switching speed.
6. High reliability

The main types of programmable logic devices are

- ROM
- CPLD
- FPGA
- PLA
- PAL.

Programmable logic device as a black box.

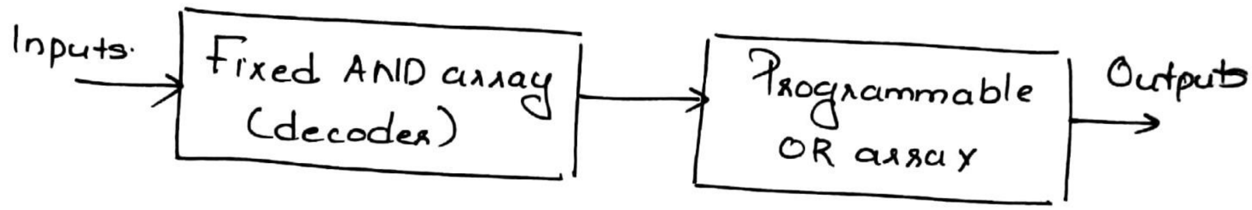


1) CPLD [Combinational Programmable Logic Devices].

A CPLD is an IC with programmable gates divided into an AND array and an OR array to provide AND-OR sum of products (SOP) implementation. There are 3 major types of combinational PLDs and they differ in the placement of the programmable connection in the AND-OR array.

a) PROM (Programmable read only memories)
b) PAL (Programmable Array Logics)
c) PLA (Programmable Logic Arrays).

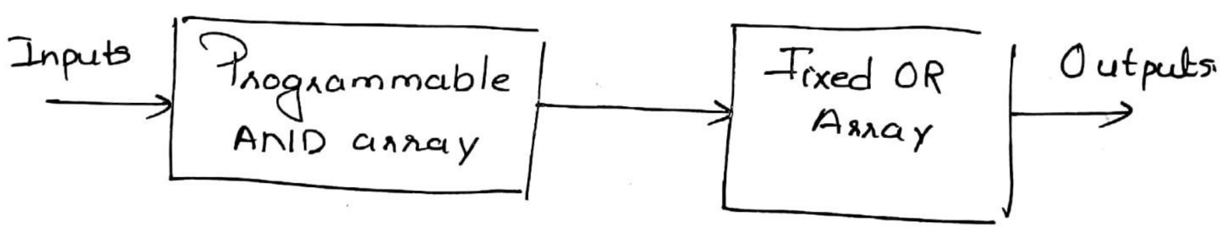
a) PROM



PROM has a fixed AND array constructed as a decoder and a programmable OR array.

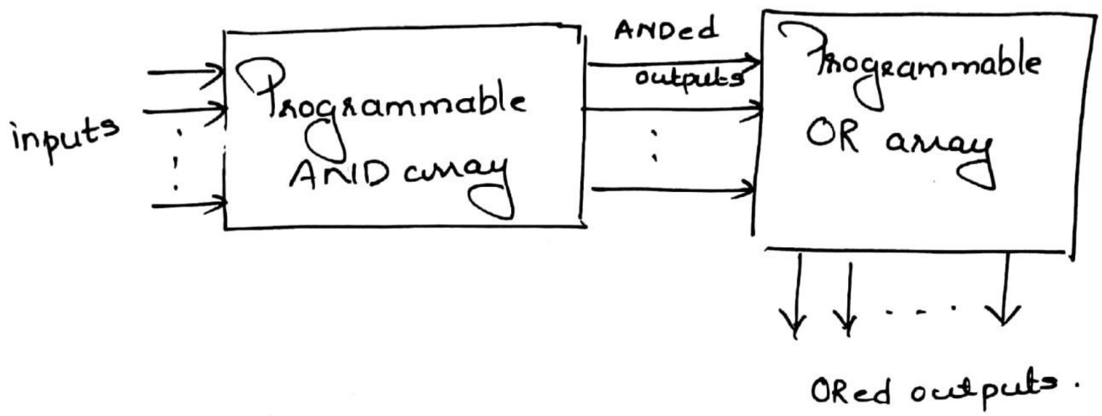
b) PAL [Programmable Array Logic].

PAL has programmable AND array and a fixed OR array.



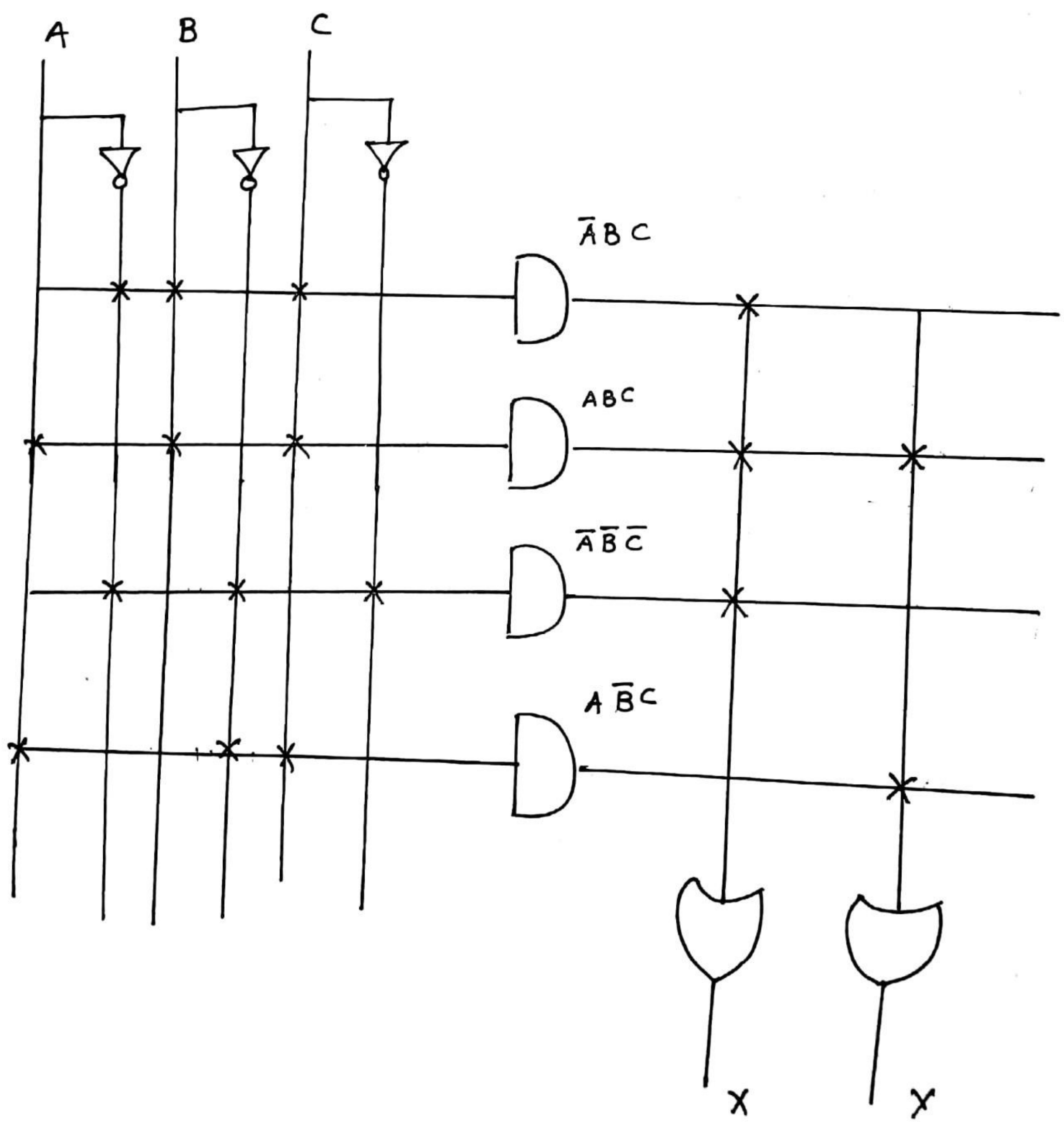
☆ c) PLA [Programmable Logic Array].

The most flexible PLD is the PLA, where both the AND and OR arrays can be programmed. The product terms in the AND array may be shared by any OR gate to provide the required SOP implementation.



Design a PLA to realize the function.

eg: $X = \bar{A}BC + ABC + \bar{A}\bar{B}\bar{C}$ / $Y = ABC + A\bar{B}C$



FIELD PROGRAMMABLE GATE ARRAYS (FPGA)

FPGA provide the next generation in the programmable logic devices. The word Field in the name refers to the ability of the gate arrays to be programmed for a specific function by the user instead of by the manufacturer of the device. The word Array is used to indicate a series of columns and rows of gates that can be programmed by the end user.

As compared to standard gate arrays, the field programmable gate arrays are larger devices. The basic cell structure for FPGA is somewhat complicated than the basic cell structure of standard gate array. The programmable logic blocks of FPGA are called Configurable Logic Block (CLB). Logic blocks can be programmed to perform the function of basic logic gates such as AND, and XOR, or more complex combinational functions such as decoders or mathematical functions. In most FPGAs, the logic blocks also include memory elements, which may be simple flip-flops or more complete blocks of memory.

