1 Introduction

Syllabus

Overview of VLSI design methodology, VLSI design flow, Design hierarchy, Concept of regularity, Modularity, and Locality, VLSI design style, Design quality, package technology, introduction to FPGA and CPLD, computer aided design technology.

Contents

1.1 Overview of VLSI Design Methodology

1.2 VLSI Design Flow ..... Summer-15, 21, ... Winter-21, ... Marks 7

1.3 Design Hierarchy

1.4 Concept of Regularity, Modularity and Locality Summer-21, Winter-21, ... Marks 3

1.5 VLSI Design Styles ..... Winter-21, Summer-21 .... Marks 4

1.6 Design Quality ..... Winter-14, Summer-15, ... Marks 7

1.7 Package Technology

1.8 Architecture of FPGA ..... Summer-21, ... Marks 7

1.9 Complex Programmable Logic Device (CPLD) ..... Winter-14, 21, Summer-21, Marks 7

1.10 Computer-Aided Design Technology

1.11 Short Questions and Answers

1.1 Overview of VLSI Design Methodology

- Because of IC technology it is possible to design powerful and flexible processors and hence highly intelligent devices for the user. IC memories provide the essential element for these processors.
- Improvement in IC technology has greatly enhanced space exploration, made automobiles more efficient, revolutionized nature of warfare, brought vast libraries of information to web browsers.
- The technique to increase number of devices per chip is called level of integration. The level of integration of chips has resulted in rapid transition from Small Scale Integration (SSI), Medium Scale Integration (MSI), Large Scale Integration (LSI), through Very Large Scale Integration (VLSI) technologies. The next phase will be Ultra Large Scale Integration (ULSI).

Integration levels	
SSI	10 Gates
MSI	1000 Gates
LSI	10,000 Gates
VLSI	> 10 kGates

- This evolution is based on the number of devices per chip as well as size of the chip and the technology used.

Moore's Law

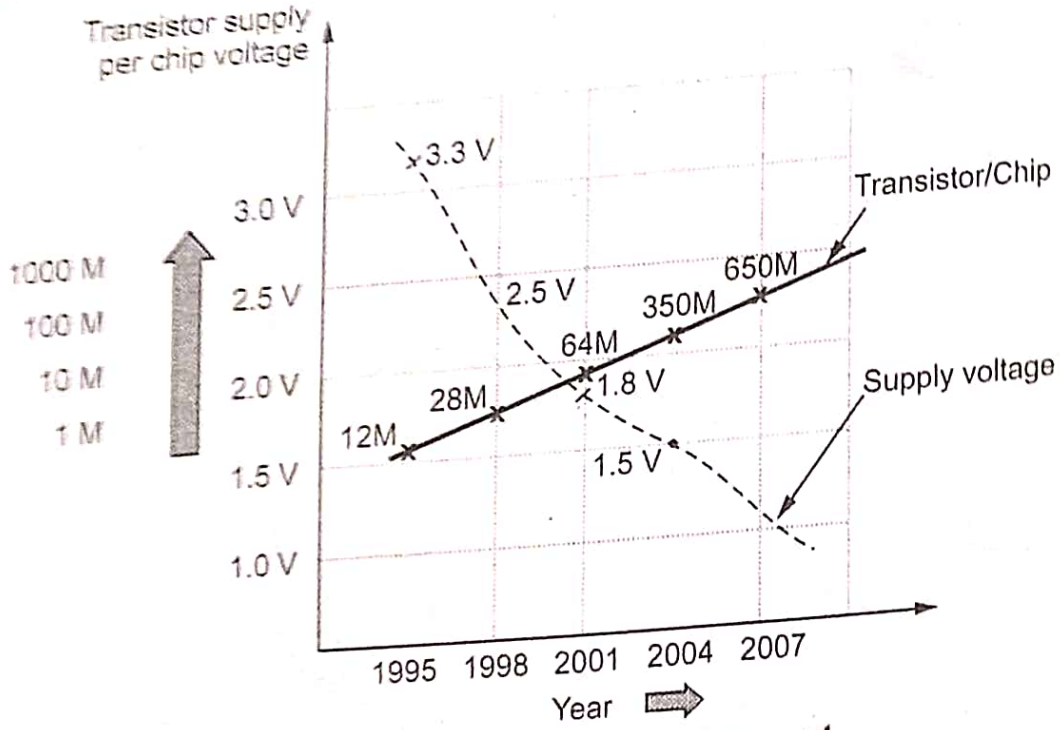


Fig. 1.1.1 IC technology development

- Gorden Moore (of Intel) in 1965 found that plotting the number of transistor that can be most economically fabricated on a chip is a straight line on semilog scale. He also devised that lowering the supply voltages and channel lengths together with increasing transistor per chip ratios will make system more efficient and smaller in size and can be packed at significantly higher package density. Moore also stated that Silicon - technologies will double the number of transistors per chip every 18 months.
- Table 1.1.1 shows technology roadmap for semiconductors.

Parameter	1997 (250 nm)	1999 (180 nm)	2002 (130 nm)	2005 (100 nm)	2008 (70 nm)	2011 (60 nm)
DRAM (bits)	256 M	1 G	4 G	16 G	64 G	256 G
MPU (transistors/cm^2)	3.7 M	6.2 M	18 M	39 M	84 M	180 M
DRAM (Chipsize-mm^2)	280	400	560	790	1120	1580
MPU (Chipsize-mm^2)	300	340	430	520	620	750

Table 1.1.1

1.1 VLSI Design Styles

- During design cycle the circuit performance is improved by design improvements. The circuit performance is improved more rapidly in the begining and then gradually untill the performance finally saturates. The circuit performance level within a design time is much dependent on efficiency of design methodologies and design style. The two VLSI design styles - Semi - custom and full - custom are compared for their relative merits in the design.
- For full-custom design style it takes longer design time until maturity but has more oppertunity for circuit performance improvement. The product resulting has a high level of performance i.e. high processing speed, lower power dissipation and small silicon area.

  In semi-custom design style (standard-cell based design or EPGA) allows shorter designtime until maturity and less opportunity for performance improvement. The circuit performance is higher than full custom design as the components are already optimized.

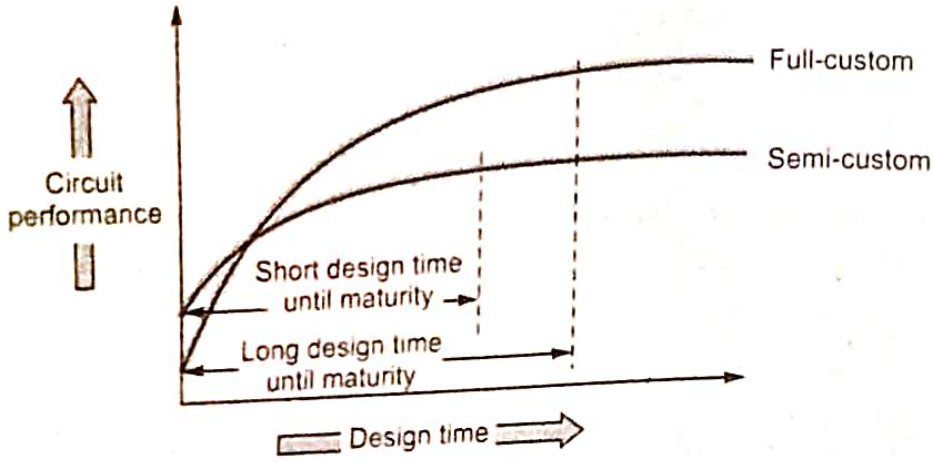


Fig. 1.1.2

- The choice of particular design style for VLSI product depends on performan requirements, technology used, expected life time of product and product cost.

1.2 VLSI Design Flow

GTU : Summer-15,21, Winter-21

- The design process starts with the given set of requirements. Initial design developed and tested for the given requirements. The design is improved

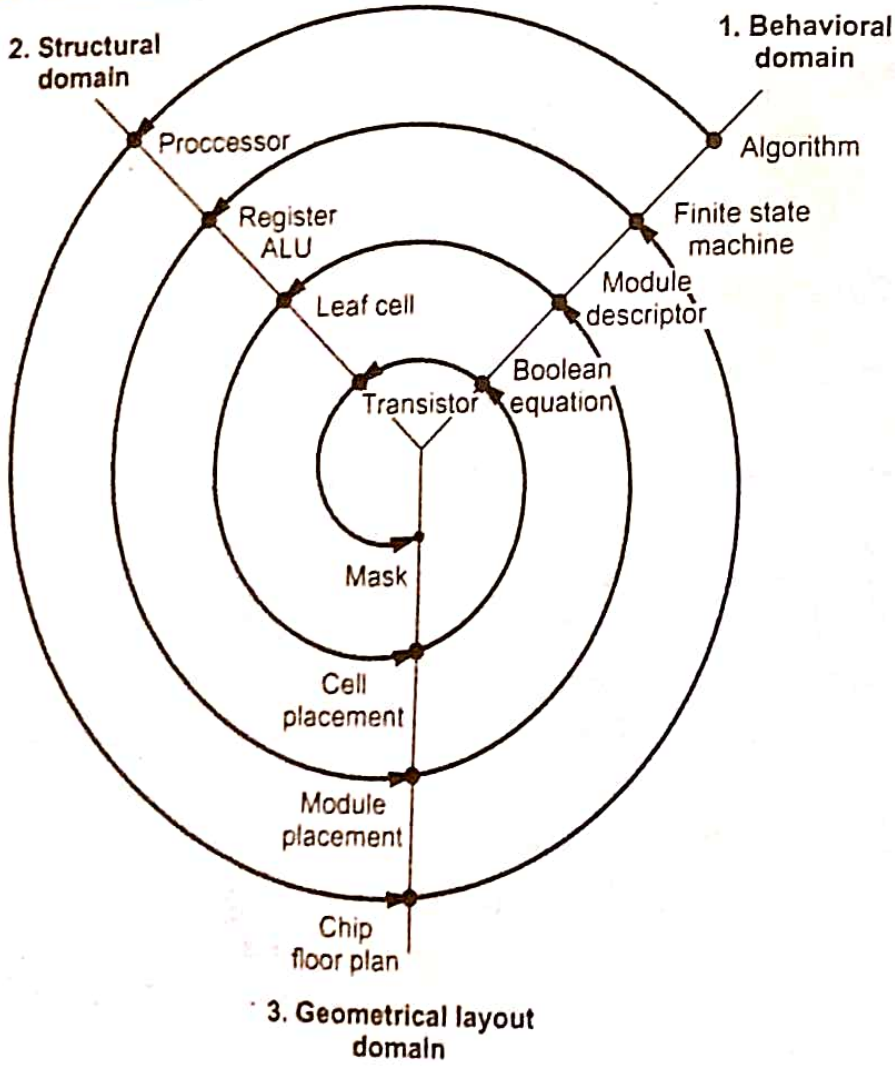


Fig. 1.2.1 VLSI design flow in three domain

needed. A simplified design flow was introduced by D. Gajski applicable for most logic chips. It is also called Y-chart which resembles letter "Y". Each branch represents three different domains. These domains are :

1. Behavioral domain.
2. Structural domain.
3. Geometrical layout domain.

- The design flow starts from algorithm which describes the behavior of target chip, it the first evolution. The second evolution in this domain defines finite state machines. The third evolution starts with module descripter and the last evolution deals with Boolean equation of leat cells and mask generation.

2.1 Simplified VLSI Design Flow

- Typical VLSI design flow is shown in Fig. 1.2.2.

Design Specifications

- Specification of a design is as a guide to choose the right technology and for knowing the needs of the vendor. Specifications allows each engineer to understand the entire design. It helps the engineer for designing correct interface with rest of the circuit or system. It reduces time required for design and also misassumptions if any.
- Any specification includes following information :

1. A block diagram providing details how designed chip fit into the entire system.
2. Internal block diagram for every subsection and its function.
3. Input threshold levels of all input pins and driving capability of output pins.
4. Timing specifications like setup and hold times, propagation delays and clock-cycle time.
5. Package type required.
6. Total gate count of the system under design.
7. Total power consumption of the circuit.
8. Test procedures for different tests.
9. Total cost of the target design chip.

Design Entry

User can enter a design with a schematic editor or any other text-based software tool, either a hardware description language (VHDL or VERILOG).

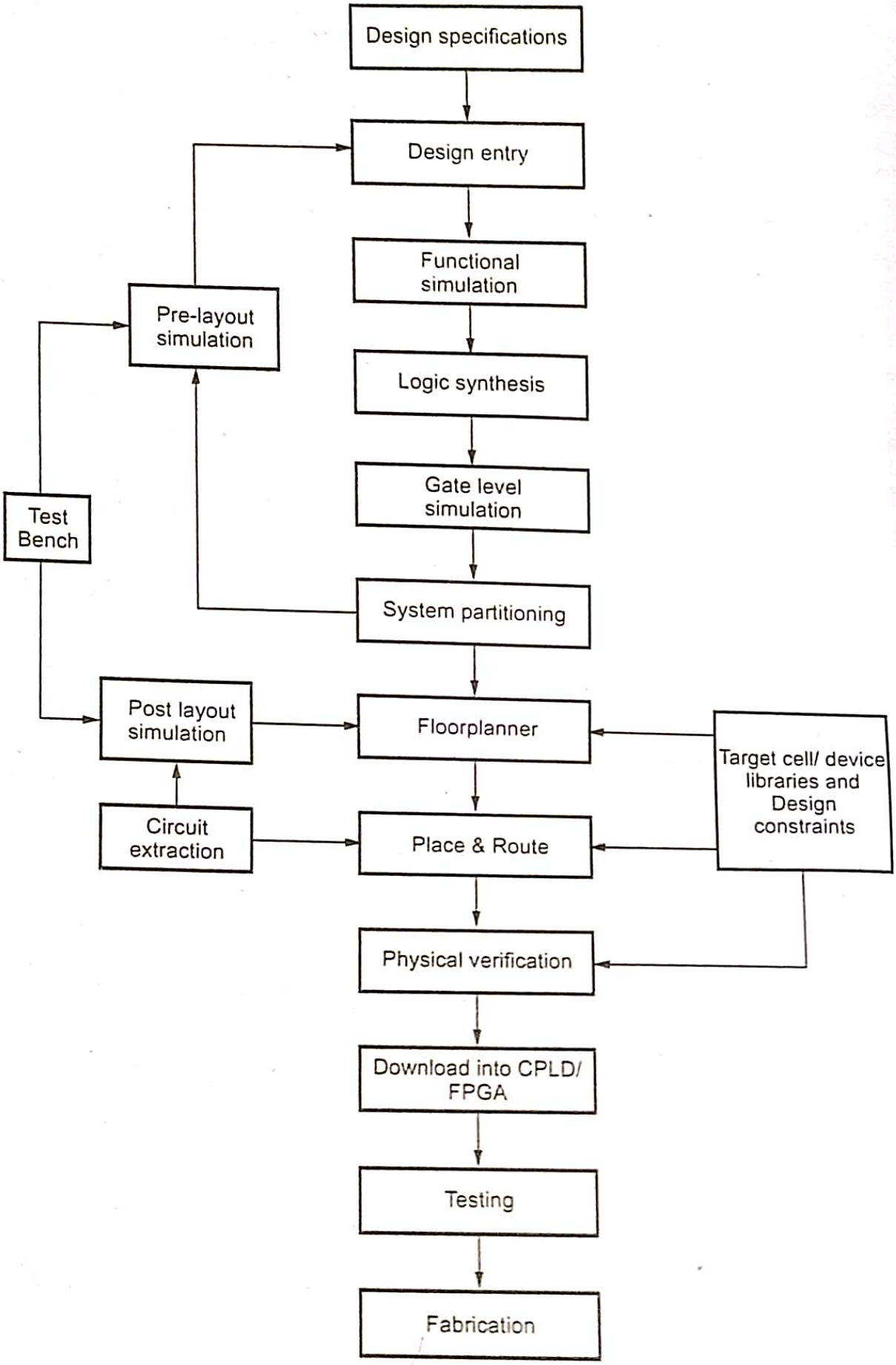


Fig. 1.2.2 VLSI design flow

matic Entry

It provides a graphical interface for design entry. A design can be build by a user with individual gates or he can combine gates to create functional blocks.

Entry

This entry supports mixed level description where gate and netlist constructs both are used along with functional descriptions.

unctional Simulation

It is the process where logic in the design is checked before user implements it in a device. As the timing information is not available at this early stage of design flow, functional simulator tests the logic of design using unit delays.

.ogic Synthesis

Here logic synthesis tool is used which produces Netlist (textual information) from synthesis process. Logic cells and their interconnections are described in detail in the Netlist. Netlist is an EDIF (Electronic Data Interchange Format) file. Thus during synthesis behavioural information in the HDL file is translated into a structural netlist.

ystem Partitioning

- System partitioning is the process of dividing a large and complex system into smaller modules.

Prelayout Simulation

- This is required for verification of a circuit design through software programs. Here stimuli is applied to design over a specific time period and recording, analyzing the respective response from the model.

Floorplanner

- The main function of floorplanner is to estimate the required chip area that will be used for each standard cell or module of the design. It is responsible for performance improvement of the design. Floorplanner is a tool that lets user generate and edit hierarchical floorplans.