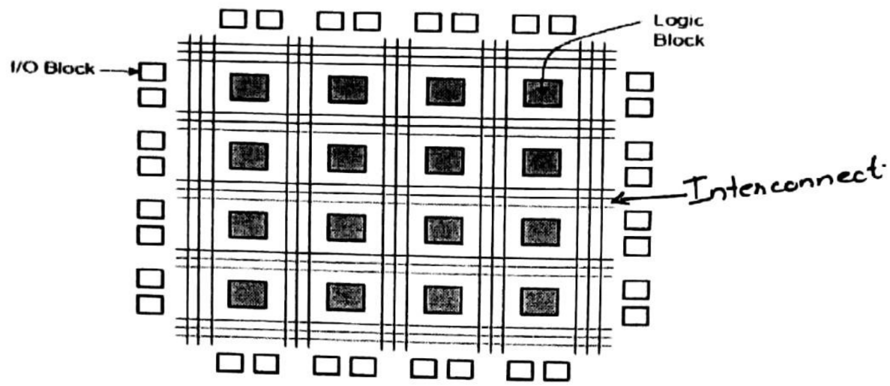
The FPGA architecture consists of three types of configurable elements.
(i)IOBs – a perimeter of input/output blocks
(ii)CLBs- a core array of configurable logic blocks
(iii)Resources for interconnection

The IOBs provide a programmable interface between the internal; array of logic blocks (CLBs) and the device's external package pins. CLBs perform user-specified logic functions, and the interconnect resources carry signals among the blocks.

A configurable program stored in internal static memory cells determines the logic functions and the interconnections. The configurable data is loaded into the device during power-up reprogramming function. FPGA devices are customized by loading configuration data into internal memory cells. The FPGA device can either actively read its configuration data out of an external serial or byte-wide parallel PROM (master modes), or the configuration data can be written to the FPGA devices (slave and peripheral modes).

Architecture of FPGA:

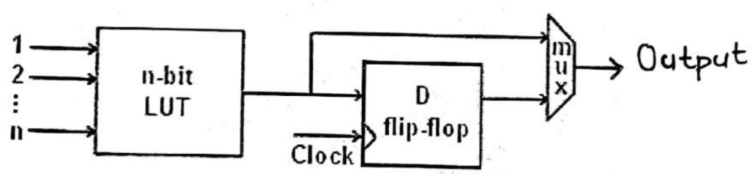
The fig .1 shows the general structure of FPGA chip. It consists of a large number of programmable logic blocks surrounded by programmable I/O block. The programmable logic blocks of FPGA are smaller and less capable than a PLD, but an FPGA chip contains a lot more logic blocks to make it more capable. As shown in fig. the logic blocks are distributed across the entire chip. These logic blocks can be interconnected with programmable inter connections.



The programmable logic blocks in the Xilinx family of FPGAs are called Configurable Logic Blocks(CLBs).The Xilinx architecture uses, CLBs, I/O blocks switch matrix and an external memory chip to realize a logic function. It uses external memory to store the interconnection information. Therefore, the device can be reprogrammed by simply changing the configuration data stored in the memory.

CLB (Configurable Logic Blocks):

The CLB consists of an n-bit look-up table (LUT), a flip-flop and a 2x1 mux. The value n is manufacturer specific. Increase in n value can increase the performance of a FPGA. Each CLB has n-inputs and only one input, which can be either the registered or the unregistered LUT output. The output is selected using a 2x1 mux. The LUT output is registered using the flip-flop (generally D flip-flop). The clock is given to the flip-flop, using which the output is registered.



FPGAProgramming

The design is first coded in HDL (Verilog or VHDL), once the code is validated (simulated and synthesized). During synthesis, typically done using tools like Xilinx ISE and a technology-mapped net list is generated. The net list can then be fitted to the actual FPGA architecture using a process called place-and-route, usually performed by the FPGA Company's proprietary place-and-route software. The user will validate the map, place and route results via timing analysis, simulation, and other verification methodologies. Once the design and validation process is complete, the binary file generated is used to (re)configure the FPGA. Once the FPGA is (re)configured, it is tested. If there are any issues or modifications, the original HDL code will be modified and then entire process is repeated, and FPGA is reconfigured.

UNIVERSITY QUESTIONS SOLVED

1. Design an AND-OR PLA with outputs,

$F_1 = m_1 + m_6$

$F_2 = m_0 + m_5 + m_6 + m_7$

$F_3 = m_3 + m_4 + m_7$.

A	B	C	
0	0	0	$m_0 = \bar{A}\bar{B}\bar{C}$
0	0	1	$m_1 = \bar{A}\bar{B}C$
0	1	0	$m_2 = \bar{A}B\bar{C}$
0	1	1	$m_3 = \bar{A}BC$
1	0	0	$m_4 = A\bar{B}\bar{C}$
1	0	1	$m_5 = A\bar{B}C$
1	1	0	$m_6 = AB\bar{C}$
1	1	1	$m_7 = ABC$.

Given, $F_1 = m_1 + m_6$

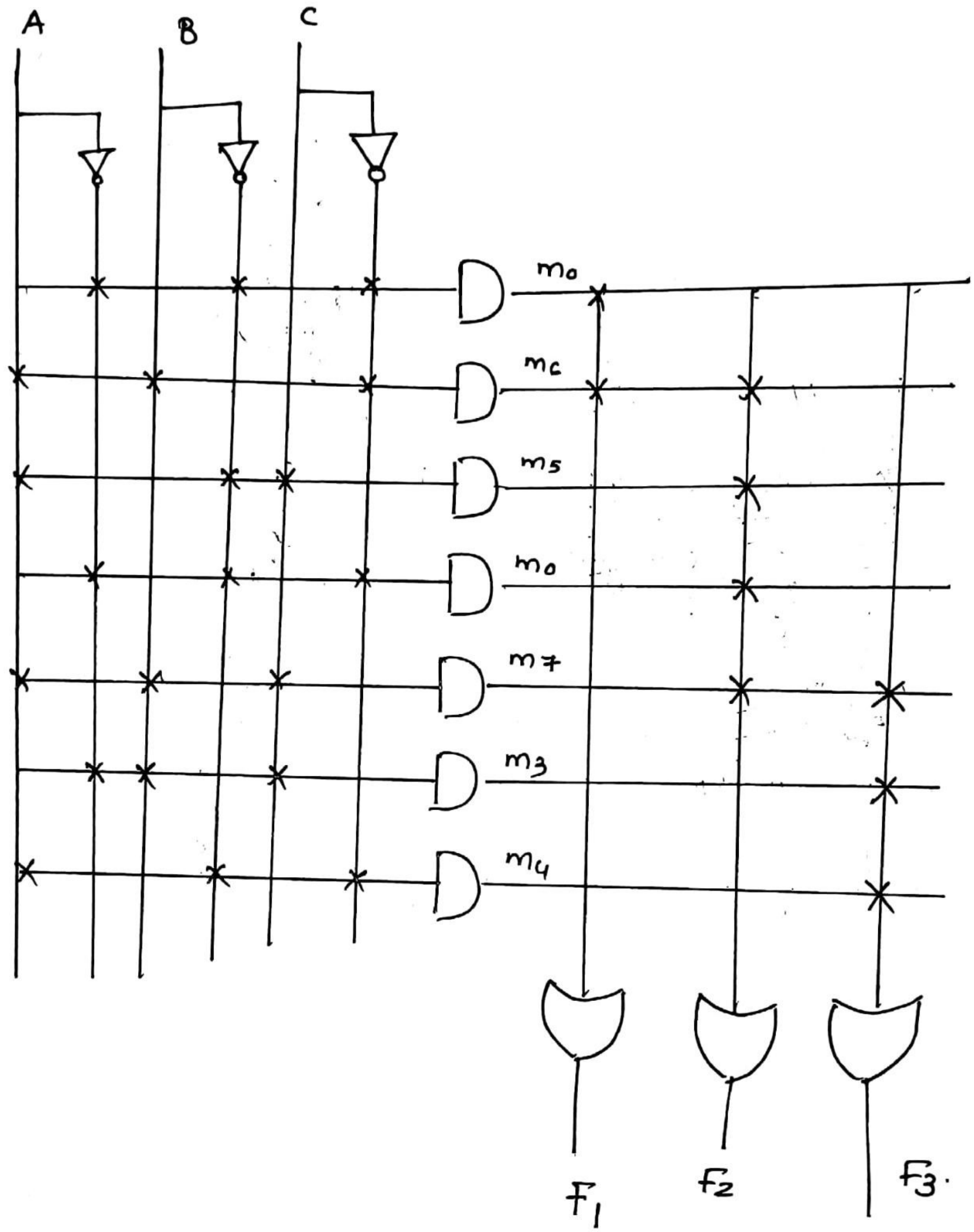
$= \bar{A}\bar{B}C + AB\bar{C}$

$F_2 = m_0 + m_5 + m_6 + m_7$

$= \bar{A}\bar{B}\bar{C} + A\bar{B}C + AB\bar{C} + ABC$

$F_3 = m_3 + m_4 + m_7$

$= \bar{A}BC + A\bar{B}\bar{C} + ABC$.



A
B
C
mo
mc
m5
mo
m7
m3
m4
F1
F2
F3.

2. Design an AND-OR PLA with outputs

$F_1 = m_0 + m_2 + m_6$

$F_2 = m_0 + m_5 + m_6$

$F_3 = m_3 + m_4 + m_7$

A	B	C	
0	0	0	$- m_0 = \bar{A}\bar{B}\bar{C}$
0	0	1	$- m_1 = \bar{A}\bar{B}C$
0	1	0	$- m_2 = \bar{A}B\bar{C}$
0	1	1	$- m_3 = \bar{A}BC$
1	0	0	$- m_4 = A\bar{B}\bar{C}$
1	0	1	$- m_5 = A\bar{B}C$
1	1	0	$- m_6 = AB\bar{C}$
1	1	1	$- m_7 = ABC$

Given, $F_1 = m_0 + m_2 + m_6$

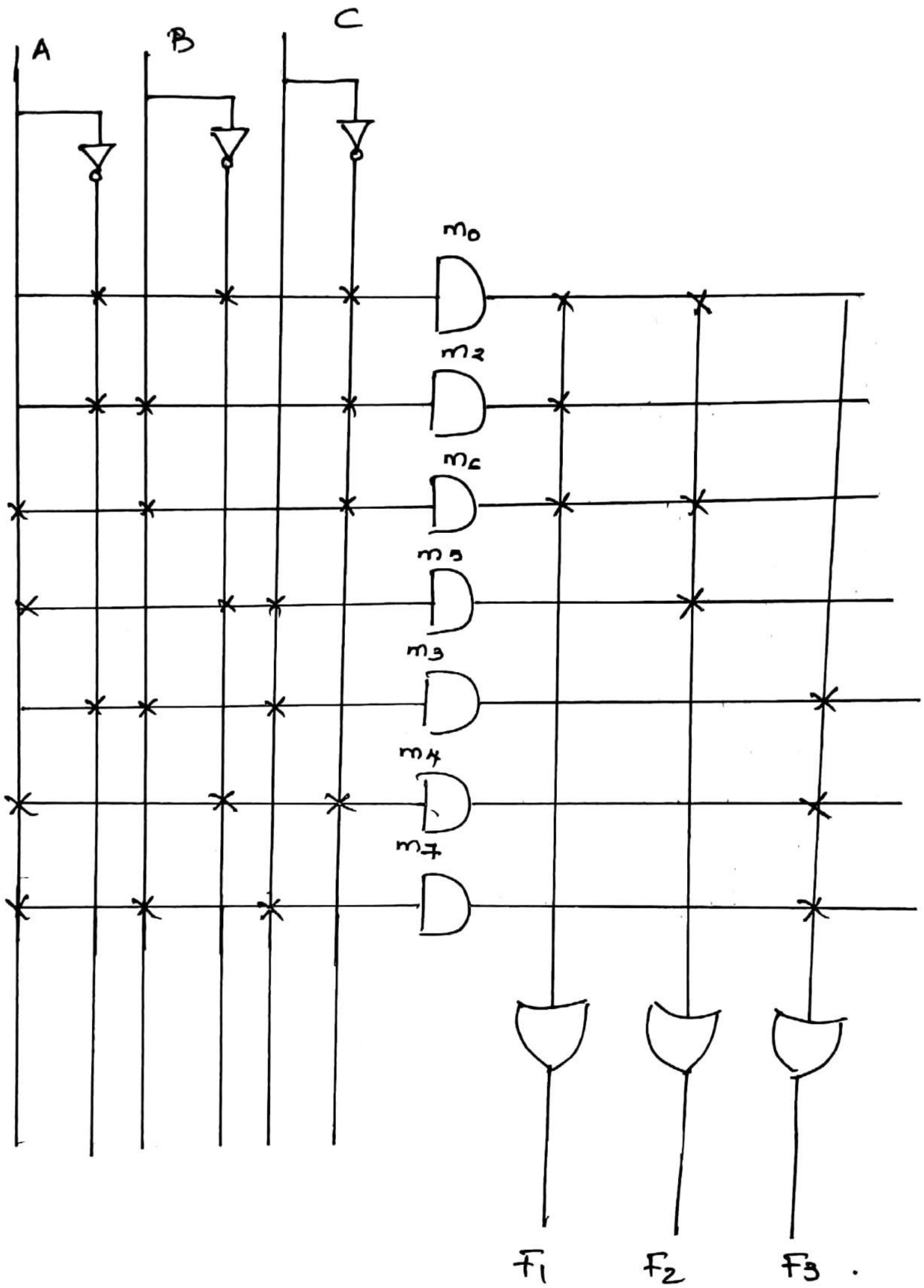
$= \bar{A}\bar{B}\bar{C} + \bar{A}B\bar{C} + AB\bar{C}$

$F_2 = m_0 + m_5 + m_6$

$= \bar{A}\bar{B}\bar{C} + A\bar{B}C + AB\bar{C}$

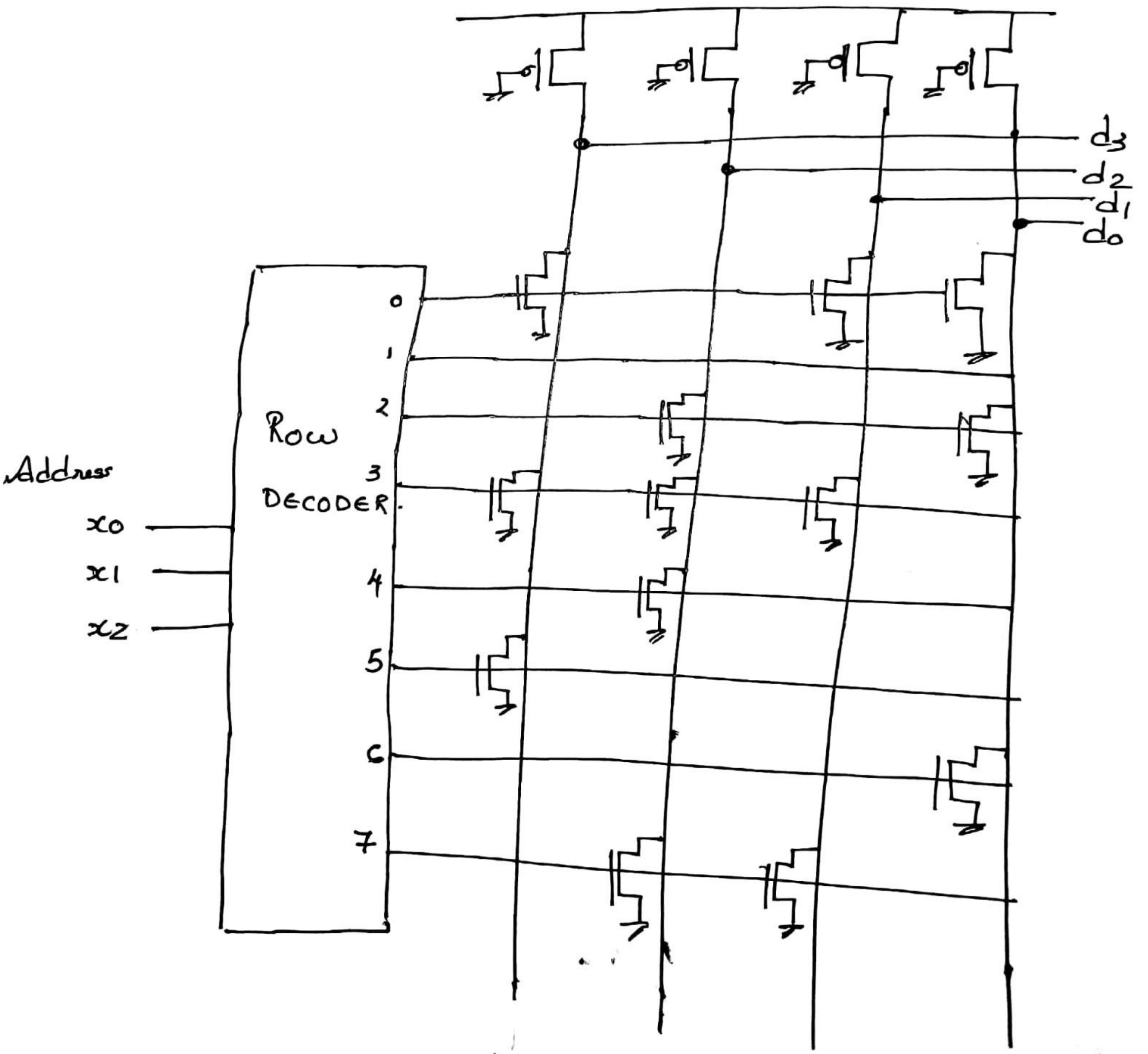
$F_3 = m_3 + m_4 + m_7$

$= \bar{A}BC + A\bar{B}\bar{C} + ABC.$

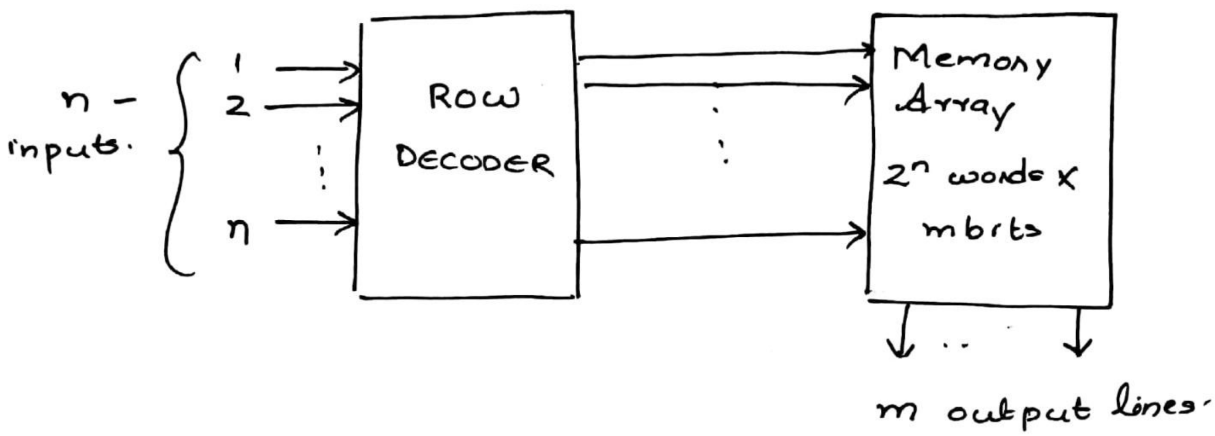

A
B
C
m0
m2
m6
m5
m3
m4
m7
F1
F2
F3

1. Design an FET programmable ROM that contains the following data.

Address	0	1	2	3	4	5	6	7
Data	0100	1111	1010	0001	1011	0111	1110	1001



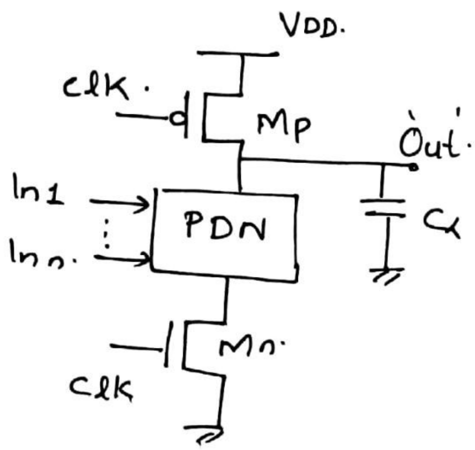
Basic ROM Structure



Truth Table

Address			Data			
x_0	x_1	x_2	d_3	d_2	d_1	d_0
0	0	0	0	1	0	0
0	0	1	1	1	1	1
0	1	0	1	0	1	0
0	1	1	0	0	0	1
1	0	0	1	0	1	1
1	0	1	0	1	1	1
1	1	0	1	1	1	0
1	1	1	1	0	0	1

Dynamic CMOS Design



Precharge [CLK=0]

→ 'Out' is precharged to VDD by pmos transistor Mp.