) Place and Route

- After design mapping, flow engine places and routes the design. All logic blocks, including the Configurable Logic Blocks (CLB) and Input-Output Blocks (IOB) are assigned specific locations on the die at place stage.
- In the route stage, the logic blocks are assigned, particular interconnect elements on die.

Circuit Extraction

- This process determines the resistances and capacitances of all the interconnections.

x) Post Layout Simulation

- After physical place and route, this simulation is carried out. While carrying this simulation propagation delays of logic cells and interconnection delay interconnect are taken into account. If post layout simulation results ful-fill design specifications, designer can proceed for chip finishing part.

xi) Physical Verification

- After placement and routing and full custom editing physical verification is car out. It is the process of interpreting the physical layout data to determine whe it conforms to the electrical design rules, physical design rules and so schematic. Design Rule Check (DRC), Electrical Rule Check (ERC), Antenna ch and short circuit check are the processes which comes under physical verificatio

xii) Testing

- During production of chips, it is necessary to have some sort of built-in tests designed system which continuously tests the system over long period of t Chip will fail because of some electrical or mechanical problems that will usu show up with such testing procedure.

xiii) Chip Fabrication

- Before submitting design for fabrication, input-output pads should be include the design and it's connectivity should be verified.
- Then appropriate package selection for the design and selecting bonding plan the package is required. Details of how each pad of design is connected to pin of package is required.

1.2.2 Top-down and Bottom-up Design Approach

- Top-down and Bottom-up design approaches are two floorplanning methodolo These methodologies are explained in following paragraphs.

i) Top-down methodology

- This methodology is used for optimizing the design cells which are at the level of design hierarchy.
- Floorplanning top-down methodology allows user to determine how si should be bussed and assigned at the top level of the design.
- After floorplanning, place and route of top-level cells, layout of lower-level in the hierarchy is completed.
- Here an accurate estimate of layout area is generated by placement and routin
- In this methodology, user can change the floorplan easily, but layout area i accurate at each stage.

ii) Bottom-up methodology

- This methodology is used for optimizing the design cells which are at the b level of design hierarchy.

- Here first floor planning is carried out on leaf cells, which contain no underlying hierarchy.
- Important features of bottom-up approach are

a) In the higher levels of the design hierarchy, placed and routed leaf cells are considered as instances.

b) At every stage, layout area is exact, but layout process might be lengthy as more rework is necessary in order to achieve required results.

c) Global optimization of layout area is not carried out.

d) Optimization of layout area is carried out for each leaf cell.

iversity Questions

1. *Discuss VLSI design flow in detail.* **GTU : Summer-15, Winter-21, Marks 7**
2. *Describe VLSI design flow.* **GTU : Summer-21, Marks 7**
3. *Explain VLSI design flow using Y-chart.* **GTU : Winter-21, Marks 3**

3 Design Hierarchy

- For VLSI design, approach similar to software development is used where large problems are split into smaller sections until simple subroutines, with well defined functions and interfaces can be written. This is known as hierarchy or divide or 'divide and conquer' technique. It involves dividing modules into sub-modules and then repeating this process on sub-modules until complexity of smaller parts becomes manageable.
- The example of sturctural hierarchy is CMOS 4-bit adder. Fig. 1.3.1 shows the structural decomposition of 4-bit adder into its components.
- The 4-bit adder can be decomposed into 1-bit adders progressively. Then each adder is decomposed into separate sum and carry circuits and finally into individual logic.

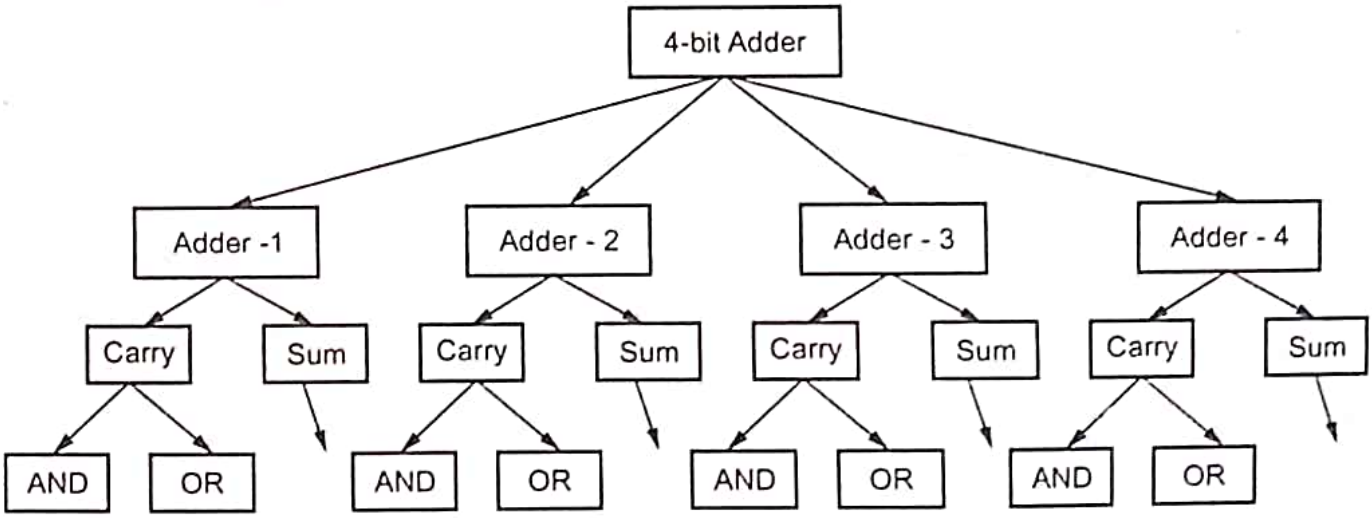


Fig. 1.3.1 Structural hierarchy of 4-bit adder

1.4 Concept of Regularity, Modularity and Locality

GTU : Summer-21, Winter-21

1. Regularity

- The hierarchical decomposition of a complex system should result in simple an similar blocks. It is called as regularity. Regularity reduces number of differen modules needed to be designed and verified. The regularity can exists at differen abstraction levels.
- An example of regularity is design of array structures consisting of identical cell such as parallel multiplication arry.

2. Modularity

- The modularity in design implies that the various functional blocks of a large system must have well-defined functions and interfaces.
- Modularity allows designer to design each module relatively independent fro each other as there is no ambignity about the function and the signal interface o these blocks.
- Modularity enables the parallelization of the design process. All the blocks can b easily combined to form a large system.

3. Locality

- The concept of locality signifies that connections are mostly between neighbourin modules and long distance connections must be avoided. It avoids lon interconnect delays.
- In certain systems the time critical operations are performed locally i.e. withou accessing modules or signals at long distance. Sometimes the replication of logic i done for local use.

University Questions

1. *Define following VLSI design concepts - regularity, modularity and locality.* GTU : Summer-21, Marks 3
2. *Explain the concept of regularity and modularity in brief.* GTU : Winter-21, Marks 3
3. *Describe : (1) Regularity (2) Modularity (3) Locality.* GTU : Summer-21, Marks 3

1.5 VLSI Design Styles

GTU : Winter-21, Summer-21

1.5.1 Programmable Logic Devices (PLDs)

- A PLD is a general-purpose chip for implementing logic circuitry. It contains collection of logic circuit elements which can be customized in different ways.

- The different versions of PLDs are explained below :

a) Programmable Logic Array (PLA)

- Several types of PLDs are commercially available.
- It provides a single level of logic e.g. array of multi-input AND gates.
- General structure of a PLA is shown in Fig. 1.5.1.

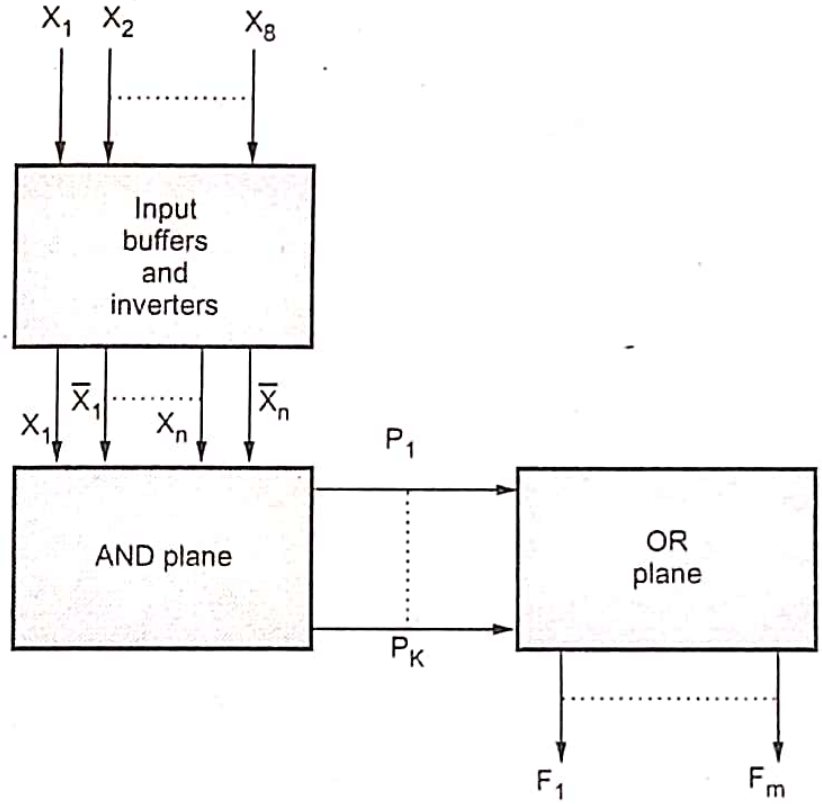


Fig. 1.5.1 General structure of a PLA

- As logic functions can be realized in sum-of-products form, a PLA comprises a collection of AND gates which feeds a set of OR gates.
- As shown in the Fig. 1.5.1, the PLA's inputs $x_1, \ldots x_n$ pass through a set of buffers into a circuit block called an AND plane, or AND array.
- AND plane produces set of product terms $p_1, \ldots p_k$.
- Each of these terms can be configured to implement any AND function of $x_1, \ldots x_n$.
- The product terms serve as the inputs to an OR plane, which produces the outputs $f_1, \ldots f_m$. Each output can be configured to realize any sum of $p_1, \ldots p_k$ and hence any sum-of-products function of the PLA inputs.
- PLA is efficient in terms of the area needed for its implementation on an integrated circuit chip. For this reason, PLAs are often included as part of larger chips, such as microprocessors.

b) Programmable Array Logic (PAL)

- In case of PLA both the AND and OR planes are programmable.
- But programmable switches presented two difficulties for manufactures of these devices : They were hard to fabricate correctly, and they reduced the speed performance of circuits implemented in the PLA.
- These drawbacks led to the development of a similar device in which the AND plane is programmable, but the OR plane is fixed. Such a chip is known as a programmable array logic (PAL) device.
- Because they are simpler to manufacture and thus less expensive than PLAs, and offer better performance, PALs have become popular in practical applications.
- An example of a PAL with three inputs, four product terms and two outputs is given in Fig. 1.5.2.

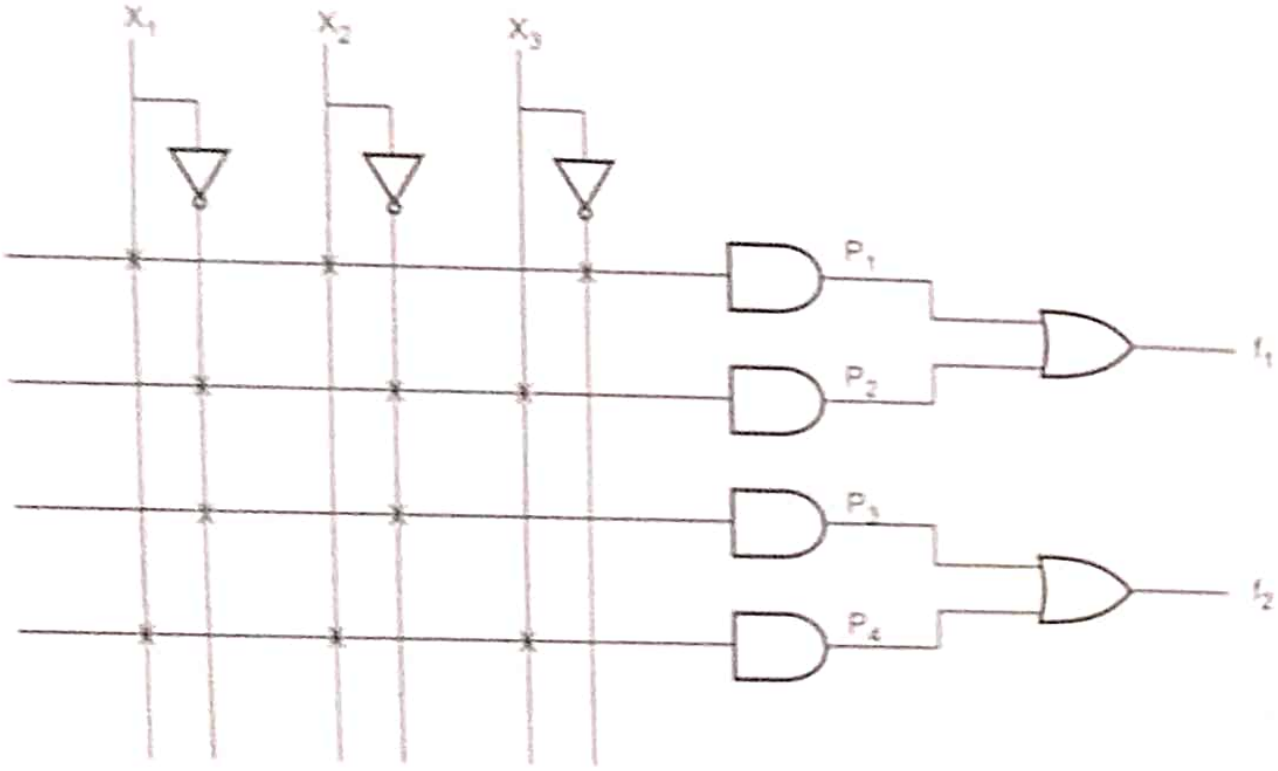


Fig. 1.5.2 An example of a PAL

- The product terms p_1 and p_2 are hardwired to one OR gate, and p_3 and p_4 are hardwired to the other OR gate.
- The PAL shown in Fig. 2.2.6 is programmed to realize two logic functions $f_1 = x_1 x_2 \bar{x}_3 + \bar{x}_1 x_2 x_3$ and $f_2 = \bar{x}_1 \bar{x}_2 + x_1 x_2 x_3$.
- Compare to PLA, PAL offers less flexibility, as PLA allows up to four product terms per OR gate, whereas OR gates in the PAL have only two inputs.

- To compensate for reduced flexibility, PALs are manufactured in a range of sizes, with various numbers of inputs and outputs and different numbers of inputs to the OR gates.
- In many PALs extra circuitry is added at the output of each OR gate to provide additional flexibility. It is customary to use the term macro cell to refer to the OR gate combined with the extra circuitry.
- A number of companies manufacture PLAs or PALs or other similar types of simple PLDs (SPLDs).

Programmable Read Only Memory (PROM)

- Many different types of memory blocks exist. In a ROM stored information can be read out of the storage cells, but it cannot be changed.
- Another type of ROM allows information to be read out of storage cells and stored, or written into them.
- Reading its contents is the normal operation, whereas writing requires a special procedure. Such a memory block is called a programmable ROM (PROM).
- The storage cells in a PROM are usually implemented using EEPROM transistors.
- PROM consists of fixed AND array and programmable OR array with programmable interconnect.

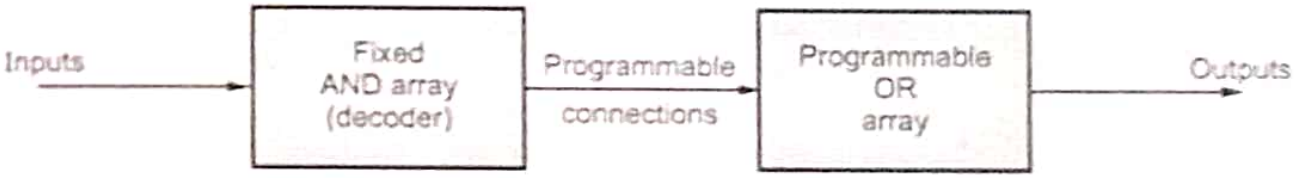


Fig. 1.5.3 PROM block diagram

Programmable Logic Sequencer (PLS)

- Provides two levels of logic, usually an AND/OR array. Programmed by blowing fusible links at specified points.
- In a PLD, software is used to decide how to connect the groups of transistors (gates) or the groups of gates (macro cells) together in a particular way.
- PLDs are programmed in three ways
  i) Connections between gates are linked by physically breaking fuses.

ii) Alternatively, connections are made by melting some lines.

iii) Third method connects the gate by using software controlled switches within the PLDs.