→ Mn → off.

Evaluate [CLK = 1]

→ Precharge transistor Mp - OFF

→ Mn turned ON., the 'Out' is discharged to GND. (when PDN is ON.

→ If PDN turned OFF, the precharged value remains stored on C_L.

Design Problems

1. Charge Leakage.

→ If the PDN is off, the o/p should remain at the precharged state of VDD during Evaluation phase.

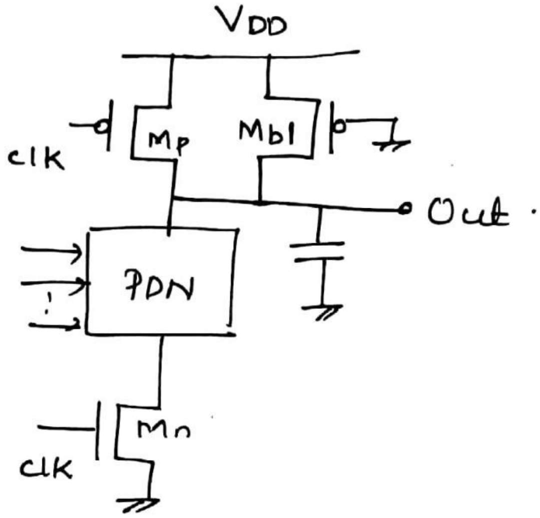
→ However, this charge gradually leaks away due to leakage currents.

→ Leakage is caused by high-Impedance state of output node during evaluate mode, when PDN is off.

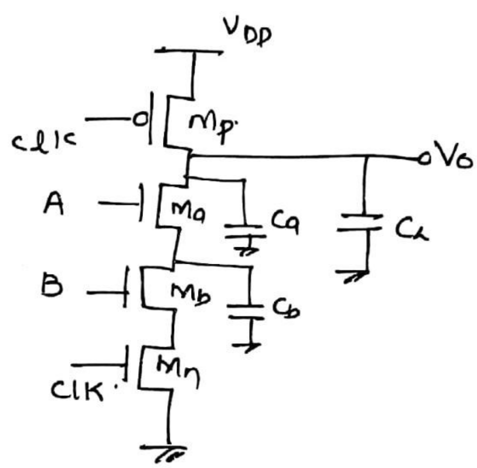
→ Leakage problem can be ~~reduce~~ avoid by reducing the output impedance on output node during evaluation.

→ This is done by adding a - Bleeder Transistor.

→ Fn of bleeder - an nMOS style pull-up device, is to compensate for the charge lost due to pull-down leakage paths.



2. Charge Sharing



→ Output charged upto VDD ⇒ Precharge phase.

→ When all inputs are zero, Ca is discharged.

→ Evaluation, A=1, B=0 ⇒ Ma ON

Charge stored in C_L distributed over C_L and Ca. It cause a drop in output voltage.

$$V_{out} = V_{DD} \cdot \frac{C_a}{C_a + C_b}$$

→ Voltage drop due to charge sharing can be reduced by pre charging the internal nodes.

MODULE-6

CMOS System Design

Adders

Addition is the most commonly used arithmetic operation.

The truth table of binary full adder is given below, where A and B are inputs, C_i is the carry input, S is the sum output, C_o is the carry output.

Truth Table - Full Adder

A	B	C_i	S	C_o	Carry status	
0	0	0	0	0	Delete.	i.e. $C_o = 0 \Rightarrow$ (carry is deleted)
0	0	1	1	0	Delete	
0	1	0	1	0	Propagate	$C_o = C_i$ (incoming carry propagate to output)
0	1	1	0	1	Propagate	
1	0	0	1	0	Propagate	
1	0	1	0	1	Propagate.	
1	1	0	0	1	Generate/propagate	$C_o = 1$ (carry is generated)
1	1	1	1	1	Generate/propagate	

$$S = A \oplus B \oplus C = \bar{A}\bar{B}C_i + \bar{A}B\bar{C_i} + A\bar{B}\bar{C_i} + ABC_i$$

$$C_o = AB + BC_i + AC_i$$

It is often useful to define 'S' and C_o as functions of some intermediate signals G (generate), D (delete) and P (propagate).

G = 1 ensures that a carry bit will be generated at C_o independent of C_i

D = 1 ensures that a carry bit will be deleted at C_o independent of C_i

P = 1 guarantees that an incoming carry will propagate to C_o.

Expressions for these signals can be derived from the truth table.

$$G = A \cdot B$$

$$P = A \oplus B \text{ (sometimes } A + B\text{)}.$$

$$D = \bar{A}\bar{B}$$

We can rewrite S and C_o as functions of P and G.

$$C_o = AB + C_i(A+B)$$

$$= G + C_i P$$

$$S = A \oplus B \oplus C_i = P \oplus C_i$$

$$\therefore \quad C_o(G,P) = G + PC_i$$

$$S(G,P) = P \oplus C_i$$

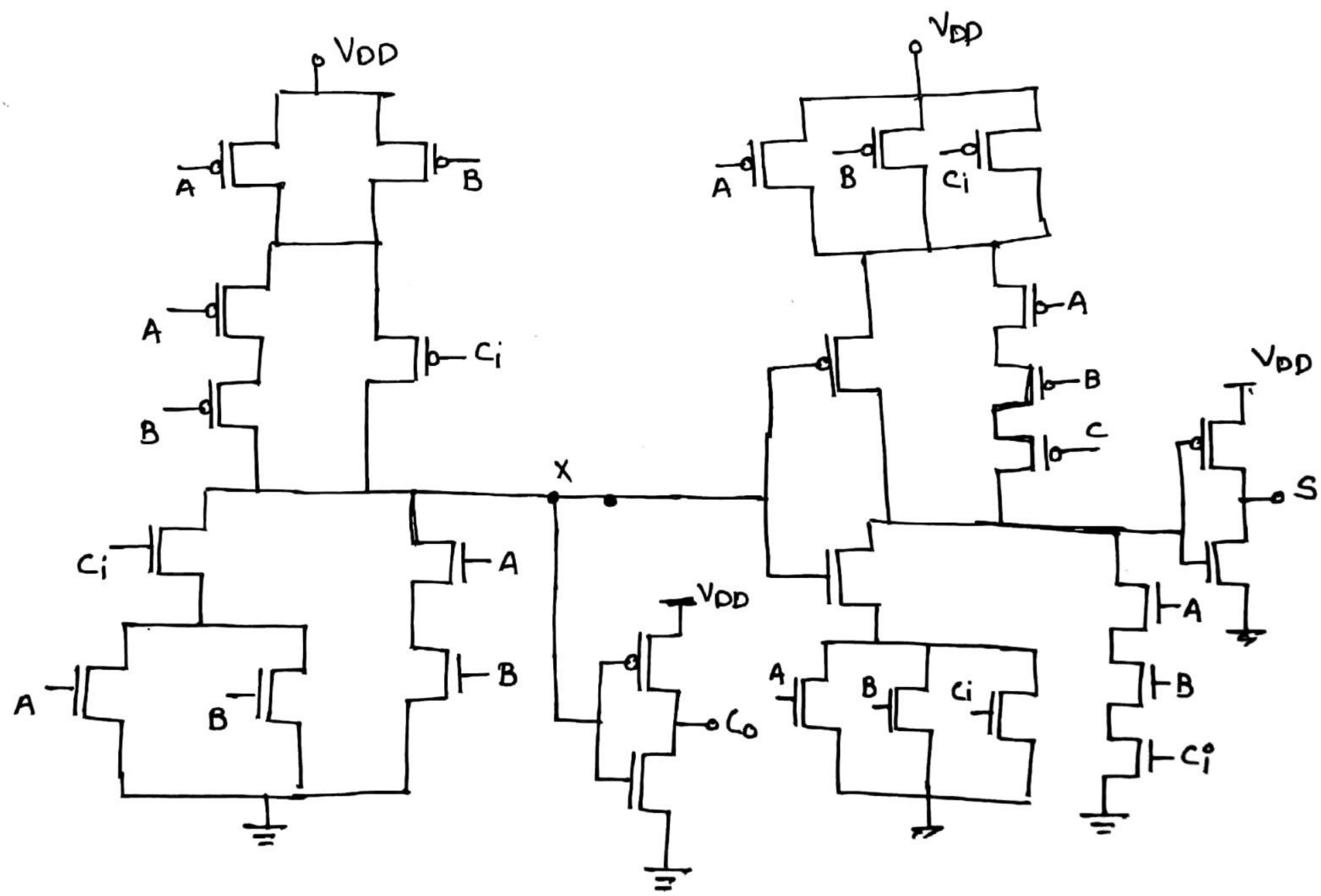
STATIC ADDER CIRCUIT

One way to implement the full adder circuit is to take logic equations and translate them directly into complementary CMOS circuitry.

Some logic manipulations can help to reduce the transistor count.

$$S = ABC_i + \overline{C_o}\,(A+B+C_i)$$

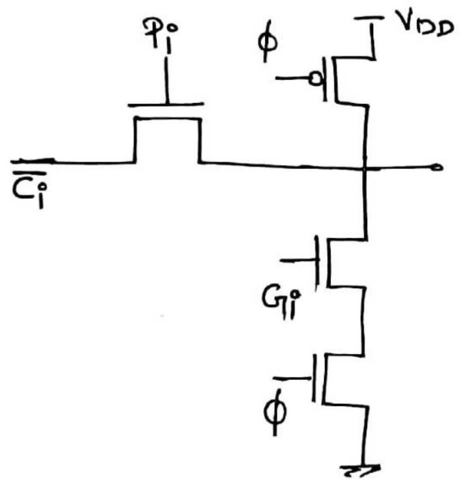
$$C_o = AB + C_i\,(A+B)$$



• Complementary Static CMOS implementation of full adder

DYNAMIC ADDER [MANCHESTER CARRY CHAIN ADDER]

Dynamic implementation using only propagate and generate signals. The dynamic implementation use only propagate and generate signals. Since the transitions in a dynamic circuit are monotonic, (i.e. only high to low transition during the evaluation phase), the transmission gates can be replaced by NMOS - only pass transistors.



When $\phi = 0$ [precharge], o/p is always pulled to V_{DD}.

$\overline{C_o} = 1 \Rightarrow C_o = 0$ [Similar to delete].

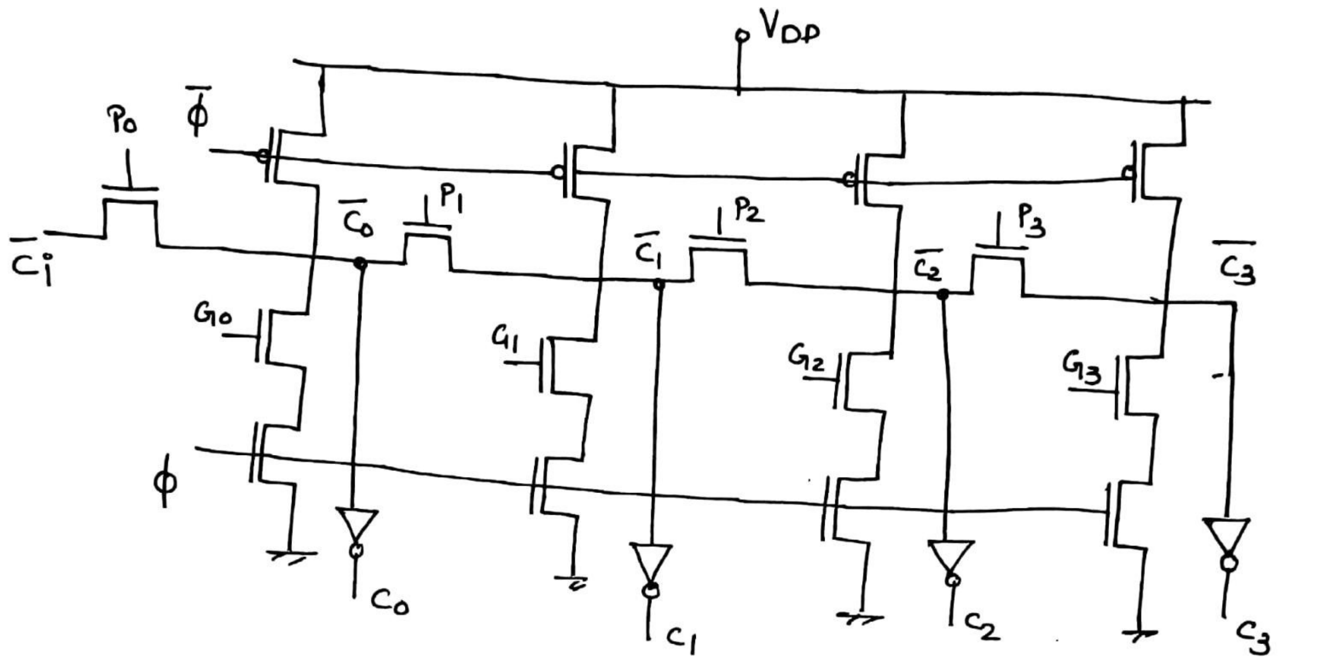
When $\phi = 1$ [evaluate]

Case 1: if $P_i = 1$, $\overline{C_o} = \overline{C_i} \Rightarrow C_o = C_i$

Case 2: if $G_i = 1$, $\overline{C_o} = 0 \Rightarrow C_o = 1$

P_i and G_i are not become high/low simultaneously.

eg: Manchester carry chain adder in dynamic logic [4 bit].



During the precharge phase ($\phi=0$), all the intermediate nodes of the pass-transistor carry chain are precharged to VDD.

During evaluation, the C_k node is discharged when there is an incoming carry ($C_i=1$) and the propagate signal P_K is high, or when generate signal for stage k (G_K) is high.

CARRY BYPASS ADDER [CARRY SKIP CIRCUIT]

From truth table of full adder.

$$G = A \cdot B$$

$$D = \bar{A}\,\bar{B}$$

$$P = A \oplus B$$

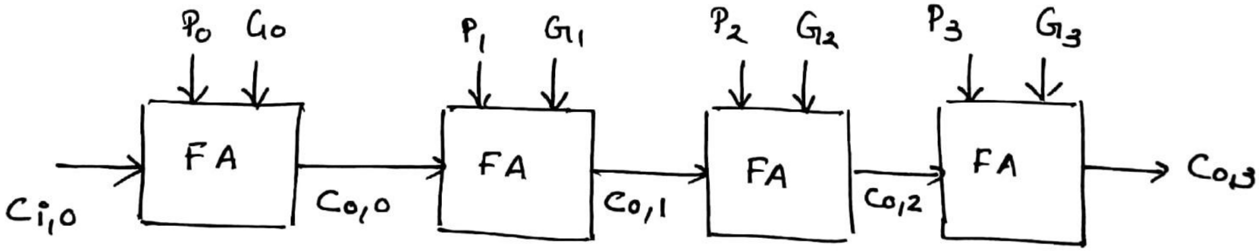
$$C_o(G, P) = G + PC_i$$

$$S(G, P) = P \oplus C_i$$

Consider the 4 bit adder block in the fig(1) below. Suppose that the values of A_K and B_K $(K = 0, 1, 2, 3)$ are such that all propagate signals P_K $(K = 0, 1, 2, 3)$ are high. An incoming carry $C_{i,0} = 1$, propagate under those conditions through the complete adder chain and causes an outgoing carry $C_{o,3} = 1$.

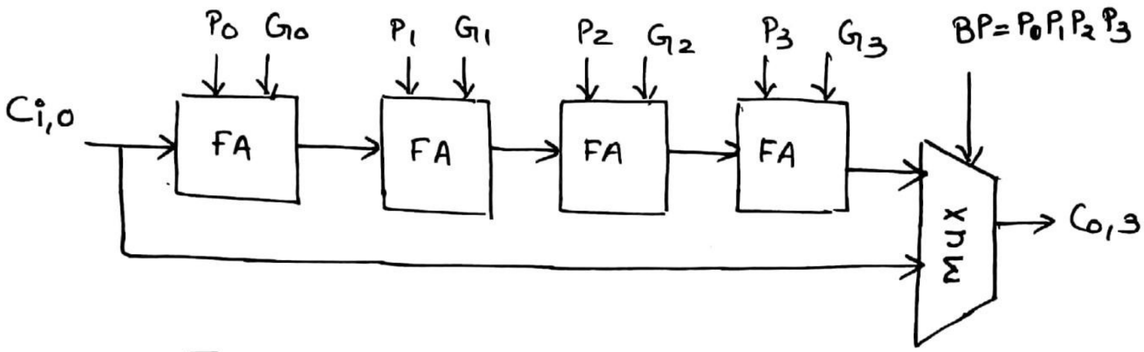
If $P_0 P_1 P_2 P_3 = 1$, then $C_{o,3} = C_{i,0}$

else either DERETE or GENERATE occur.



Fig(1): Carry Propagation

This information can be used to speed up the operation of adders, as shown in fig (2).