Examples on PALs, PLAs, PGAs and PLDs :-

Example for PLAs

- Every possible connection is made when PLA is manufactured and desired connections are removed by blowing fuses, this type of PLA is called FPLA.
- In field programming logic array user can program active level of each output therefore programming FPLA consists of :
  - Programming AND matrix
  - Programming OR matrix
  - Programming active level of each output.
- Let's see how to implement a circuit for code conversion from XS3 (Excess-3) to NBCD (Natural BCD) into PLA.
- The truth table for excess-3 to BCD conversion is shown below.
- Truth table

XS3				NBCD			
A	B	C	D	P	Q	R	S
0	0	1	1	0	0	0	0
0	1	0	0	0	0	0	1
0	1	0	1	0	0	1	0
0	1	1	0	0	0	1	1
0	1	1	1	0	1	0	0
1	0	0	0	0	1	0	1
1	0	0	1	0	1	1	0
1	0	1	0	0	1	1	1
1	0	1	1	1	0	0	0
1	1	0	0	1	0	0	1

- **Minimization using K-maps**

'ϕ' represents the "can't happen" terms in XS3 code and are used in simplification rocess.

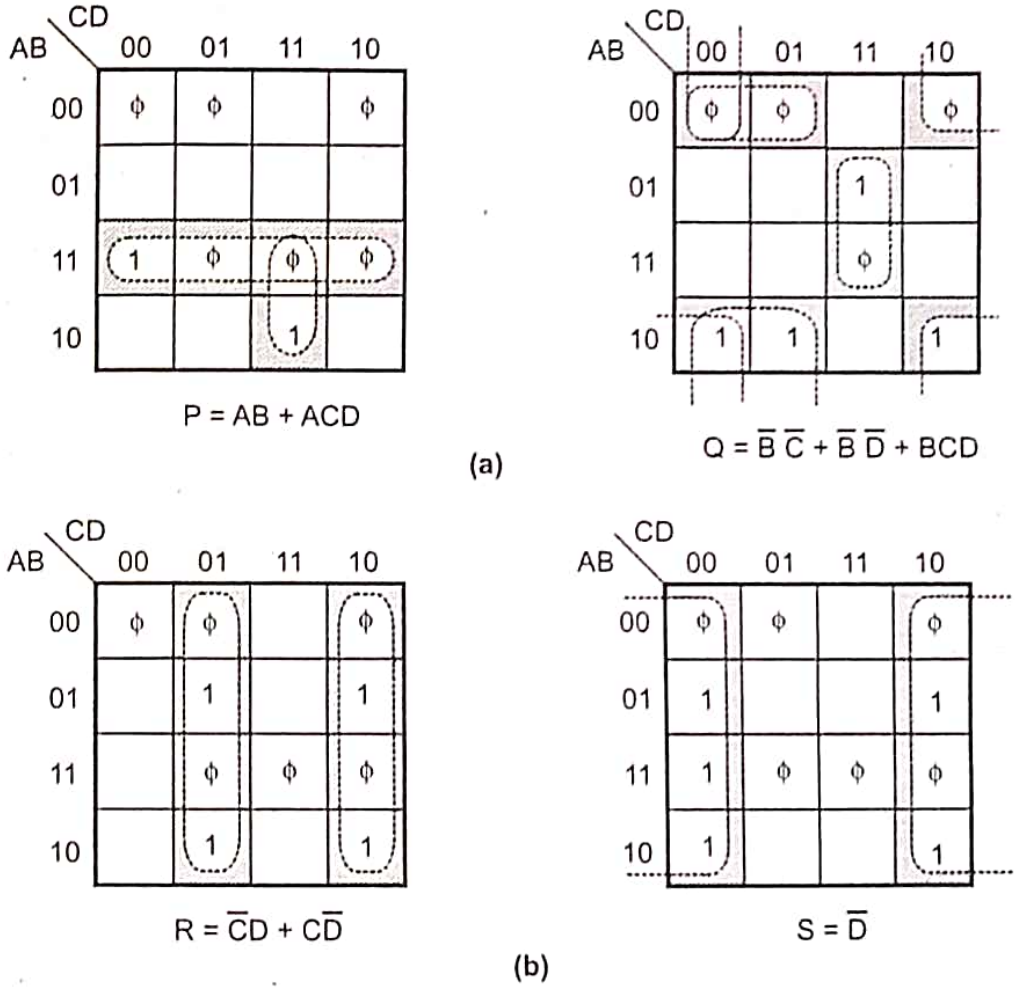


Fig. 1.5.4

ıplementation

- There are eight separate products in the four equations. Since there are four input variables and eight product terms, PLA must provide eight 8-input AND gates, four 8-input OR gates and four XOR gates i.e.

$$8'8)+(4'8)+4 = 100$$

$$64+32+4 = 100 \text{ fusible links}$$

- Implementation of above logic in PLA can be done as shown in Fig. 1.5.5.

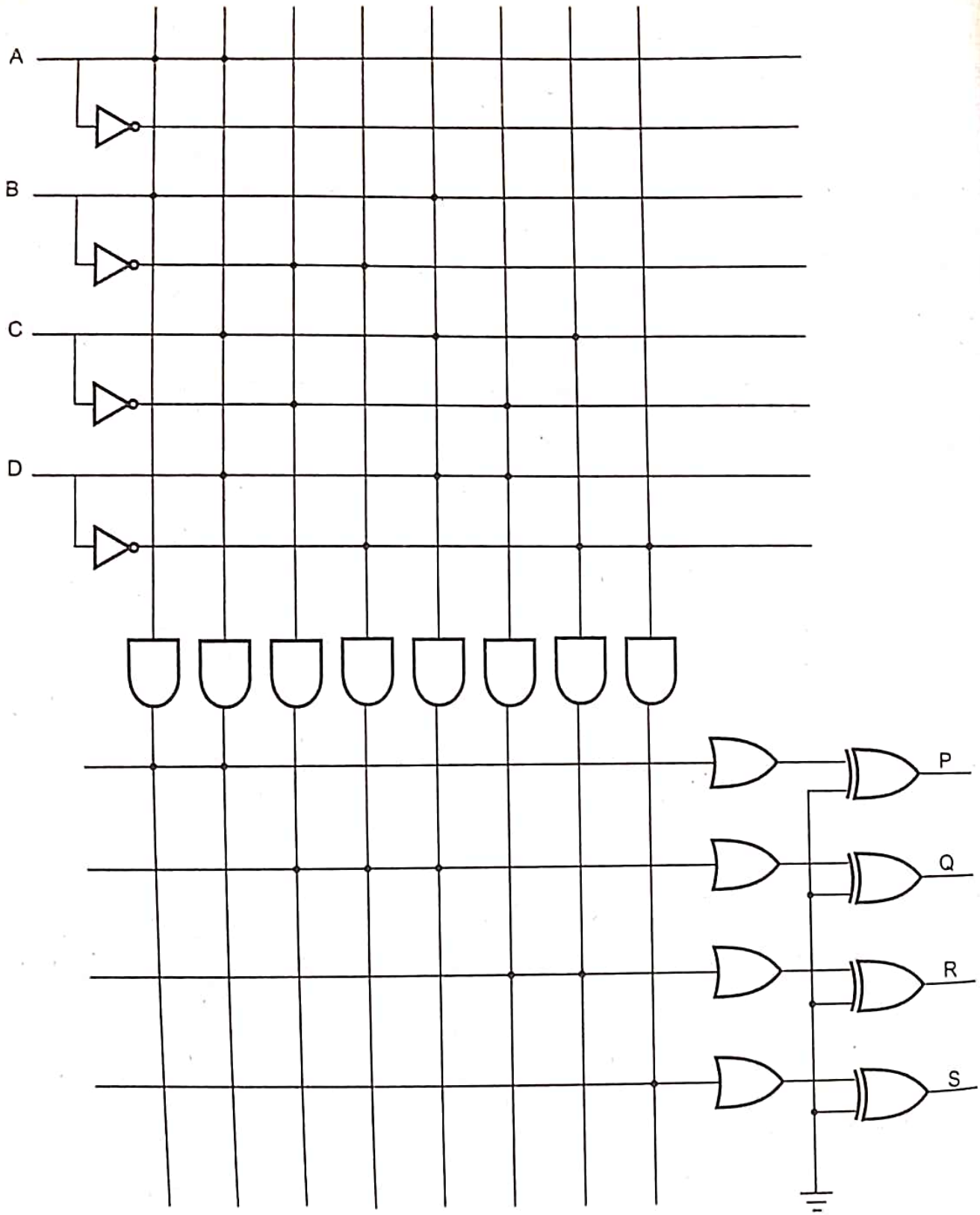


Fig. 1.5.5 Implementation of logic in PLA

- To implement the same four functions using a programmable ROM would requi a storage capacity of 128 bits and 128 fusible links, one for each single-bit memo element.

1.5.2 Field Programmable Gate Array (FPGA)

- A FPGA is a programmable logic device which supports implementations of relatively large logic circuits.
- FPGAs offer a number of configurable logic blocks (CLBs) which contains programmable combinational logic and registers for sequential circuits.
- There is also a set of input/output blocks which can be configured as fixed input, fixed output or bidirectional.
- Outputs have tristate nature and registers can be used for latching incoming or outgoing data.

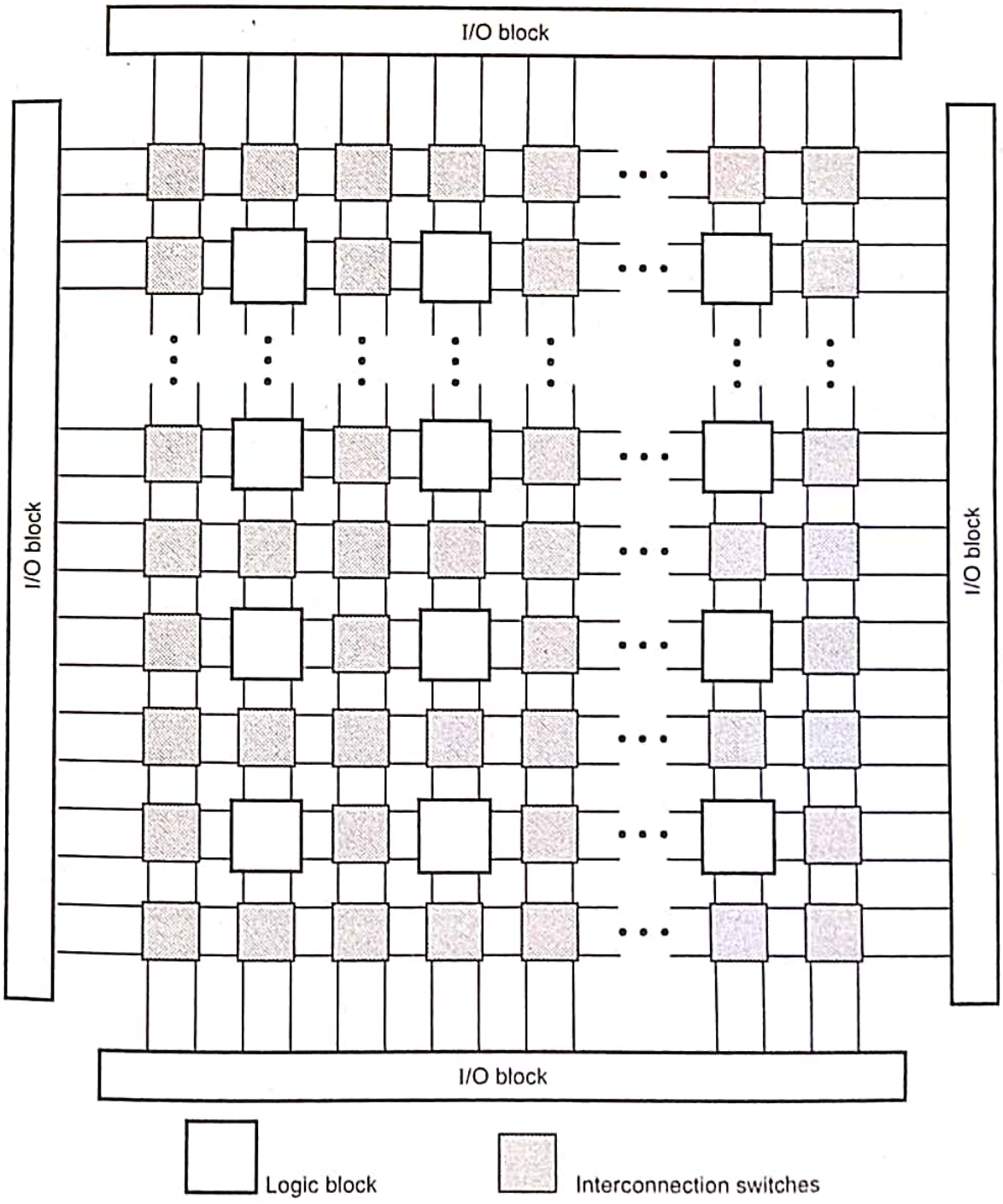
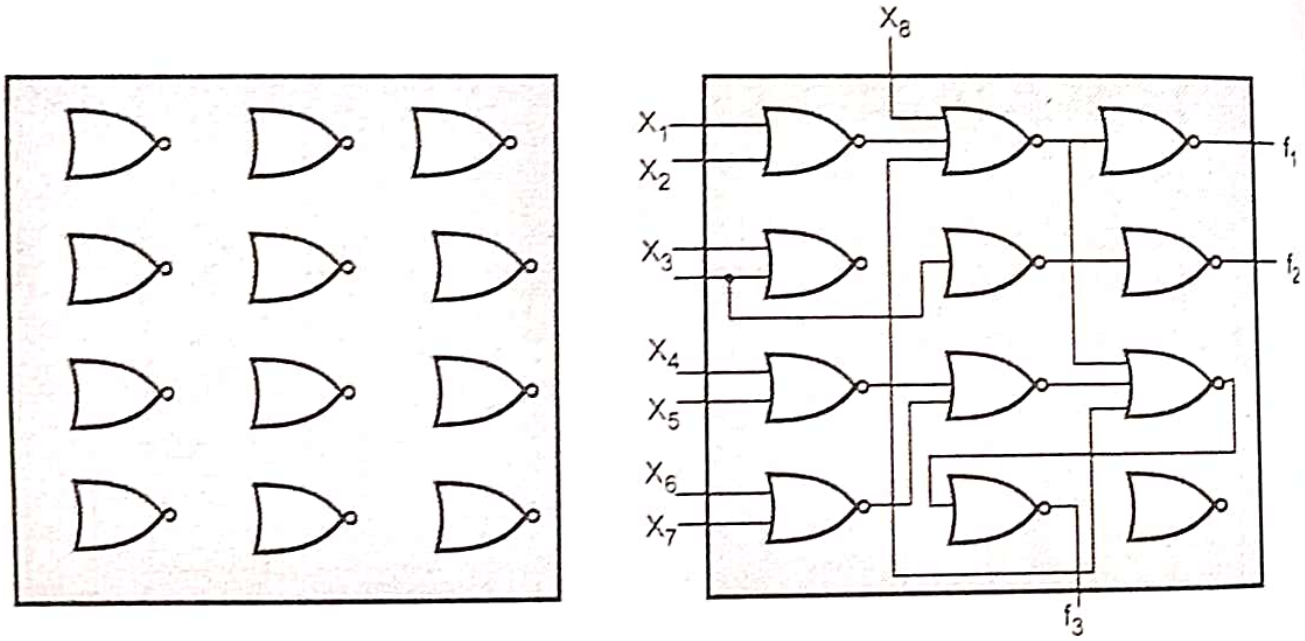


Fig. 1.5.6 General structure of FPGA

- FPGAs use a grid of logic gates, similar to that of an ordinary gate array, bu programming is done by the customer, not by the manufacturer.
- In FPGA, all the configurable logic blocks and input/output blocks can b programmably interconnected to implement virtually any logic circuit.
- FPGAs are generally programmed after being soldered down to the circuit board in the same way as larger CPLDs.
- In larger FPGAs the configuration is volatile and must be re-loaded into the devic whenever power is applied or different functionality is required.

1.5.3 Gate Array Design

- A gate array is an IC chip on which gates are placed in matrix form without an connection among the gates, as shown in Fig. 1.5.7. (a) By connecting gates, us can realize logic networks, as shown in the Fig. 1.5.7. (b) But actually logic gat are not realized in a gate array.



(a) Before making connection

(b) After making connection

Fig. 1.5.7

- Instead of gates, cells each of which consists of unconnected components, i arranged in matrix form, and each cell can realize one of few different types gates by connecting these components.
- By connecting these gates as shown in Fig. 1.5.7 (b), networks can be realiz Only two or three masks for connections and contact windows have to custom-made, instead of all two dozen masks required for full-custom design.
- Because only the connection layout, along with the placement of gates, needs to considered, CAD can be effectively used, thereby greately reducing the lay time.

- Thus, the design with gate arrays is very inexpensive and quick, compared to full custom design. Gate arrays of CMOS have been extensively used in many computers.
- In Fig. 1.5.8, connections among gates are run in narrow strips of space between columns or rows of gates. These strips of space are nothing but routing channels.