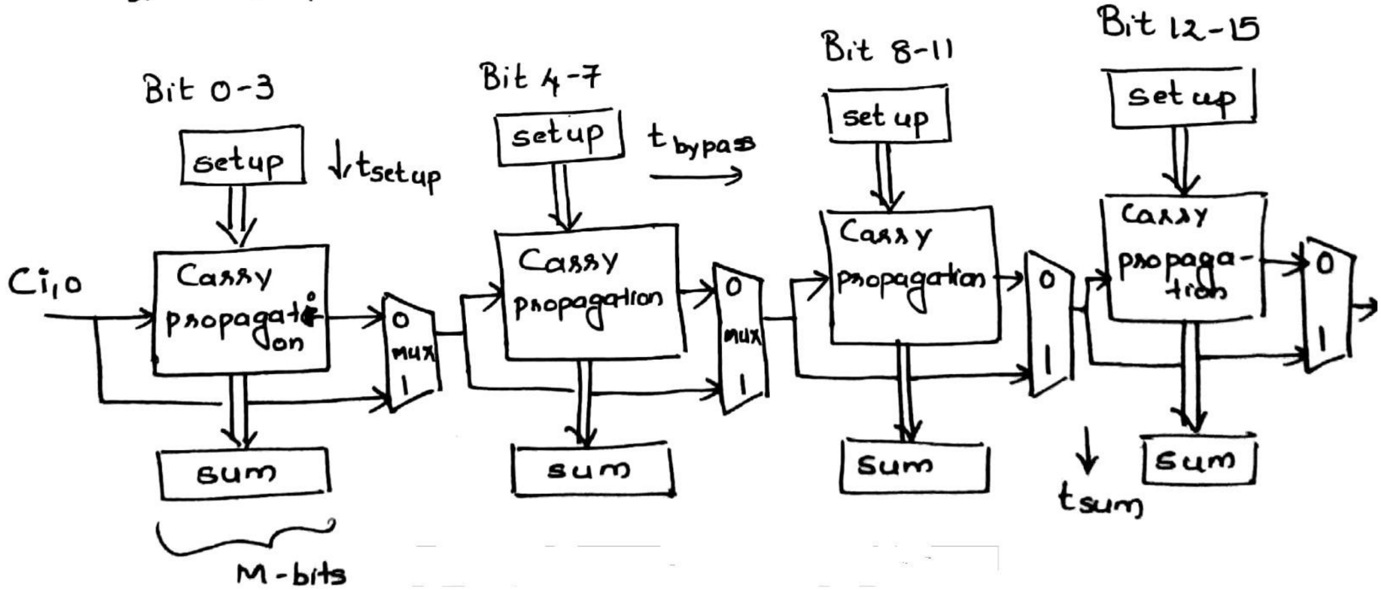
Fig(2): Adding a bypass

When $BP = P_0 P_1 P_2 P_3 = 1$, the incoming carry is forwarded immediately to the next block through the bypass transistor, hence the name carry bypass adder or carry skip adder.

eg: Design a N-bit adder [N=16]. and also calculate the delay.

At first the total adder, is divided in (N/M) equal-length bypass stages, each of which contain 'M' bits

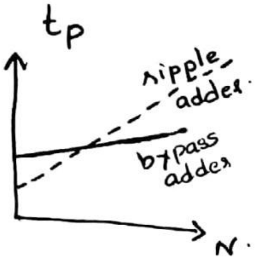
$2^M = N$, $\therefore 2^M = 16$ $\therefore M = 4$



Delay, $t_p = t_{setup} + \left(\frac{N}{M} - 1\right) t_{bypass} + (M-1) t_{carry} + t_{sum}$

$+ M \cdot t_{carry}$.

where t_{setup} → time to create the generate and propagate signals.



t_{bypass} → propagation delay through MUX

t_{carry} → propagation delay through a single bit

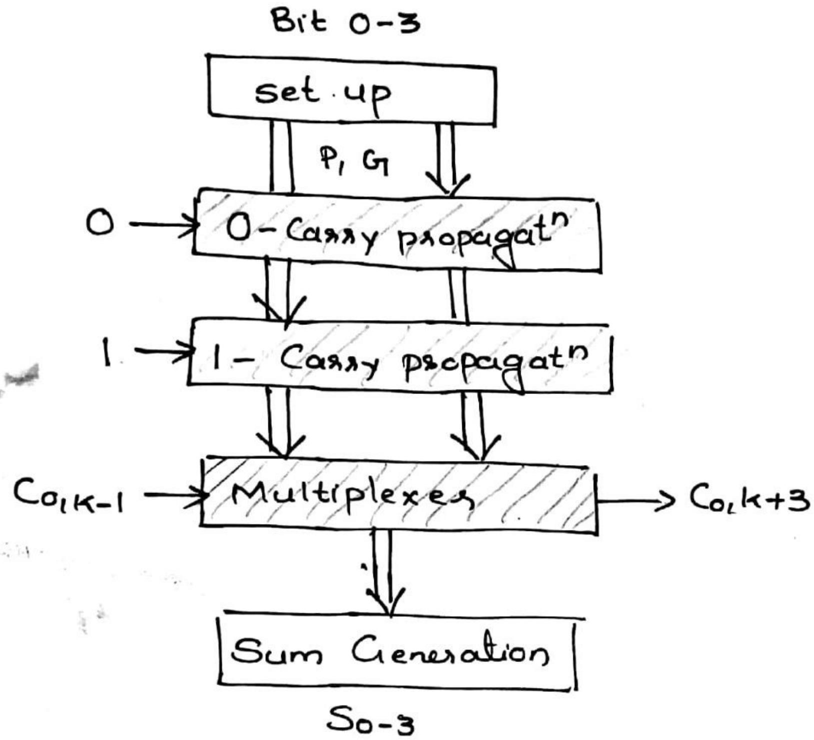
t_{sum} → time to generate the sum of final stage.

CARRY SELECT ADDER

In a ripple carry adder, every full adder cell has to wait for the incoming carry before an outgoing carry can be generated. By calculating the possible value of carry inputs and evaluating the result of both possibilities in advance, the linear dependency problem is eliminated in carry select adder. And then use a multiplexer to select between the output choices.

The carry select adder does this with a pair of M-bit adders in each group. One adder calculates the sum assuming a carry-in of '0' while the other calculates the sum assuming a carry-in of '1'. The actual carry triggers a multiplexer that chooses the appropriate sum.

4 bit carry select adder



• Shaded portion represent the critical path.

The worst-case propagation delay is

$$t_{add} = t_{setup} + M\,t_{carry} + \left(\frac{N}{M}\right) t_{mux} + t_{sum}$$

where t_{setup}, t_{sum}, t_{mux} are fixed delays.

$N \rightarrow$ total number of bits

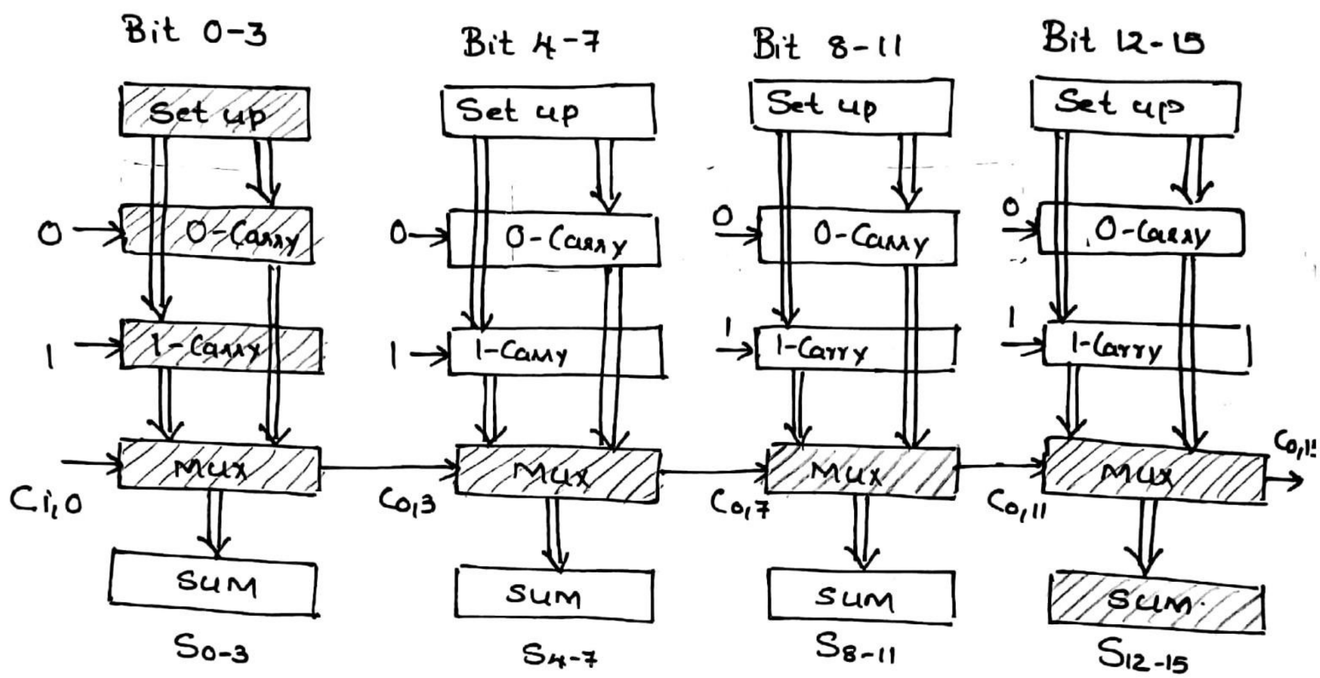
$M \rightarrow$ number of bits per stage

$t_{carry} \rightarrow$ delay of the carry through a single full-adder cell.

The carry delay of single block $\propto$ length $= M\,t_{carry}$.

Propagation delay $\propto N$.

16-bit Carry Select Adder.