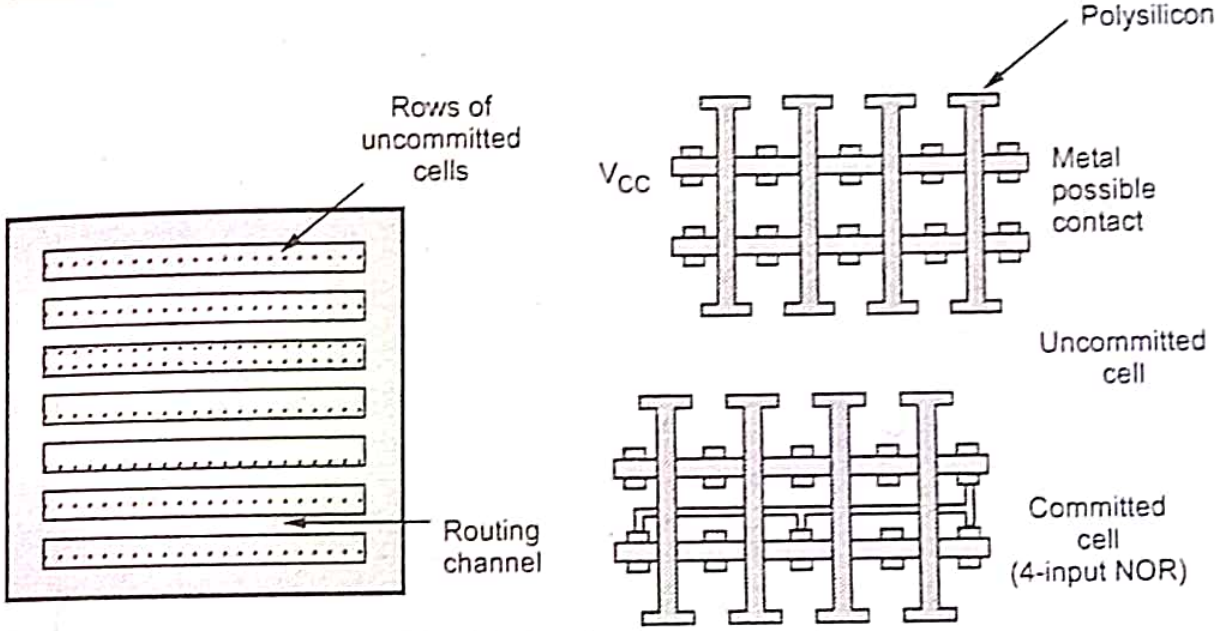


Fig. 1.5.8 Gate array network

- Types of gate arrays
  a) Channeled gate arrays.
  b) Channel less gate array or sea of gate arrays.
  c) Structured gate array.

1.5.4 ASIC

- ASIC stands for **Application Specific Integrated Circuit.**
- Application Specific Integrated Circuit (ASIC) is a full-custom circuit. In which every mask is defined by the customer or a semi-custom circuit (gate array) where only a few masks are defined.
- ICs are made on a wafer. Circuits are built up with successive mask layers. The number of masks used to define the interconnect and other layers is different between full-custom ICs and programmable ASICs.
- ASIC is the custom design integrated circuit for particular application such as processing unit of mobile phone or any electronic toy etc.
- An ASIC is a device which is manufactured by an ASIC vendor in generic form i.e. ASICs, more specifically, are designed by the end user to perform some proprietary application.

- This initial manufacturing process is the most complex, time consuming an expensive part of total manufacturing process. The result is silicon chips with a array of unconnected transistors.
- The final manufacturing process of connecting the transistors together is the completed when a chip designer has a specific design wishes to implement in the ASIC. An ASIC vendor can usually do this in couple of weeks and is known as the **turnaround time.**

A silicon chip or **integrated circuit** (IC) is more properly called a die

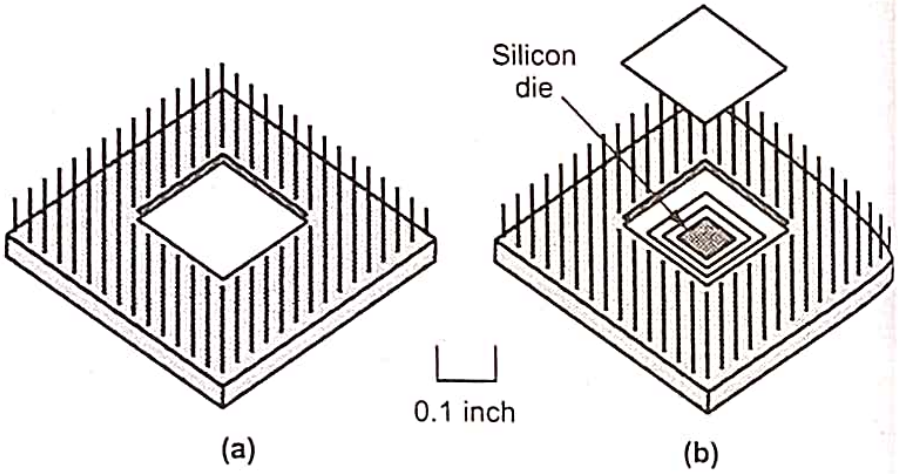


Fig. 1.5.9

- ASIC is useful for high volume production.

1.5.4.1 Types of ASIC

- ASICs can be divided into two basic categories :
  1) Full custom ASICs. 2) Semi-custom ASICs.
- Fig. 1.5.10 shows categories of ASIC and its subtypes.

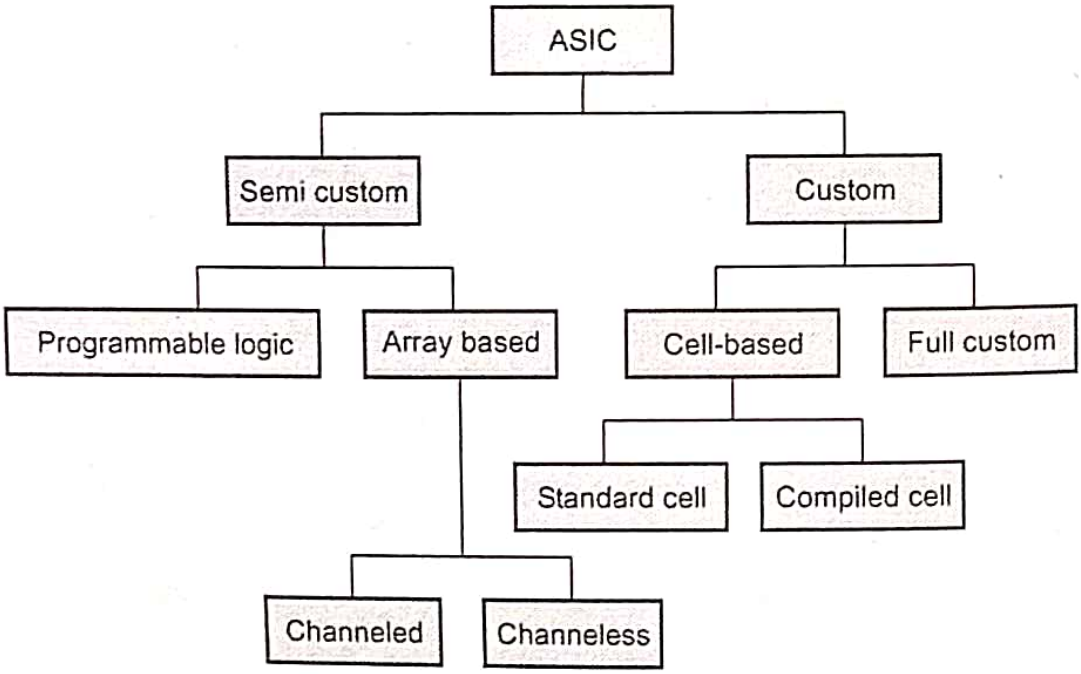


Fig. 1.5.10 ASIC classification

5.4.2 Full Custom ASIC

- A full custom ASIC all logic cells are customized by the designer.
- Bipolar technology has used widely for full custom analog design because of it's improved precision.
- CMOS technology is used for full custom ASIC because it is easier to design analog as well as digital circuits on the same chip with better performance.
- These are designed and processed like standard products with all sets of masks for all fabrication layers.
- A standard cell library contains hundreds of different logic cells including combinational functions with multiple inputs as well as latches and flip-flops with different combinations of reset, preset and clock options. It only makes sense to design a full-custom IC if there are no libraries available.
- In a full custom design, the entire mask design is done a new without use of any library. However, the development cost of such a design style is becoming prohibitively high.
- Thus, the concept of design reuse is becoming popular in order to reduce design cycle time and development cost.
- The most popular full custom design can be the design of a memory cell, be it static or dynamic. Since the same layout design is replicated, there would not be any alternative to high density memory chip design.
- In real full custom layout in which the geometry, orientation and placement of every transistor is done individually by the designer, design productivity is usually very low typically 10 to 20 transistors per day, per designer.
- Full-custom offers the highest performance and lowest part cost (smallest die size) with the disadvantages of increased design time, complexity, design expense and highest risk.
- Common examples of full-custom ICs or ASICs are requirements for high-voltage (automobile), analog /digital (communications) or sensors and actuators.

5.4.3 Semi-custom ASIC

- Semi-custom ASICs are either partly or sometimes fully prefabricated or they are structured like standard and custom ICs.
- Few mask patterns are predesigned.
- Semi-custom types of ASICs can be classified as :
  1. Standard cell based ASICs
  2. Gate arrays
  3. Mixed signal and analog ASIC
  4. Programmable Logic Devices (PLDs)

1.5.4.4 Comparison of Full Custom and

Sr. No.	Parameters	Full custom	Standard cell	Semicustom	FPGA
				Gate array	
1.	Area	Compact	Between compact and moderate	Moderate	Large
2.	Cell size	Variable	Fixed height	Fixed	Fixed
3.	Type of cell	Variable	Variable	Fixed	Programmab
4.	Area	Compact	Compact to moderate	Moderate	Large
5.	Interconnections	Variable	Variable	Variable	Programmab
6.	Performance	High	High to moderate	Moderate	Low
7.	Fabrication layer	All layers	All layers	Only routing layers	No layers
8.	Design cost	High	Medium	Medium	Low

1.5.5 Standard Cell based ASICs

- Here predesigned logic cells like gates, multipliers, flip-flop etc. known standard cells are used in standard-cell based ASICs.
- Here ASIC designer defines only the placement of the standard cells a interconnection in a CBIC (Cell Based Integrated Circuits).
- However, standard cells can be placed anywhere on the silicon that means ent mask layers of CBIC are customized and are unique to a particular customer.
- Advantage of CBIC is that designer can save time and reduces the risk by maki use of predesigned standard cell library.
- A standard cell requires less silicon for a given function as macros are ve compact.
- Important features of this ASICs are :
  a) All mask layers are customized transistors which are interconnected.
  b) Custom blocks are embedded.
  c) Manufacturing lead time is about weeks.
- Each standard cell in a library is constructed using full custom design metho The standard cells are designed to fit together like bricks. Power and GND bu run horizontally on metal lines inside the cells.

- Standard cell design allows the automation of the process of assembling an ASIC. Groups of standard cells fit horizontally together to form rows. The rows stack vertically to form flexible rectangular blocks.
- Standard cell structure is shown in Fig. 1.5.11.

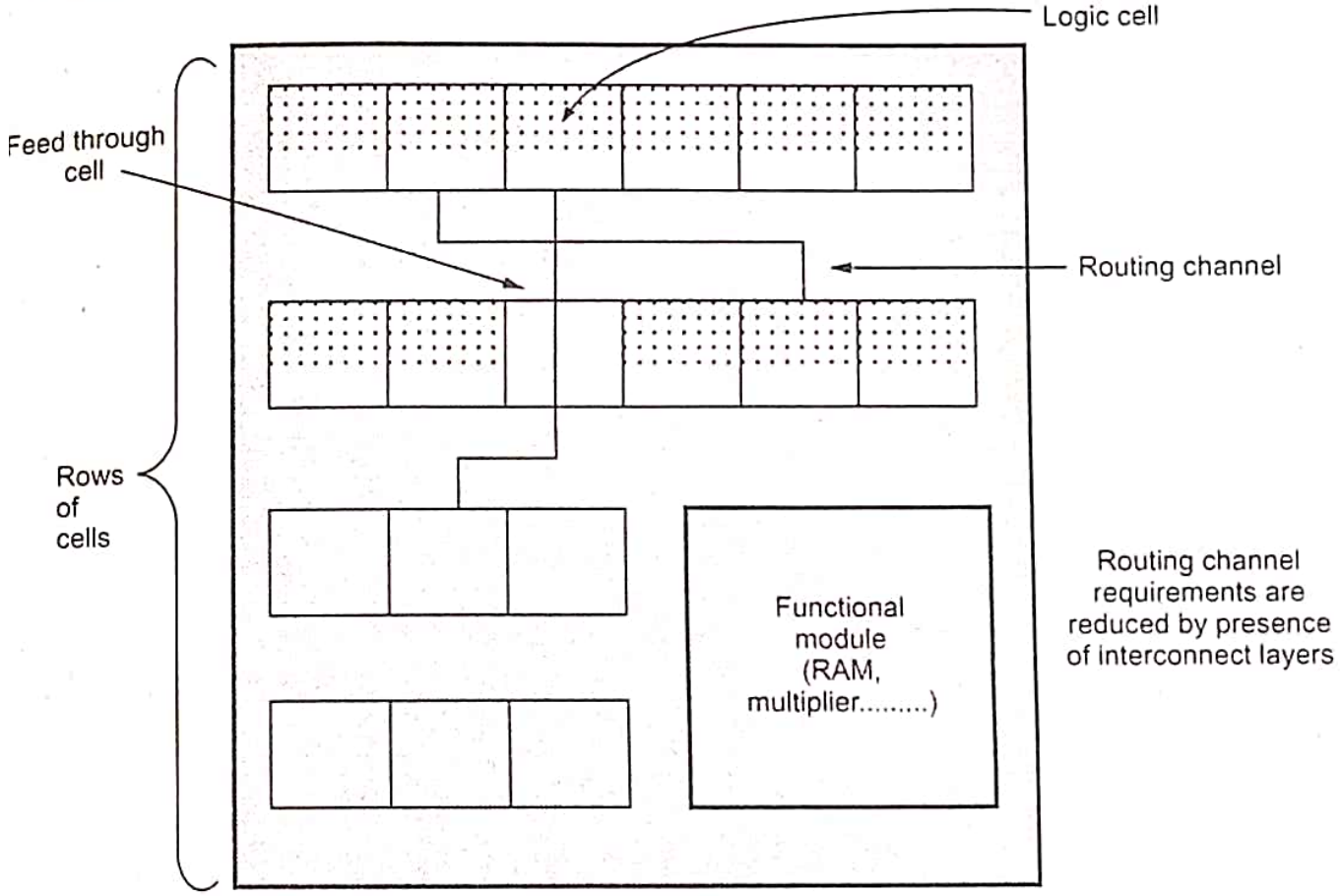


Fig. 1.5.11 Standard cell structure

- A typical library may contain a few hundred cells including inverters, NAND gates, NOR gates, complex AOI, OAI gates, D latches and flip-flops. Each gate type can have multiple implementations to provide adequate driving capability for different fanouts.
- For example, the inverter gate can have standard size transistors, double size transistors and quadruple size transistors so that chip designer can choose the proper size to achieve high circuit speed and layout density.
- Characterization of each cell is done for several different categories. It consists of,
  a) Delay time Vs load capacitance. b) Circuit simulation model.
  c) Timing simulation model. d) Fault simulation model.
  e) Cell data for place-and-route. f) Mask data.
- To enable automated placement of the cells and routing of inter-cell connections, each cell layout is designed with a fixed height, so that a number of cells can be abutted side-by-side to form rows.

- Power and ground lines run parallel to the upper and lower boundries of the c thus, neighbouring cells share a common power and ground bus.
- The input and output pins are located on the upper and lower boundries of cell.
- In many VLSI chips, such as microprocessors and digital signal processing ch standard-cells based design is used to implement complex control logic modu Some full custom chips can be also implemented exclusively with standard cells

1.5.6 System on Chip (SoC)

- SoC is an idea of integrating all components of any electronic system into a si integrated circuit (chip).
- The SOC may include digital, analog, mixed-signal and often radio-freque functions - all on one chip.
- A typical application is in the area of embedded systems and in mobile system

University Questions

1. *Discuss the impact of different VLSI design styles upon the design cycle time and the achiev circuit performances.* **GTU : Winter-21, Marks**
2. *Differentiate full custom and semi custom design flow.* **GTU : Summer-21, Winter-21, Marks**

1.6 Design Quality

GTU : Winter-14, Summer-

- The design quality is measured in order to improve chip design. the impo criteria for quality of design are :
  1. Testability
  2. Yield and manufacturability.
  3. Reliability
  4. Technology updation.

1. Testability

- The fabricated chip should be tested to ensure correct funtionality without fai Then the chip is inserted into package or base die. This requires following goa
  1. Generation of good test vectors.
  2. Availability of reliable test fixture at speed.
  3. Design of testable chip.

2. Yield and manufacturability

- The chip yield is computed by dividing number of good tested chips by number of tested chips. The chip yield is affected due to -
  - mishandling of waferchip.
  - uncontrollable process variations

- Also poor chip design can cause processing problems resulting in drop-outs during manufacturing.
- The chip yield can be categerized into two types
  a. Functional yield
  b. Parametric yield
- The functional yield indicates the problem of shorts, opens, leakage currents and design fants. The functional yield is obtained by testing the functionality of chip at lower speed than specified.
- The parametric yield performs delay testing, variation in chip performance and manufacturing faults testing

3. Reliability

- The reliability of chip depends on design and process conditions. The chip reliability causes are characterized by -

1. Electrostatic discharge (ESD)	2. Electrical overstress (EOS)
3. Electronigration	4. Latchup in CMOS I/O
5. Hot-carrier induced aging.	6. Oxide breakdown.
7. Onchip noise and crosstalk.	8. Power and ground bouncing.