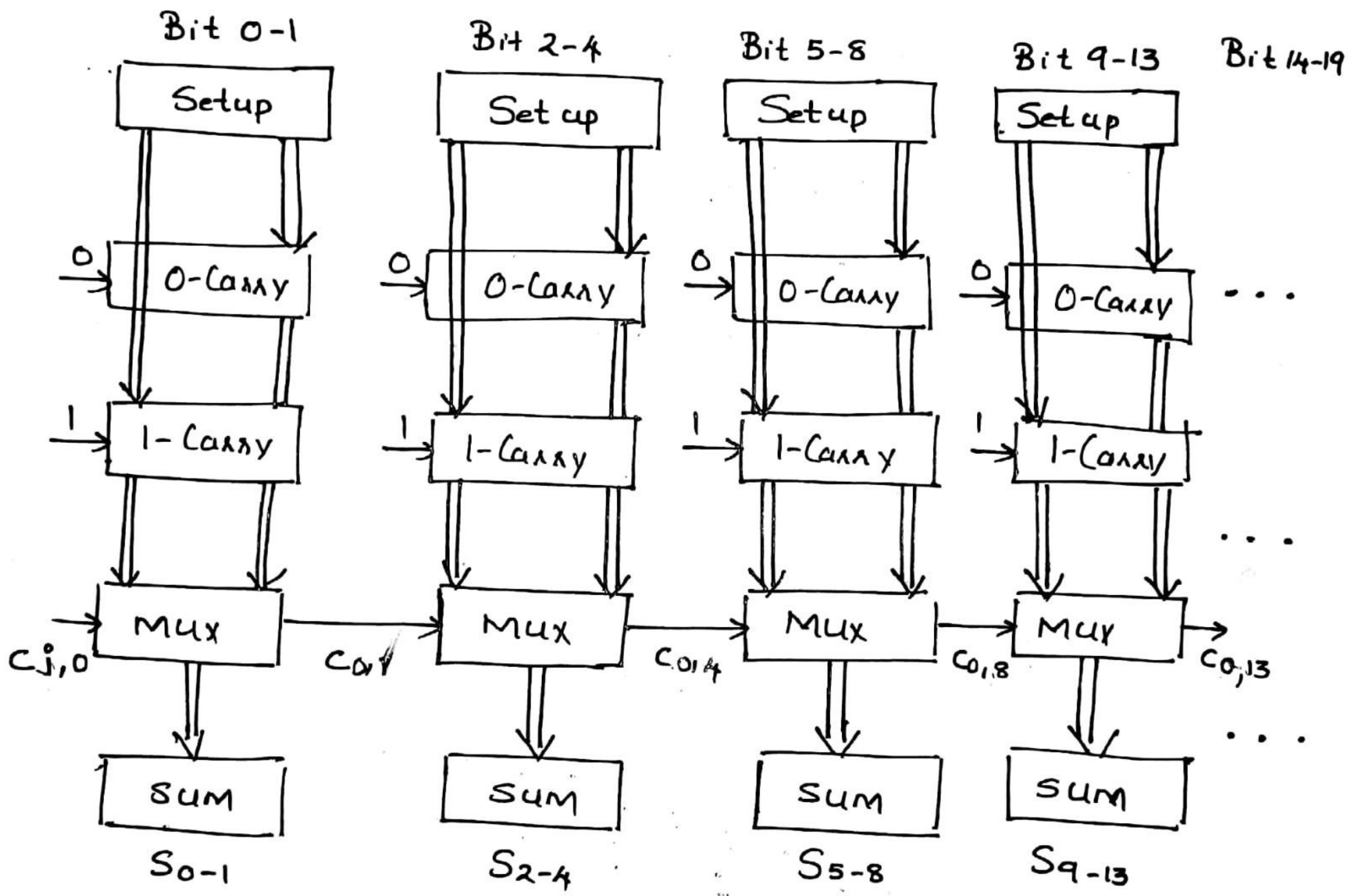


The shaded portion represent the critical path.

SQUARE ROOT CARRY SELECT ADDER

Consider the multiplexer gate in the last adder stage [16 bit carry select adder]. The inputs to this multiplexer are the two carry chains of the blocks and block-multiplexer signal from the previous stage. A major mismatch between the arrival times of the signals can be observed. The results of the carry chains of the blocks are stable before the mux signal arrives. It makes sense to equalize the delay through both paths.

This can be achieved by adding more bits to the subsequent stages in the adder, requiring more time for the generation of the carry signals.
For eg:, the first stage can add 2 bits, the second contains 3, the third has 4 and so forth.



- Square root Carry Adder.

Delay Calculation

Assume that an N-bit adder contain P stages and the first stage adds M-bits. An additional bit is added to each subsequent stage.

$$N = M + (M+1) + (M+2) + \cdots + (M+P-1)$$

$$= MP + \frac{P(P-1)}{2} = \frac{P^2}{2} + P\left(M - \frac{1}{2}\right) \longrightarrow (1)$$

If $M << N$ (eg: $M=2$ and $N=64$), the first term dominates and eqn (1) becomes

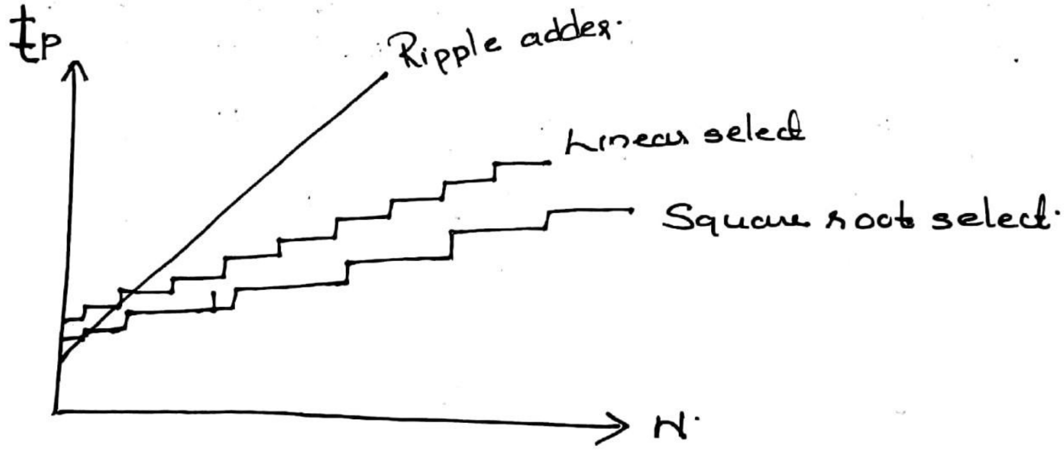
$$N = \frac{P^2}{2} \longrightarrow (2)$$

$$\therefore P = \sqrt{2N}$$

$$\therefore t_{add} = t_{setup} + M t_{carry} + (\sqrt{2N}) t_{mux} + t_{sum}$$

The delay $\propto \sqrt{N}$ for large adders.

Larger the value of N, tadd become almost a constant



- Propagation delay of square-root carry select adder Vs linear ripple and select adders.

CARRY LOOKAHEAD ADDER (CLA)

Carry lookahead adders (CLA) are designed to overcome the latency the first fulladder has to wait until the carry from the previous halfadder has arrived at its input.

CLA are used for fast addition by incorporating the addition of carry term and making them available at the input of respective adders at the same time.

$$C_{i+1} = A_i B_i + C_i (A_i \oplus B_i)$$

$$C_{i+1} = G_i + P_i C_i$$

$$S_i = P_i \oplus C_i$$

For a 4 bit adder,

$$C_1 = G_0 + P_0 C_0$$

$$C_2 = G_1 + P_1 C_1 = G_1 + P_1 (G_0 + P_0 C_0)$$

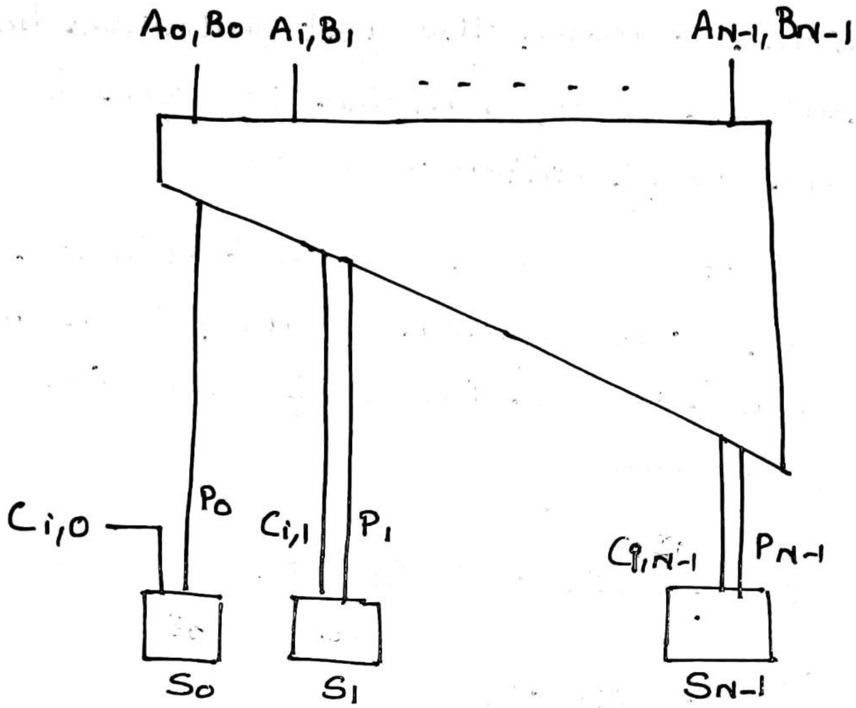
$$C_3 = G_2 + P_2 C_2 = G_2 + P_2 [G_1 + P_1 (G_0 + P_0 C_0)]$$

$$C_4 = G_3 + P_3 C_3$$

For every bit the carry and sum outputs are independent of the previous bits.

The ripple effect has thus been effectively eliminated and addition time should be independent of the no. of bits.

N-bit Carry-lookahead adder.



Large fan-in of the circuit makes it prohibitively slow for larger values of N.

MULTIPLIER

Multiplications are expensive and slow operations. The performance of many computational problems often is dominated by the speed at which a multiplication operation can be executed. Multipliers are complex adder arrays.

Consider two unsigned binary numbers X and Y that are M and N bit wide, respectively.

Express X and Y in binary representation

$$X = \sum_{i=0}^{M-1} x_i 2^i \quad , \quad Y = \sum_{j=0}^{N-1} y_j 2^j$$

$x, y \in \{0, 1\}$.

The multiplication operation is then defined as

$$Z = X * Y$$

$$= \left(\sum_{i=0}^{M-1} x_i 2^i \right) \left(\sum_{j=0}^{N-1} y_j 2^j \right)$$

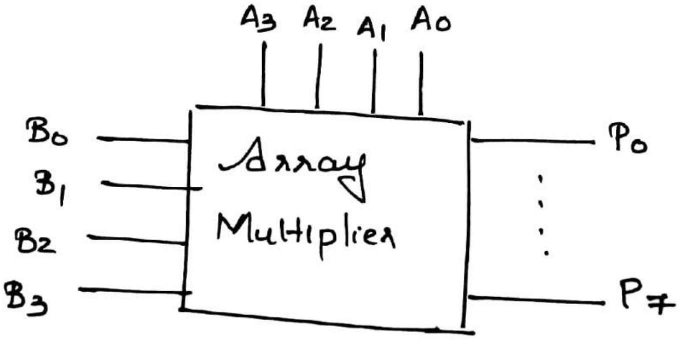
$$Z = \sum_{i=0}^{M-1} \left[\sum_{j=0}^{N-1} x_i y_j 2^{i+j} \right]$$

The simple way to perform multiplication is by using shift and add algorithm, for multiplication adds together M-partial products.

Each partial product is generated by multiplying the multiplicand with a bit of multiplier.
All partial product are generated at the same time and organized in an array. A multioperand addition is applied to compute the final product.

```
           1 0 1 0 1 0 ×   (multiplicand)
               1 0 1 1     (multiplier)
     ---------------------
           1 0 1 0 1 0  ⎫
         1 0 1 0 1 0    ⎪
       0 0 0 0 0 0      ⎬  Partial products.
     1 0 1 0 1 0        ⎭
     ---------------------
     1 1 1 0 0 1 1 1 0     (Result)
```

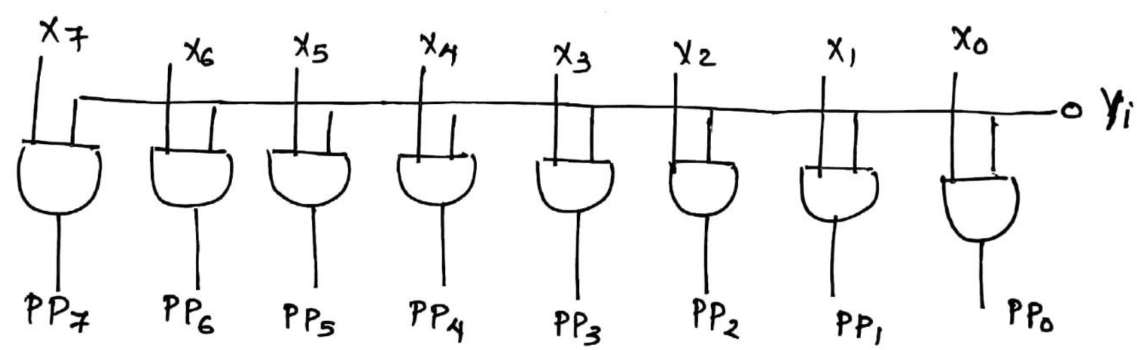
This set of operation can be mapped directly into hardware. The resulting structure is called an array multiplier and it combines the 3 functions: partial-product generation, partial-product accumulation and final addition.



PARTIAL PRODUCT GENERATION

Partial products result from logical AND of multiplicand X with a multiplier bit y_i. Each row in the partial product array is either a copy of the multiplicand or a row of zero's.

In most cases, the partial product array has many zero rows that have no impact on the result and thus represent a waste of effort when added. In case of a multiplier consisting of all ones, all the partial products exist, while in the case of all zeros, there is none. This observation allow us to reduce the no: of generated partial product by half.



PARTIAL PRODUCT ACCUMULATION

After partial products are generated, they must be summed. This accumulation is essentially a multioperand addition. One method to accumulate partial products is by using a number of adders, that will form an array - hence the name <u>array multiplier</u>.

ARRAY MULTIPLIER

For M x N array multiplier, it needs

a) M * N - AND gates

b) N → Halfadders

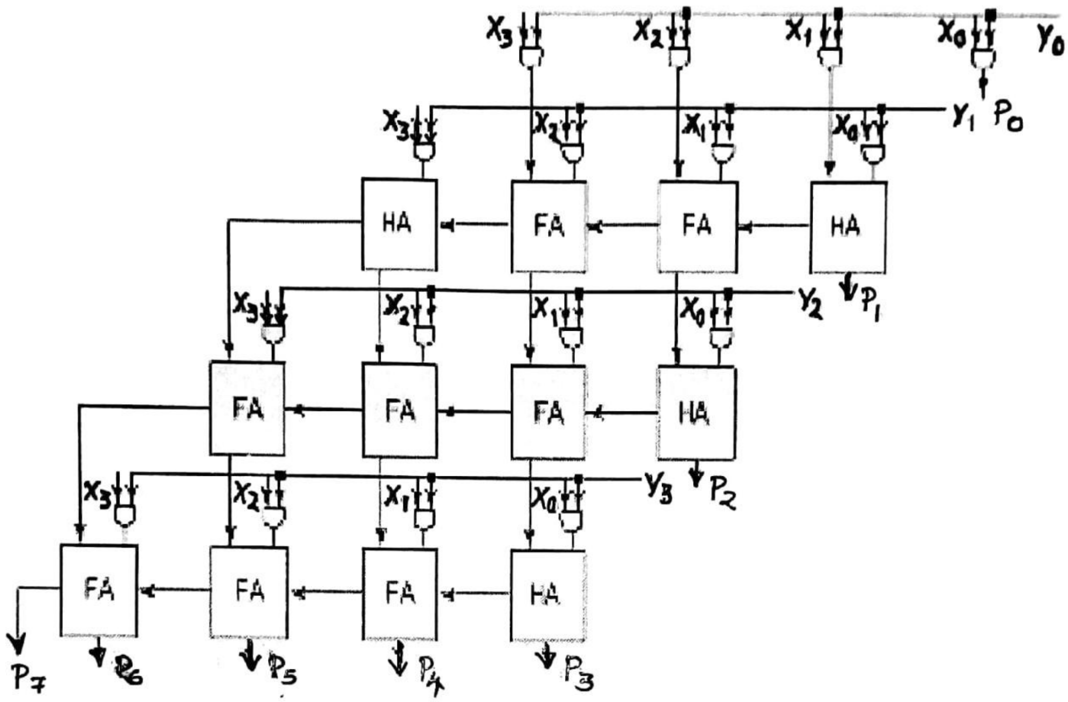
c) (M-2) N → Full adders

Total (M-1) N → Adders.

Principles of Array Multiplier.

4 x 4 bit multiplication

			x_3	x_2	x_1	x_0	
			y_3	y_2	y_1	y_0	
			x_3y_0	x_2y_0	x_1y_0	x_0y_0	+
		x_3y_1	x_2y_1	x_1y_1	x_0y_1		
	x_3y_3	x_2y_2	x_1y_2	x_0y_2			
x_3y_3	x_2y_3	x_1y_3	x_0y_3				



The adders are arranged in a carry save chain, the carry-out bits are fed to the next available adder in the column to the left. The array multiplier accepts all of the input simultaneously. The longest delay in the calculation of the product bits depends on the speed of adders. The carry chain in P_7 that originates from the carry bits from the P_1 column and propagate through P_2-P_6 quantities would be an obvious problem.

Delay

$$t_{mult} = [(M-1)+(N-2)]\,t_{carry} + (N-1)\,t_{sum} + t_{and}.$$

t_{and} - delay of AND gate
t_{carry} - delay b/w input and output carry.

FINAL ADDITION

The final step for completing the multiplication is to combine the result in the final adder. The choice of adder style depends on the structure of accumulation array. A carry-lookahead adder is the preferable option if all input bits to the adder arrive at the same time, as it have the smallest possible delay.

WALLACE TREE MULTIPLIER

The partial sum adders can be rearranged in a treelike fashion, reducing both critical path and numbers of adder cells needed.

Consider the simple example of four partial products each of which is four bits wide. The number of full adders needed for this operation can be reduced by observing that only column 3 in the array has to add four bits. All other columns are less complex. Fig(b) represents the original matrix of partial product into tree shape.

Partial products.

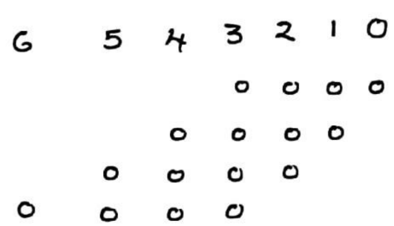


Fig (a): Pa.

First stage: Bit position

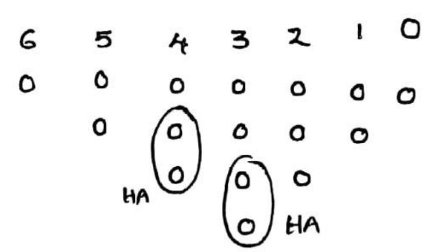
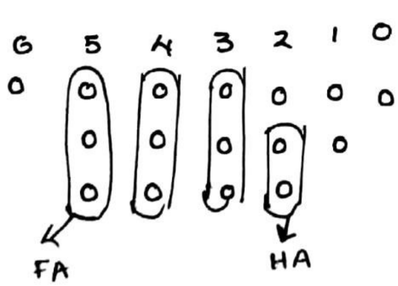


Fig (b).

Second stage



Fig(c)

Final adder

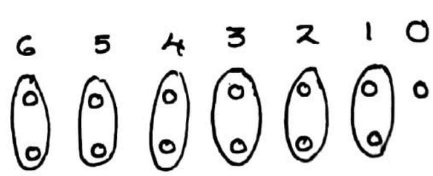


Fig (d)

Full adder:- it take 3 inputs and produce 2 outputs, the sum located in the same column and carry located in the next one.

Half adder:- it take 2 inputs and produce 2 outputs.

In the first step, we introduce HAs in columns 4 and 3 (fig b). The reduced tree is shown in fig(c). A second round of reductions creates a tree of depth 2.