4. Technology updation

- The chip product has to be technology updated to new design rules. The updation of technology should be done without any change in chip's functionality. Thus the design style should be chosen such that the technology update of chip or frictional modules for design reuse can be achived with least cost and quickly.

University Questions

1. *Which are the four general criteria to measure design quality of a fabricated integrated circuit (chip) ? Briefly explain each of them.* **GTU : Winter-14, Marks 7**
2. *Briefly describe yield and manufacturability.* **GTU : Summer-15, Mark 1**

1.7 Package Technology

- Due to advent of devices with high frequency, high power dissipation leads to introduction of new packages. IC packaging can be classified into following major categories :
  i) Surface mounting package

ii) Chip-scale packaging

iii) Bare die

iv) Through-hole packages

v) Module assemblies

1.7.1 Surface Mounting Package

- Surface mounting packages are available in plastic packages, ceramic packages a metal packages. Plastic packages are light weight, small, able to withstand sh and inexpensive. Plastic packages are hydroscopic and absorb moisture. T moisture can vaporize during heating but causes package cracking (popc effect). Also moisture can allow contaminants to enter the IC and cause failure a later time.
- Ceramic packages are preferred in extreme environmental conditions such military and space where shorter life and operation is not compromised. reliability of ceramic package makes them most suitable for critical communicat navigation and avionic system.
- Surface mounting packages are available in
  - Ball Grid Array (BGA)
  - Quad Flat Packs (QFP)
  - Biometric Package
  - Ceramic Column Grid Array (CCGA)
  - Ceramic Leadless Chip Carrier (CLCC)
  - Vertical Mount Package (VPAK)

1.7.2 Chip-Scale Packaging

- Chip scale packaging are small, easy for handling and testability. In flip-chip of packages I/O terminations are in the form of bumps on one side of pa (bumped chip).
- Chip scale packaging are available in
  - Bottom Leaded Package (BLP)
  - Flip Chip Attach (FCA)
  - Micro BGA
  - Micro Pak.

1.7.3 Bare Die

- These are compact sized with no signal delay package. Available bare die packages are
  - Controlled Collapse Chip Connection BGA (C4BGA)
  - Chip Scale Package (CSP)
  - Chip On Board (COB)
  - Tape Automated Bonding (TAB)

1.7.4 Through-Hole Packages

- Through-hole packages are mounted on plated holes on PCB. Available through-hole packages are
  - Ceramic Dual In-line Package (CER DIP)
  - Pin Grid Array (PGA)
  - Quad In-line Package (GUTP)
  - Single In-line Module (SIM)
  - Water On Water (WOW)
  - Vertical In-line Package (VIL)

1.7.5 Module Assemblies

- It uses packaged parts or bare die for assembly. It can be mounting parts on substrate as surface mount or through-hole. Available module assemblies are
  - Dual In-line Memory Module (DIMM)
  - Multichip Module (MCM)
  - Memory Cube
  - Single In-line Memory Module (SIMM)

1.8 Architecture of FPGA

GTU : Summer-21

- The building blocks of an FPGA organized as a 2-D array of cells is shown in Fig. 1.8.1. The basic architecture of FPGA consists of -
  1. Configurable Logic Block (CLB)
  2. I/O block
  3. Programmable interconnect.

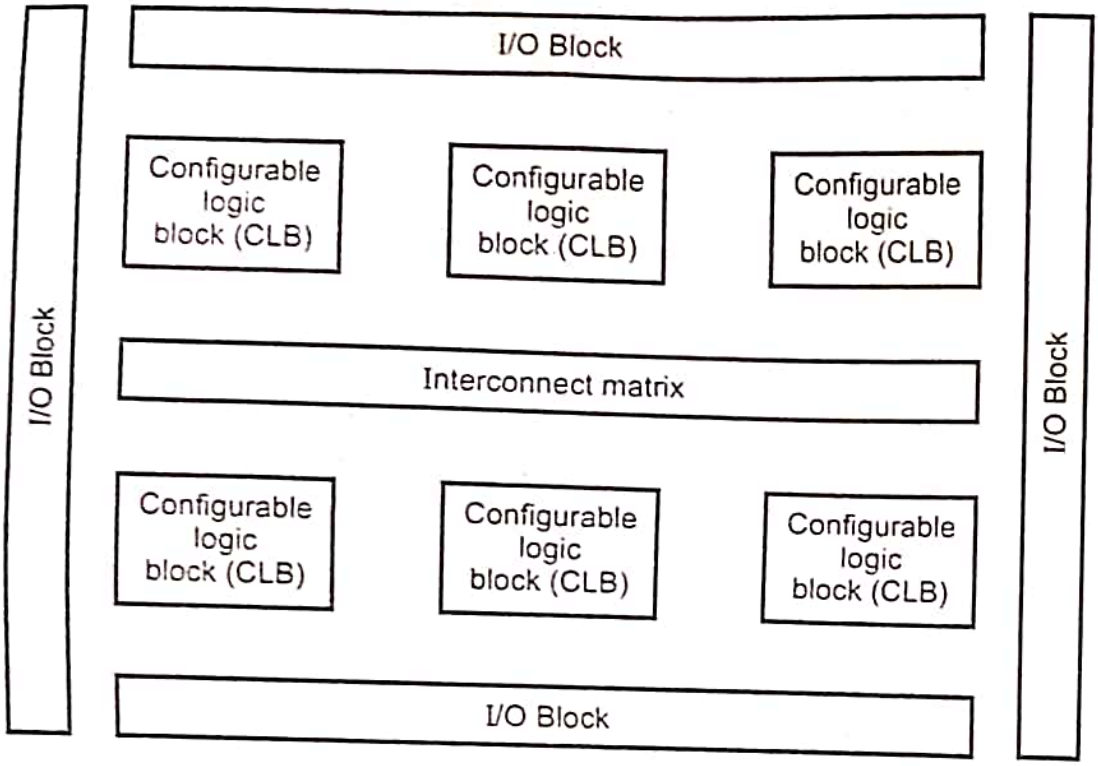


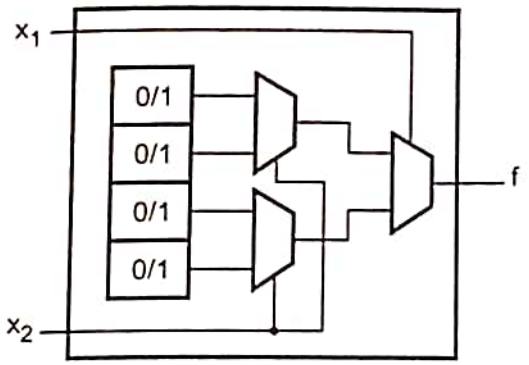
Fig. 1.8.1 Architecture of FPGA

1.8.1 Configurable Logic Block (CLB)

- Each logic block is an FPGA typically has a smaller number of inputs and outputs. A variety of FPGA products are on the market, featuring different types of logic blocks.
- The most commonly used logic block is a lookup table (LUT), which contains storage cells that are used to implement a small logic function.
- Each cell is capable of holding a single logic value, either 0 or 1. The stored value is produced as the output of the storage cell.
- LUTs of various sizes may be created, where the size is defined by the number of inputs.

Two Input LUT

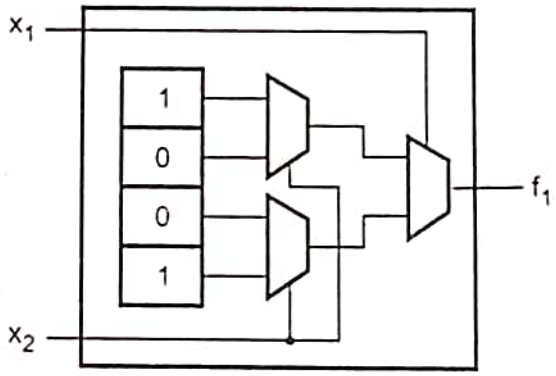
- Fig. 1.8.2 shows the structure of a small LUT. It has two inputs, x_1 and x_2, and one output, f.
- It is capable of implementing any logic function of two variables. Because a two-variable truth table has four rows, this LUT has four storage cells. One cell corresponds to the output value in each row of the truth table.
- The input variables x_1 and x_2 are used as the select inputs of three multiplexers, which depending on the valuation of x_1 and x_2, select the content of one of the four storage cells as the ouptut of the LUT.



(a) Circuit for a two-input LUT

x_1	x_2	f_1
0	0	1
0	1	0
1	0	0
1	1	1

(b) $f_1 = \bar{x}_1\bar{x}_2 + x_1x_2$



(c) Storage cell contents in the LUT

Fig. 1.8.2 Circuit for a two-input LUT

- To see how a logic function can be realized in the two-input LUT, consider the truth table shown in Fig. 1.8.2 (b). The function f_1 from this table can be stored in the LUT as illustrated in Fig. 1.8.2 (c).
- The arrangement of multiplexers in the LUT correctly realizes the function f_1. When $x_1 = x_2 = 0$, the output of the LUT is driven by the top storage cell, which represents the entry in the truth table for $x_1\ x_2 = 00$.
- Similarly, for all valuations of x_1 and x_2, the logic value stored in the storage cell corresponding to the entry in the truth table chosen by the particular valuation appears on the LUT output.
- Providing access to the contents of storage cells is only one way in which multiplexers can be used to implement logic functions.

ıree Input LUT

- Fig. 1.8.3 shows a three input LUT. It has eight storage cells because a three-variable truth table has eight rows.
- In commercial FPGA chips, LUTs usually have either four or five inputs, which require 16 and 32 storage cells, respectively.

- Some FPGAs, usually have extra circuitry, besides a LUT, in each logic block. Fig. 1.8.4 shows how a flip-flop may be included in an FPGA logic block.
- For a logic circuit to be realized in an FPGA, each logic function in the circuit must be small enough to fit within a single logic block. In practice, a user's circuit is automatically translated into the required form by using CAD tools.

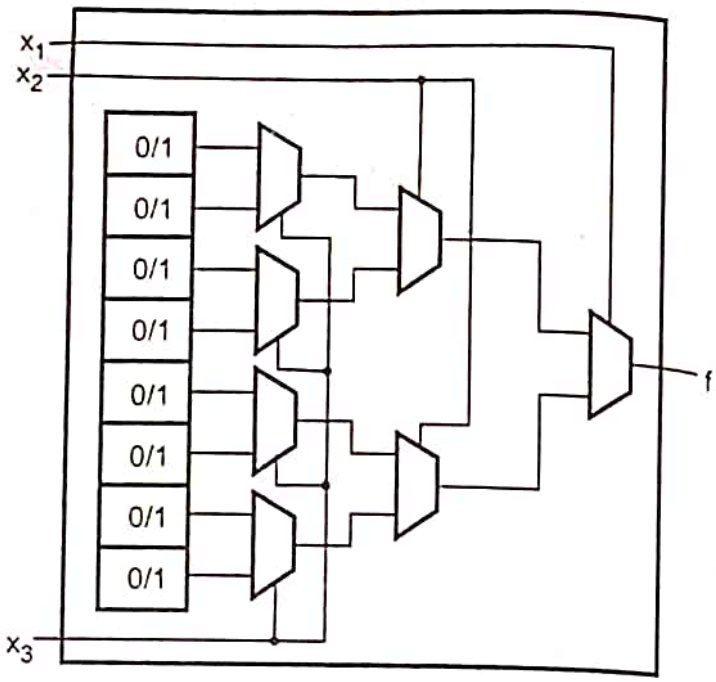


Fig. 1.8.3 A three input LUT

- When a circuit is implemented in an FPGA, the logic blocks are programmed to realize the necessary functions and the routing channels are programmed to make the required interconnections between logic blocks.

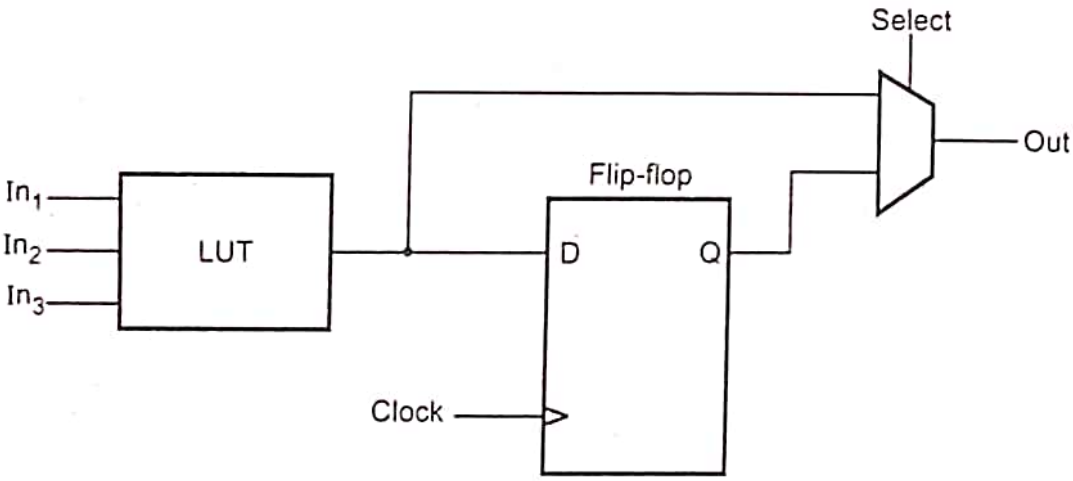


Fig. 1.8.4 Inclusion of flip-flop in an FPGA logic block

1.8.2 I/O Blocks

- Each bank can be configured individually to support a particular I/O standard.
- Allows the FPGA to work with device using multiple I/O standard.
- FPGA can actually be used to interface between different I/O standards.
- Modern FPGA output signals with fast edge rates requires termination to prevent reflections and maintain signal integrity.
- High pin count package cannot accommodate external termination registers.
- Thus a Digitally Controlled Impedance (DCI) is employed.
- DCI eliminates the need for external register and improves signal integrity.

1.8.3 Programmable Interconnect

- A programmable switch matrix form the heart of interconnect in a FPGA. The actual switching matrix employed a structure of six pass transistors per cross point. Thus connectivity can be established by controlling the transistors and various type of connections.
  1) **Single lines :** Used to connect a CLB to another CLB is one hop away. These wires have to go through a programmable switch hence add delay.
  2) **Double lines :** These wires travel past two CLB before hitting the switch, hence they provide shorter delays for longer connections.
  3) **Long lines :** Wires in long groups do not go through any programmable switch at all, instead they travel all the way across or down a row or column and are driven by tri-state drivers.
  4) **Global clock lines :** These lines are optimized for case as clock inputs to the CLB, providing short delay and minimal skew.

niversity Questions

1. *Describe FPGA architecture.* **GTU : Summer-21, Marks 4**
2. *Sketch PMOS fabrication steps.* **GTU : Summer-21, Marks 7**

.9 Complex Programmable Logic Device (CPLD)

GTU : Winter-14, 21, Summer-21

mplex PLDs (CPLDs)

- Complex Programmable Logic Device (CPLD) is an erasable programmable logic device which can be programmed with a schematic or behavioural design. CPLD is nothing but a type of complex PLD based on EPROM or EEPROM technology.
- Basic CPLD cell is called a macrocell, which is CPLD implementation of a CLB. It is composed of AND gate arrays and is surrounded by interconnect area.
- They are characterized by an architecture offering high speed, predictable timing and simple software.
- Also referred as multilevel arrays, devices which combine number of PAL type circuits on the same chip.

Logic blocks are programmable AND and fixed OR structure with fewer product terms than most PAL devices.

PALs and GALs are available in small sizes, equivalent to few hundred logic gates. For bigger circuits (logic), complex PLDs or CPLDs can be used.

CPLDs can replace thousands, or even hundreds of thousands of logic gates.

- Some CPLDs are programmed using a PAL programmer but this method resul inconvenient for devices with hundreds of pins.
- Second method of programming is to solder the device to its printed circuit boar then feed it with a serial data stream from personal computer.
- CPLD contains a circuit which decodes the data stream and configures the CPL to perform its specified logic function.

1.9.1 Architecture of CPLD

- A CPLD consists of multiple circuit blocks on a single chip with internal wirin resources to connect the circuit blocks.
- The architectural building blocks of CPLD :
  1. PAL blocks (also called an function blocks)
  2. Interconnect matrix
  3. Input /output (I/O) blocks.
- Fig. 1.9.1 shows generic CPLD architecture.

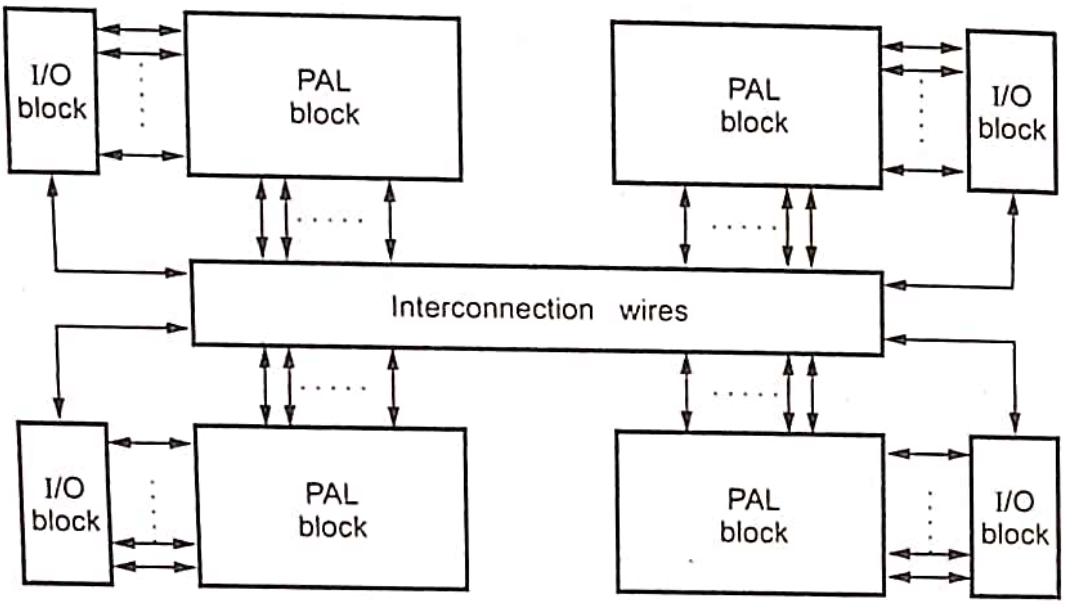
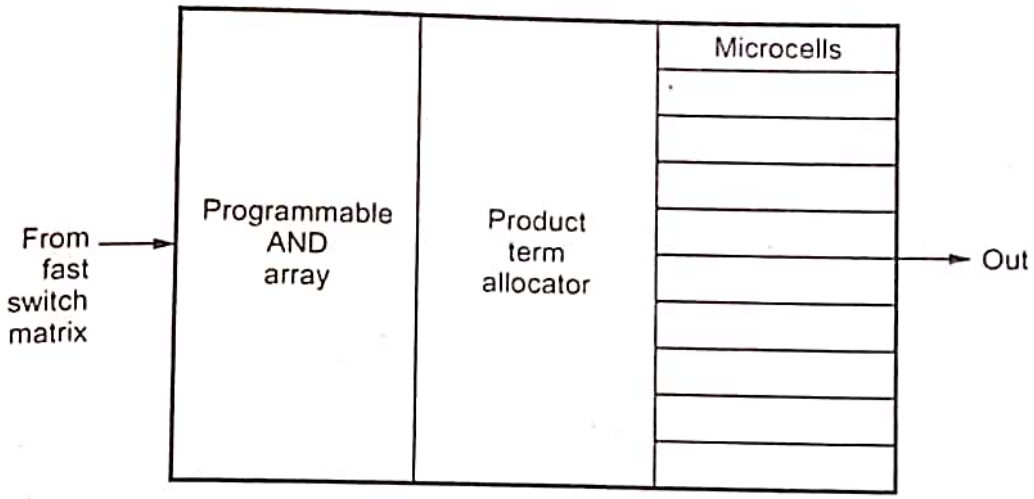


Fig. 1.9.1 Structure of CPLD

1) Function block :

- AND plane can accept inputs from the I/O blocks, other function blocks feedback from the same function block.
- The terms are then ORed together using a fixed number of OR gates and term are selected via a large multiplexer.
- The output of the MUX can then be sent straight out of the block or through clocked flip-flop.



- Each function is comprised of microcells each capable of implementing a combinational or registered function.
- The FB generates outputs that drive the fast connect switch matrix.
- Logic within FB is implemented using a sum-of-product form.

:) Interconnect matrix :

- The interconnect matrix connects signals to FB inputs, all IOB outputs and all 0. FB outputs drive the fast connect matrix. Any of these may be selected to drive each FB with a uniform delay.

) I/O block :

- The I/O Block interfaces between the internal logic and the device uses I/O Pins.
- Each I/O Block includes an input buffer, output driver, output enable selection multiplexer and user programmable ground control.
- The i/p buffer is compatible with 5 V TTL, 3.3 V CMOS and 2.5 V CMOS signals.
- Each output driver is designed to provide fast switching with minimal power noise. All output driver in the device may configured for driving either 3.3 V CMOS level.
- Each IOB provides user programmable ground pin capability. This allow device I/O pins to be configured as additional ground pins.
- Each IOB also provides for bus-hold circuitry (also called keeper) that is active during valid user operation.

9.2 Advantages of CPLD

- CPLD has many benefits as listed below :
  i) Ease of design
  ii) Faster time to market.

iii) Low development costs.

iv) Longer time in market through field upgrade ability.

v) Increased product revenue.

vi) Reduced PCB area.

vii) Decreased component inventory.

viii) Lower cost than discrete parts.

1.9.3 Applications of CPLD

1. Used in wide variety of applications, from implementing random glue logic prototyping small gate arrays.
2. CPLDs can realize reasonably complex designs, such as graphics controlle LAN controllers, UARTs, cache control and many others.
3. One of the most common uses in industry at this time, and a strong reason f large growth of the CPLD market, is the conversion of designs which consist multiple SPLDs into a smaller number of CPLDs.
4. Circuits which can exploit wide AND/OR gates, and do not need a very lar number of flip-flops are good candidates for implementation in CPLDs.
5. A significant advantage of CPLD : Simple design changes throu re-programming, within system programmable CPLDs possible to reconfigu hardware without power down.
6. Predictability of circuit implementation and speed performance is one of t strongest advantage of CPLD architectures.

1.9.4 Selection Criteria for CPLD and FPGA

- When considering CPLD for use in a design following issues should be taken i account.

1) Programming technology :

- EPROM, EEPROM or Flash EPROM. This will determine the equipment needed program the devices and whether they can be programmed only once or ma times.

2) The function block capability :

i) How many function blocks are there in the device ?

ii) How many product and sum terms can be used ?

iii) What are minimum and maximum delays through the logic ?

iv) What additional resources are there ?

v) What kind of register controls are available (clock, enable, reset, preset) ?

vi) How many are local inputs to the function block and how many are global chipwide inputs ?

vii) What kind of clock driver are in the device and what is the worst case skew of the clock signal on the chip ?

The I/O capability :

i) How many I/O are independent, used for any function and how many are dedicated for clock input ?

ii) What is the output drive capability interm of voltage level and current ?

iii) What kind of logic is included in an I/O block that can be used to increase the functionability of the design ?

9.5 Comparison of CPLD and FPGA

Sr. No.	CPLD	FPGA
1.	CPLD provide highest performance, but they also contain fewer registers than FPGAs.	FPGAs are offered in a wide density range, from few thousand gates to several million gates.
2.	CPLDs are ideal for complex blocks with a large number of inputs.	FPGA architecture is more granular compared to CPLD, making it suitable for designs with large number of simple blocks with few numbers of inputs.
3.	CPLDs can implement large functions in one pass through logic array.	FPGAs can't implement large functions in one pass through the logic array.
4.	CPLD's resources are partitioned into logic blocks, imposing restrictions on how they may be used.	FPGAs contain arrays of logic cells and are less partitioned than CPLDs.
5.	A number of input-output pins offered by CPLD are significantly higher.	A number of input-output pins offered on the FPGA are less than CPLD.
6.	Any design can fit into a 100 k gate CPLD, i.e. space required on silicon is less.	Same design require 150 k gate FPGA i.e. space required on silicon is more.
7.	CPLD based design needs less board space, board layout complexity, cost and bill of material (BOM) than FPGA.	FPGA based design needs increased board space, board layout complexity, cost and bill of material (BOM).

8.	CPLD doesn't require any memory store configuration program.	FPGA invariably require one or more configuration PROM, depending on the of FPGA used.
9.	CPLDs with their course-grain architecture and dedicated routing offer the advantage of predictable timing and end-of-use.	Due to granular architecture and segmen routing of FPGAs, it is difficult to predic the speed performance of a design until is targeted in the software.
10.	CPLD consume more power than FPGA devices.	FPGA consumes less power than CPLD.
11.	Process technologies are EPROM, EEPROM and FLASH.	Process technologies are SRAM, anti-fuse and EEPROM.
12.	Applications of CPLD bus interfaces, complex state machines, fast memory interfaces, wide detectors, PAL device integration.	FPGA applications Logic consolidation, board integration, replaces absolute devices, simple state machines, complex controllers/interfaces

University Questions

1. *Draw and explain general architecture of CPLD.* **GTU : Winter-14, Marks**
2. *Compare FPGA and CPLD.* **GTU : Summer-21, Winter-21, Marks**

1.10 Computer-Aided Design Technology

- Synthesis of logic circuits using AND, OR and NOT gates is very simple and approach can be used for design of smaller circuits. However, logic circuits fo in complex systems, such as today's computers, cannot be designed manu They are designed using sophisticated CAD tools that automatically implement synthesis techniques.
- To design a logic circuit, a number of CAD tools are needed. They are usu packaged together into a CAD system, which typically includes tools for following tasks : design entry, synthesis and optimization, simulation and phys design. A typical CAD system in its first stage is shown in the Fig. 1.10.1. (See Fig. 1.10.1 on next page)

Design entry

- Design entry is the process which involves entering into the CAD syster description of the circuit being designed. This stage is called design entry. Th are three design entry methods; using truth tables, using schematic capture, writing source code in a hardware description language. (See Fig. 1.10.2 on page 1 - 38)

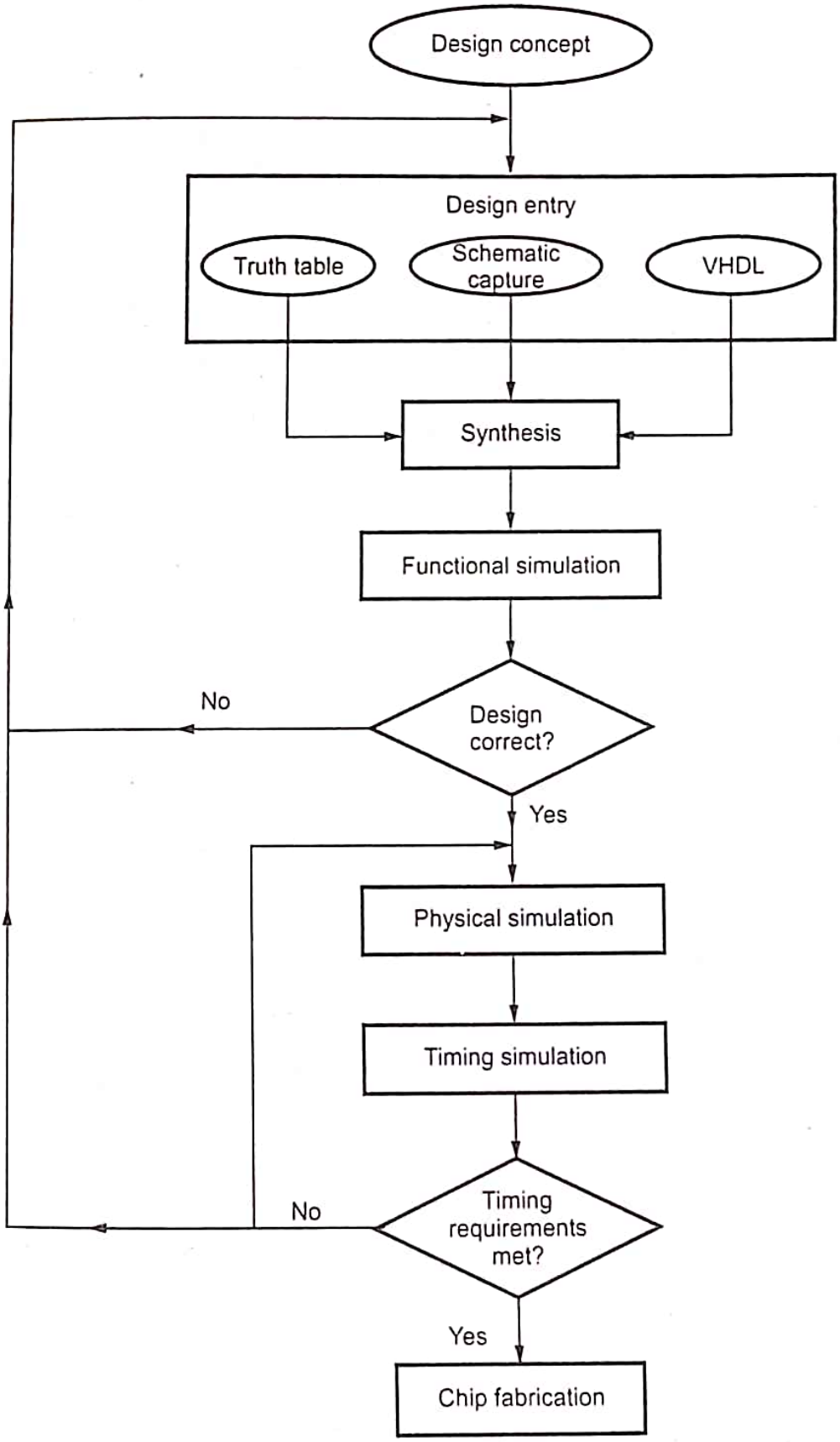


Fig. 1.10.1 A typical first-stage CAD system

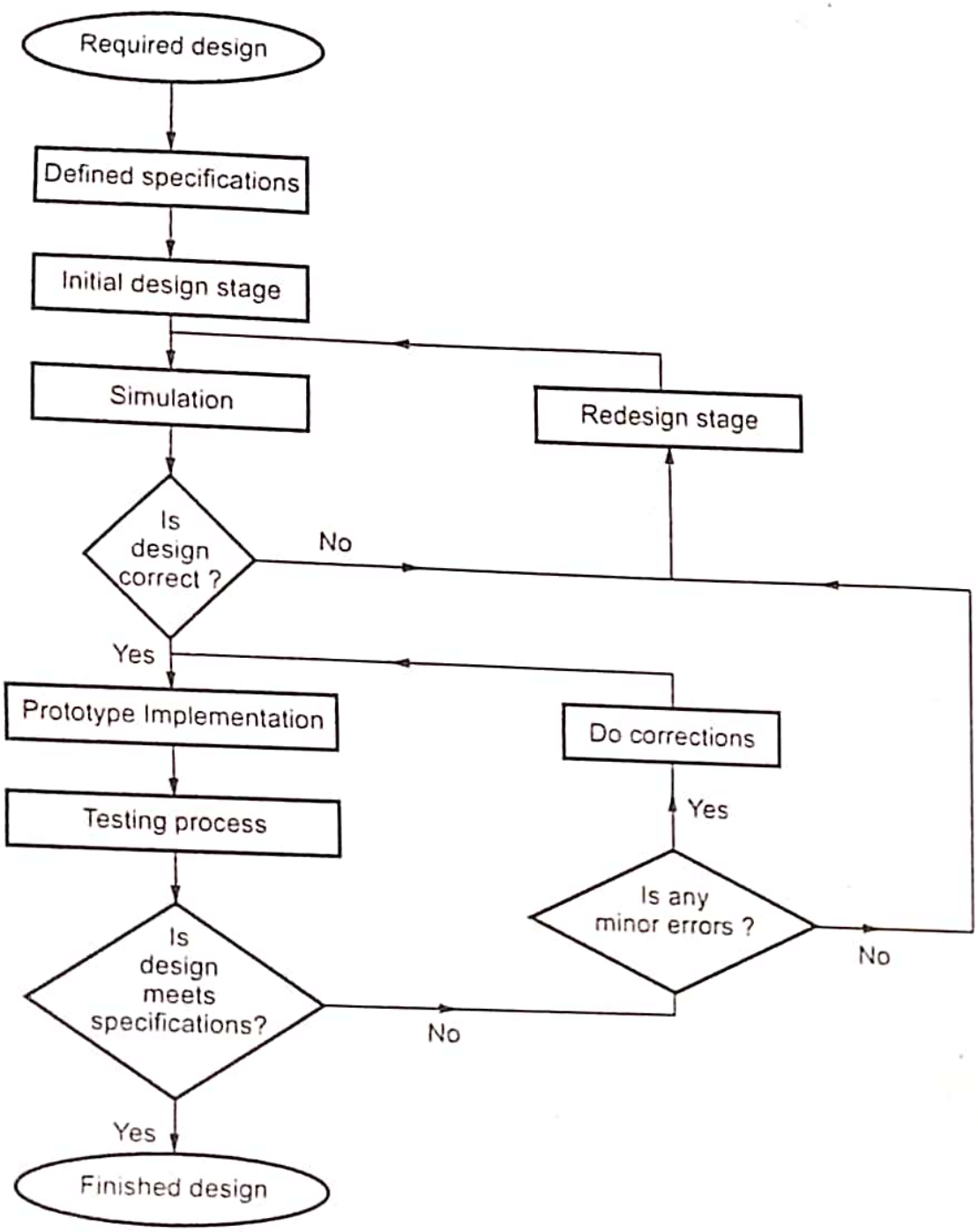


Fig. 1.10.2 A typical development process

i) Design entry with truth tables

- Many CAD systems allow design entry using truth tables, where the table specified as a plain text file. Alternatively, it may also be possible to specify truth table as a set of waveforms in a timing diagram. 'Waveform editor' is us to draw the timing diagram. CAD system is capable of transforming this timi diagram automatically into a network and logic gates. Because truth tables a practical only for functions with a small number of variables, this design ent method is not appropriate for large circuit.

esign entry with schematic capture

A schematic capture tool uses the graphics capabilities of a computer and a computer mouse to allow the user to draw a schematic diagram. To facilitate inclusion of basic gates in the schematic; the tool provides a collection of graphical symbols that represent gates of various types with different number of inputs. This collection of symbols is called library. The gates in the library can be imported into the user's schematic and the tool provides a graphical way of interconnecting the gates to create a logic network. The tool called 'Graphic Editor' of CAD system can be used to draw a schematic diagram.

Design entry with hardware description languages

A Hardware Description Language (HDL) is similar to a typical computer programming language except that it is used to describe hardware rather than a program to be executed on a computer. Two HDLs are IEEE standards : VHDL (Very High Speed integrated circuit hardware Description Language) and verilog HDL. Both languages are widely used in industry. Design entry of a logic circuit is done by writing VHDL or Verilog code.

thesis and optimization

- We know that synthesis is the process of generating a logic circuit from a truth table. Synthesis CAD tools perform this process automatically. Regardless of what type of design entry is used, the initial logic expressions produced by synthesis tools are not likely to be in an optimal form. Because these expressions reflect the designer's input to the CAD tools, it is difficult for a designer to manually produce optimal results, especially for large circuits. One of the important tasks of synthesis tools is to manipulate user's design to automatically produce equivalent circuit. This step of synthesis is called logic synthesis or logic optimization.

ulation

- Once the design entry and synthesis are complete, it is useful to verify that the designed circuit functions as per expectation. A CAD tool that performs this task is called a functional simulator. This simulator uses two types of information, first is user's initial design in terms of logic equations during synthesis and second, user specified valuations of the circuit's inputs that are applied to these equations during simulation. The output of the simulation is provided either in truth-table form or as a timing diagram.

10.1 CAD Tools for VLSI Design

i] **Front-end Tools** : Design Entry, Editors, Simulation, Synthesis, Timing Analysis, DFT Insertion, Test Generation.

ii] **Back-end Tools** : Floor Planning, Place-and-Route, Extraction, LVS (Layout vs. Schematic), LVL (Layout vs. Logic), ERC, DRC, Pattern Generators, Format Converters, Mask Graphics.

iii] **Design Capture Tools :** Editors, VHDL, SystemVerilog, SystemC, State Char FSM Capture.

iv] **Synthesis Tools :** Behavioral Synthesis, RTL Synthesis, FPGA Synthesis, Log Synthesis, Physical Synthesis, Module/Cell Generators (ROM, PLA, RAM Data-path Compiler, Adder/Multiplier Generators, DSP Synthesis.

v] **Analysis Tools :**

Checkers : DRC, ERC, Net Compare, Ratio Checker, Short-circuit Check Fan-in / Fan-out Checker, Power Checker.

Verifiers : Timing Verifier, Simulators, ICE/Hardware Simulators, Form Verifier.

vi] **Testing Related Tools :** ATPG, DFT Tools.

1.11 Short Questions and Answers

Q.1 Draw the VLSI design flow.

Ans. :

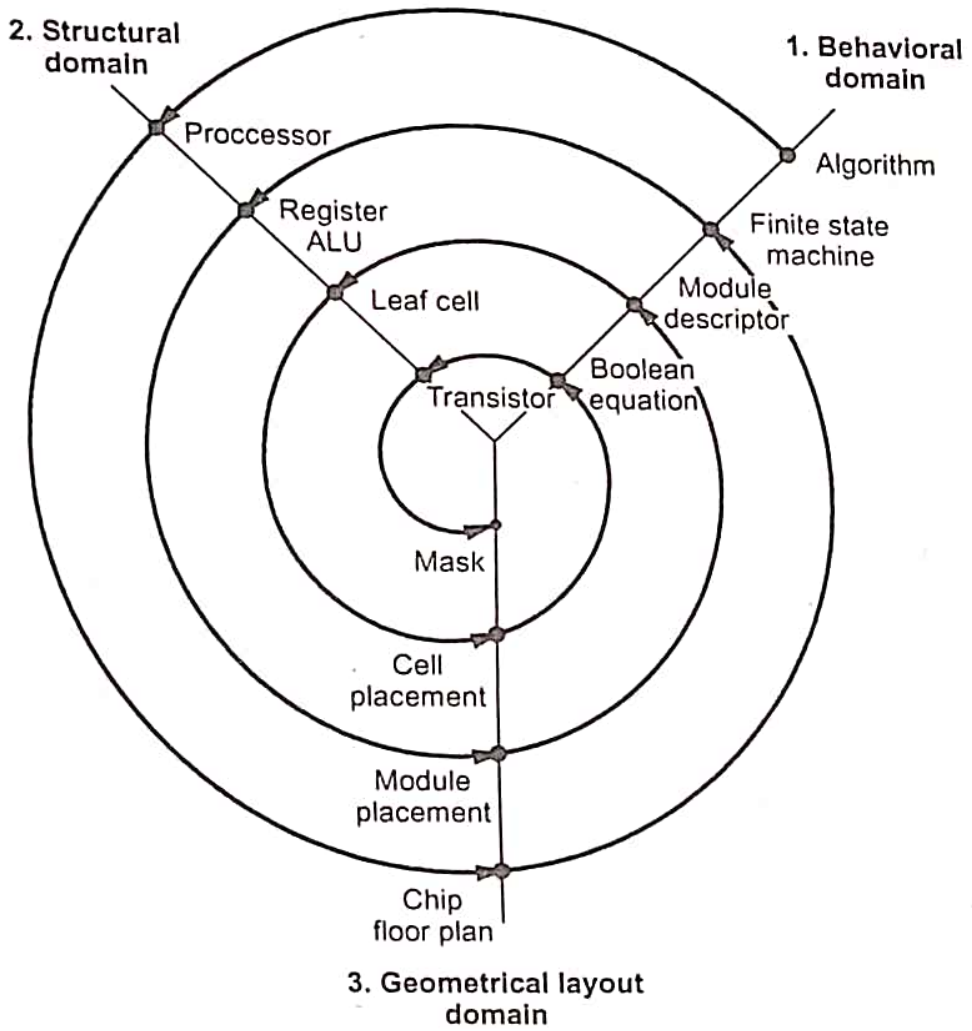


Fig. 1.11.1 VLSI design flow in three domain

Q.2 What is the fundamental goal in device modeling ?

Ans. : Goal of device modeling -

- The fundamental goal in device modeling is to obtain the functional relationship among the terminal electrical variables of the device that is to be modeled.
- These electrical characteristics depend upon a set of parameters including both geometric variables and variables dependent upon the device physics.

Q.3 Write the Comparison between FPGA and CPLD.

Ans. :

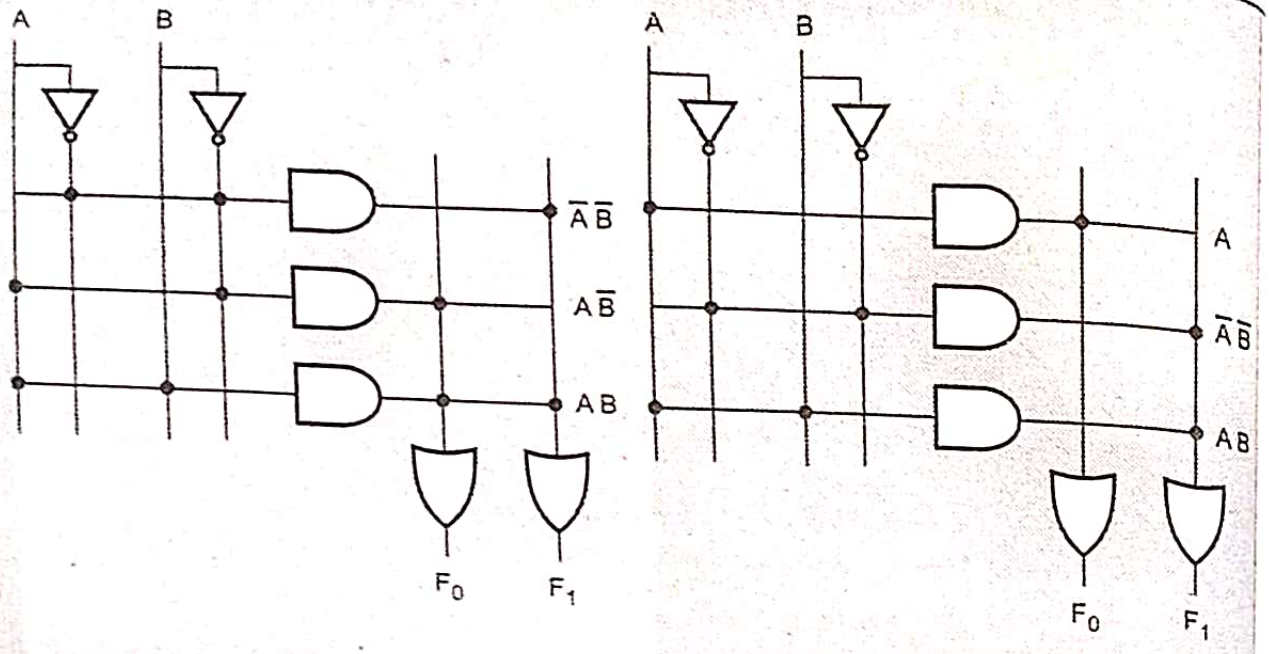
Sr. No.	CPLD	FPGA
1.	Complex Programmable Logic Device.	Field Programmable GateArray.
2.	It is collection PLDs.	It is collection of CLBs.
3.	Gate density upto 10000 gates.	1000-few hundred 1000s gates.
4.	Interconnection wise it is complex so called as CPLD.	Programmable at field or site so called as FPGA.
5.	PAL like blocks used as PLDs.	CLBs used as building blocks.
6.	AND-OR arrays are there in PAL like blocks.	LUTs are there in CLBs.
7.	Configuration context is stored in ROM.	Configuration context is stored in RAM.
8.	Configuration context is Non-volatile.	Configuration context is volatile.

Q.4 Explain difference between PLA and PAL.

Ans. :

Sr. No.	PLA	PAL
1.	Both AND and OR arrays are programmable.	OR array is fixed and AND array is programmable.
2.	PLA is a kind of programmable logic device used to implement combinational logic circuit. It has Programmable AND Gate linked with programmable OR Gate.	PAL is an combinational PLD that was developed to overcome certain disadvantage of PLA. PLA shows longer delay due to additional fusible links which results from using two programmable array and increase circuit complexity.

3.



Q.5 What is programmable logic array ?

Ans. : PLA

- A Programmable Logic Array (PLA) has a programmable AND array at the inputs and programmable OR array at the outputs. The PLA has a programmable AND array instead of hard-wired AND array.
- The number of AND gates in the programmable AND array are usually much less and the number of inputs of each of the OR gates equal to the number of AND gates.
- The OR gate generates an arbitrary Boolean function of minterms equal to the number of AND gates.

Q.6 What are the types of programmable devices ?

Ans. : Types of Programmable Logic Devices (PLDs) : PROM, PAL, PLA, CPLDs, FPGAs, etc.

1. PROM : The PROM (Programmable Read Only Memory) has a fixed AND array (constructed as a decoder) and programmable connections for the output OR gates array. The PROM implements Boolean functions in sum-of-minterms form.
2. PAL : The PAL (Programmable Array Logic) device has a programmable AND array and fixed connections for the OR array.
3. PLA : The PLA (Programmable Logic Array) has programmable connections for both AND and OR arrays. So it is the most flexible type of PLD.

4. Complex Programmable Logic Devices (CPLDs) : A CPLD contains a bunch of PLD blocks whose inputs and outputs are connected together by a global interconnection matrix.
5. Field Programmable Gate Arrays (FPGAs) : The FPGA consists of 3 main structures :

   A] Programmable logic structure, B] Programmable routing structure and

   C] Programmable Input/Output (I/O).

Q.7 State the three important blocks of FPGA architecture.

Ans. : The building blocks of FPGA are :

1. Logic Cells (LCs) grouped into Configurable Logic Blocks (CLBs).
2. I/O blocks.
3. Programmable interconnects.

❑❑❑

Notes

2 Fabrication of MOSFET

Syllabus

Introduction, Fabrication Process flow : Basic steps, C-MOS n-WellProcess, Layout Design rules, full custom mask layout design.

Contents

2.1 Introduction

2.2 Fabrication Processes **Winter-14, 21,**

..... **Summer-15, 21,** ..... Marks 7

2.3 CMOS Fabrication Process ..... **Summer-21, Winter-21,** ... Marks 7

2.4 Layout Design Rules

2.5 Full-Custom Mask Layout Design

2.6 Short Questions and Answers

2.1 Introduction

- CMOS fabrication technology involves both n-channel (nMOS) and p-chann (pMOS) transistor fabrication on same substrate.
- Special regions called tubs or wells are created on opposite substrate type. F example, p-well is created in n-type substrate and n-well is created in p-ty substrate.
- Simplified process for fabricating n-well CMOSIC with single polysilicon layer shown in Fig. 2.1.1.

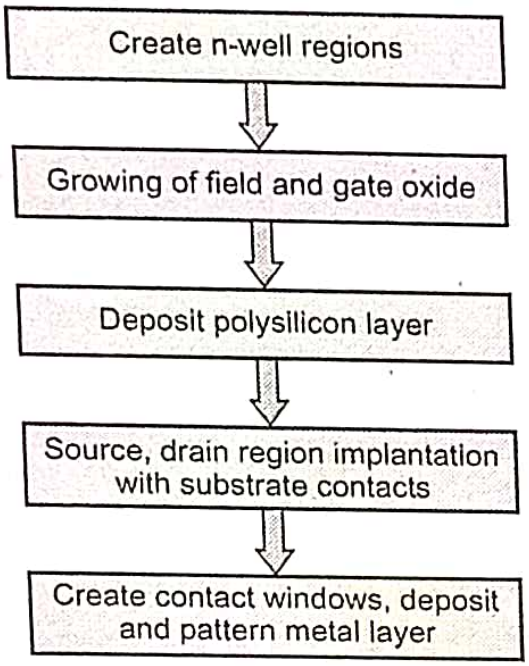


Fig. 2.1.1 Fabrication process of n-well CMOS IC

2.1.1 Difference between BJT and MOSFET

Sr. No.	BJT	MOSFET
1.	BJT is a Bipolar Junction Transistor.	MOSFET is a Metal Oxide Semiconduc Field-Effect Transistor.
2.	A BJT has three terminals called emitter, collector and base.	MOSFET has a gate, source and drain.
3.	BJTs are preferred for low current applications.	MOSFETs are for high power function

2.1.2 NMOS is Preferred over PMOS

- N-channel MOSFETs have some inherent performance advantages over p-cha MOSFET's. The mobility of electrons, which are carriers in the case of n-channel device, is about two times greater than that of holes, which are carriers in the p-channel device. Thus an n-channel device is faster tha p-channel device.

- As conductivity is proportional to mobility, NMOS offers less $R_{DS(on)}$ and gate capacitance than PMOS for same die area.
- As the mobility of the carriers in an NMOS is approximately 2 to 3 times higher than that of a PMOS, for the same $R_{DS(on)}$ value, the PMOS must be 2 to 3 times the size of the NMOS.
- In high frequency applications PMOS results into higher switching losses due to their higher gate charge requirement than NMOS for same current rating.

2.2 Fabrication Processes

GTU : Winter-14, 21, Summer-15, 21

- Various processes in fabrication are :

1. Masks :

- All processes needs use of masks during various stages of fabrication. These masks are mainly produced by optical techniques or photolithographic process. But due to geometric dimensions shrink and different chip designs on a single wafer, some other techniques are emerging. One such method of mask is by using E-beam machine.

Steps in E-beam mask process

1. A chrome-plated glass plate is coated with E-beam sensitive resist.
2. The mask description data is loaded with E-beam machine.
3. Plate is loaded into E-beam machine. Here, the plates exposed with the patterns specified by customer's mask data.
4. The plate is then introduced into developer to bring out the pattern left by E-beam in resist coating.
5. The plate is sent to bake and plasma de-summing in order to remove resist residue.
6. The chrome is etched and stripped off remaining E-beam resist.

Advantages of E-beam masks

1. Tighter layer to layer registration.
2. Smaller feature sizes.

2. Oxidation

- During the process of IC fabrication when silicon is exposed to oxygen or water vapour at high temperature, silicon rating with oxygen to form a continuous layer of silicon dioxide, this process is called **thermal** or **wet oxidation.** This oxide layer is used as mask during dopant diffusion, as a junction passivator, as an insulating oxide or as a gate dielectric in MOS transistor fabrication.

3. Lithography

- Lithography is a process of transferring patterns of geometric shapes in a mas a layer of radiation sensitive material (resist) for covering surface of semicondu wafer.
- Various lithographic process employed in integrating VLSI patterns on substrat shown in Fig. 2.2.1.

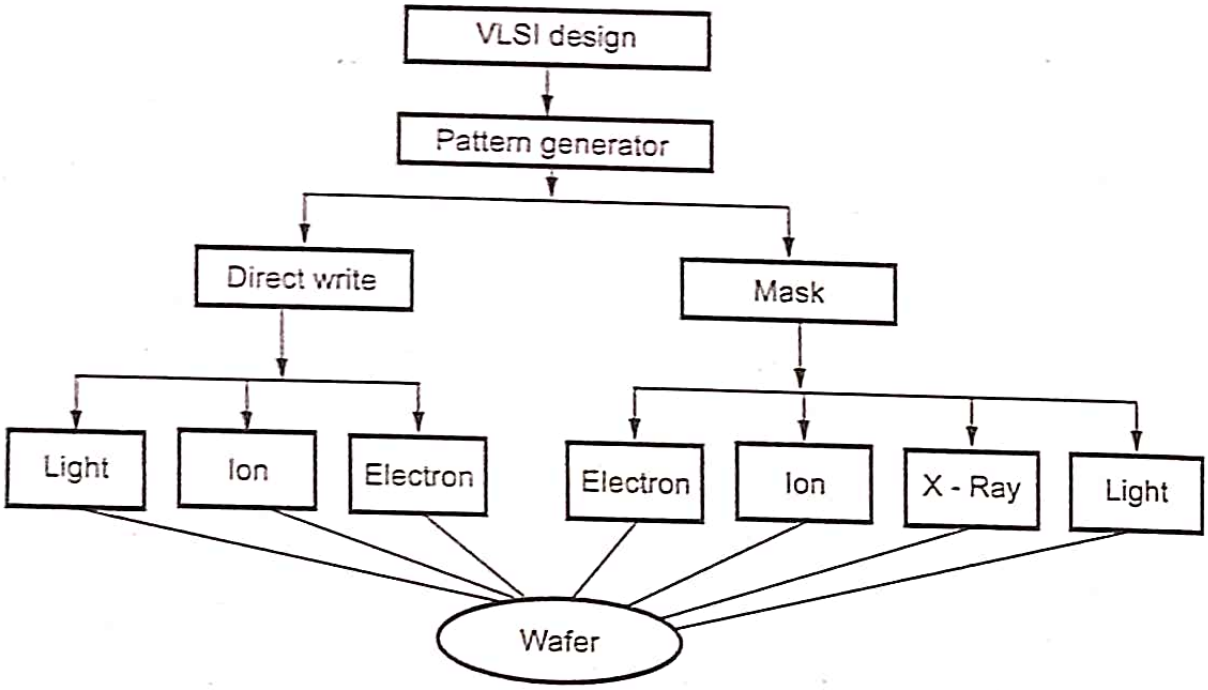


Fig. 2.2.1 Lithography process

- Lithography is process of impressing several flat embossed slabs onto a pa Every slab has definite colour. After this work, the circuit patterns are transfe on wafers.

Lithography Process by Means of Optical Replication Process

- The formation of patterns /images in a photoresist in optical lithography is du visible or ultraviolet radiations, which utilizes proximity or projection print Electron beam exposure and laser beam scanning is used for making mask.
- The exposing radiation is transmitted through the clear part of mask transferring patterns on substrate. The non transparent part blocks so radiations. The photoresists (sensitive to radiation) is coated onto substrate surf The resist image is developed on substrate.
- After the patterns are defined, an etching process is employed to selectiv remove the masked portions of the underlaying layer. Fig. 2.2.2 illustrates lithographic process in VLSI manufacturing.

Positive photo-resist

- The photoresist which is initially insoluble and becomes soluable on exposure UV light is called **positive photoresist.**

Negative photoresist

- The photoresist which is initially soluble and becomes insoluble (hardened) on exposure to UV-light is called **negative photoresist.**
- Negative photoresistive are more sensitive to light.

Functions of photoresists :

- Photoresist layers have two basic functions :
  a) precise pattern formation; and
  b) protection of the substrate from chemical attack during the etch process.

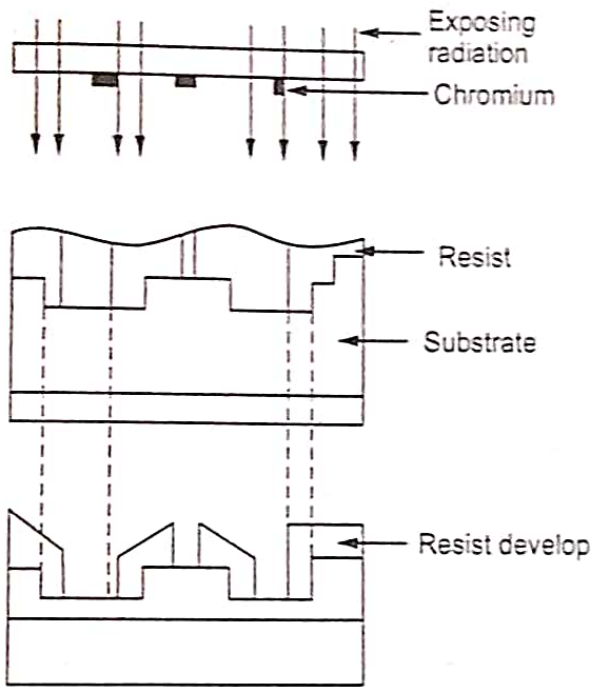


Fig. 2.2.2 Lithography by optical replication process

4. Diffusion

- The process of introducing controlled amounts of dopants into semiconductors is called **diffusion**. The conductivity of silicon is changed because of diffusion producing selectively n-type and p-type regions requires that diffusion to be carried out at higher temperature and by placing the dopant atoms on the surface of semiconductor.
- Diffusion is carried out in furnace where inert gas containing dopant is pass through the wafer at temperature range 800 to 1200 °C.
- Diffusion of boron into silicon produces p-type silicon and diffusion of arsenic or phosphorous into silicon produces n-type semiconductor.

5. Ion Implantation

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

6. Metallization

- Metallization is process of formation of metal films for interconnections, ohmic contacts and rectifying metal semiconductor contacts. The methods used for making metal films are Chemical Vapour Deposition (CVD) and Physical Vapour Deposition (PVD).

7. Passivation

- The *passivation* means to add a protective glass layer to the silicon wafer. T passivation is also called as overglass.
- Passivation is the final processing step. It prevents the ingress of contaminati To allow connection to I/O pads and test points, suitable openings are provid on the test passivation.
- The chip can be directly connected to a circuit board using plated solder bumps the pad openings. This process is called *bumping*.

8. Etching

- **Etching** is the process of material being removed from a material's surface. T two major types of etching are wet etching and dry etching (e.g., plasma etching
- The etching process that involves using liquid chemicals or etchants to take off substrate material is called **wet etching**.
- In the plasma etching process, also known as dry etching, plasmas or etch gases are used to remove the substrate material. Dry etching produces gase products, and these products should diffuse into the bulk gas and be expel through the vacuum system. There are three types of dry etching (e.g., plas etching): chemical reactions (by using reactive plasma or gases), physical remo (generally by momentum transfer), and a combination of chemical reactions a physical removal. On the other hand, wet etching is only a chemical process.

Comparison of diffusion and ion implantation method of impurity dopings

Sr. No.	Diffusion	Ion implantation
1.	Unable to achieve high dopant concentrations.	Wide range of dopant dose (10^{11} ~ 10^{17} atoms / cm^2).
2.	High temperature, hard mask.	Low temperature, photoresist mask.
3.	Isotropic dopant profile.	Anisotropic dopant profile.
4.	Cannot indepently control of the dopant concentration and junction depth.	Can independently control of the dopa concentration and junction depth.
5.	Batch process.	Both batch and single wafer process.

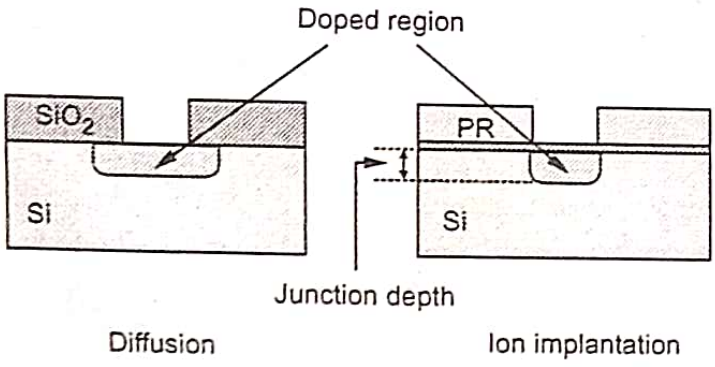


Fig. 2.2.3

2.2.1 Advantages of Ion Implantation over Diffusion

- Main advantages of ion implantation compared to diffusion for the doping of semiconductors are :
  1. Short process times, good homogeneity and reproducibility of the profiles.
  2. Exact control of the amount of implanted ions by measuring the current. This is of particular importance for low concentrations, e.g., to adjust the threshold voltage of MOS transistors.
  3. Relatively low temperatures during the process.
  4. Various materials can be used for masking, e.g., oxide, nitride, metals, and resist.
  5. Implantation through thin layers, e.g., SiO_2 is possible.
  6. Low penetration depth of the implanted ions. This allows modification of thin areas near the surface with high concentration gradients.
  7. Sequences of implantation steps with different energies and doses allow optimization of the dopant profiles.

2.2.2 Types of Photo Resist Material

- Photoresists are used in microelectronics and microsystems technologies for the fabrication of µm- and sub-µm structures. Photoresist are photosensitive material temporarily coated on wafer surface.
- Two types of photoresist materials are : Positive photoresist and negative photoresist.
- Positive photoresists produced by all resist like e.g. AR-P 1200 (spray resists), AR-P 3100, 3200, 3500, 3700 are composed of a combination of film-forming agents such as e.g. cresol novolac resins and light-sensitive components such as e.g. naphthoquinone diazide (NCD), which are dissolved in solvents such as e.g. methoxy propyl acetate (equivalent to PGMEA).
- The addition of a light-sensitive component to the alkali-soluble novolac leads to a reduced alkali-solubility. The NCD-groups block the OH-groups of the cresol novolac resigns; the alkali-solubility decreases (inhibitory effect). After exposure at 308 - 450 nm (UV range) using an exposure mask, the light-sensitive component is converted into the respective indene carbonic acid derivative, the blockage is released and the alkali-solubility of positive resists is thus enhanced by a factor of 100.
- After development, only those areas which were protected by the mask remain, while exposed areas are detached. The refractive index of novolac-based resists is in a range between 1.58 - 1.63. After development, only areas protected by the

mask remain while irradiated areas are dissolved. Photoresists provide an excelle protection against liquid etch media with pH values between 0 and 12.

- "Negative photoresists like AR-N 4200, 4300, 4400 are composed of novolacs a bisazides (4200, no CAR) or novolacs, acid generators and amine compone (4300, 4400, CAR) dissolved in safer solvents like e.g. methoxypropyl acet (PGMEA). (CAR = Chemically Amplified Resist).
- "The chemical amplification is based on the formation of acids during irradiati and a subsequent cross-linking of amine components with the novolacs. Si acids (protons) are continuously formed during the cross-linking reaction, ea proton can induce many cross-linking events which results in a high sensitivity.

2.2.3 Fabrication of nMOS Transistor

- nMOS fabrication process involves nine steps as follows :

Step - 1 : Processing is done on a thin film (wafer) of pure silicon crystal. p-impurities are introduced as crystal is grown. Typically wafer diameter is 75 to 150 mm and thickness is 0.4 mm. Impurity is boron having concentrations of $10^{15}/cm^3$ to $10^{16}/cm^3$ giving resistivity of 25 Ω - cm 2 Ω - cm.

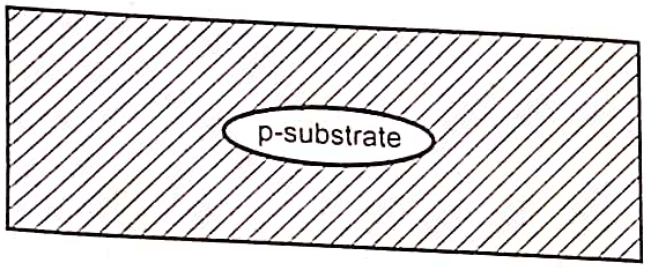


Fig. 2.2.4

Step - 2 : A layer of silicon dioxide (SiO_2) is grown over the surface of wafer for protection of surface. The thickness of silicon dioxide layer is typically 1 μm. This layer acts as barrier to dopants during processing.

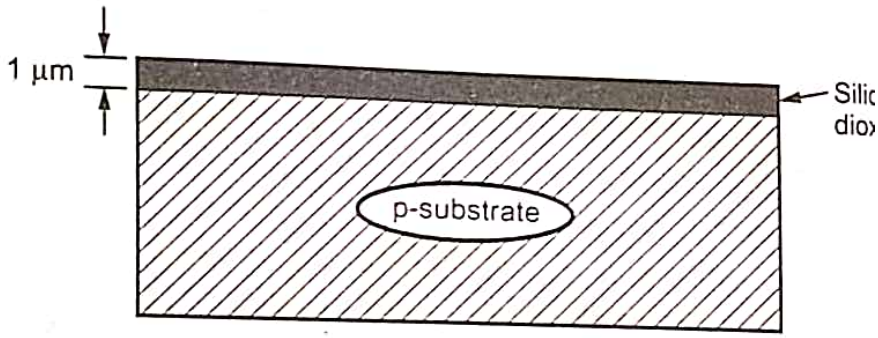


Fig. 2.2.5

Step - 3 : The SiO_2 layer is now covered with photoresist which is deposited on wafer.

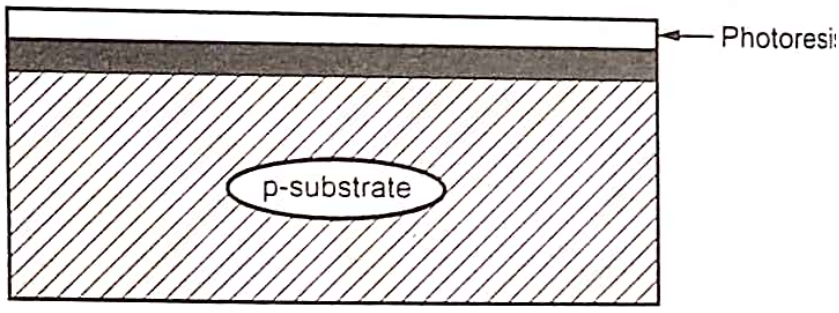


Fig. 2.2.6

Step - 4 :

The photoresist layer is exposed to UV light through a mask. The mask corresponds to regions into which diffusion is to take place with transistor channels.

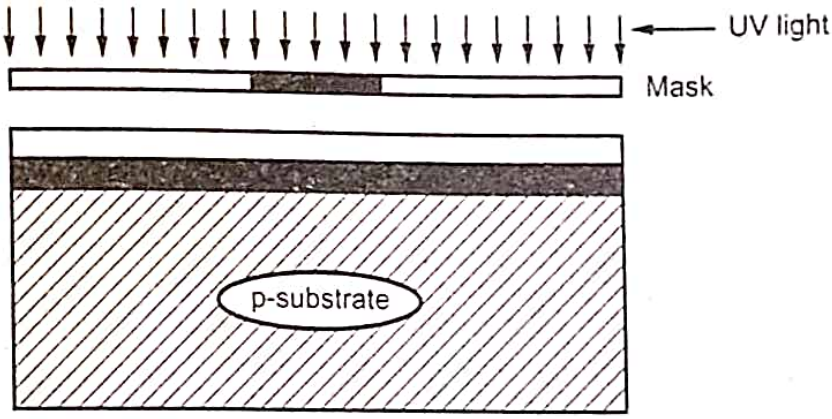


Fig. 2.2.7

Step - 5 : These regions are then etched together so that the wafer surface is exposed in window defined by mask.

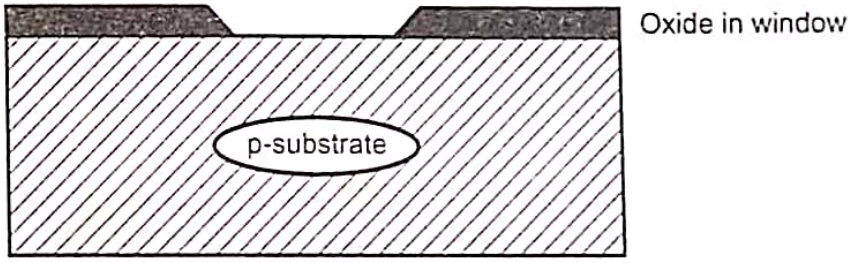


Fig. 2.2.8

Step - 6 : The photoresist layer remains on substrate is removed and a thin layer of SiO_2 (0.1 μm typical) is grown over the chip surface and then polysilicon is deposited on it to form gate structure. The polysilicon layer is fabricated by heavily doped polysilicon deposited by chemical vapour deposition (CVD). Precise control of thickness, impurity concentration and resistivity is important during fabrication process.

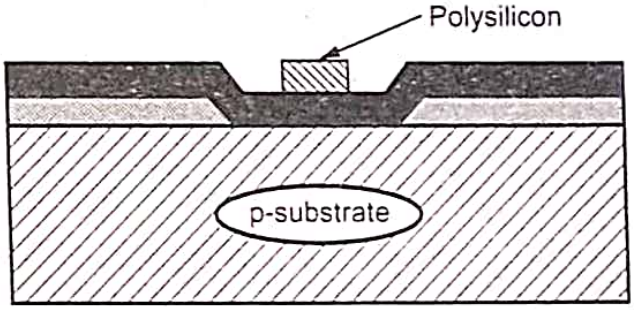


Fig. 2.2.9

Step - 7 : The thin oxide is removed to expose areas into which n-type impurities are to be diffused form the source and drain. Diffusion is achieved by heating the wafer to high temperature and passing gas containing the desired n-type impurity (Phosphorus) over the surface.

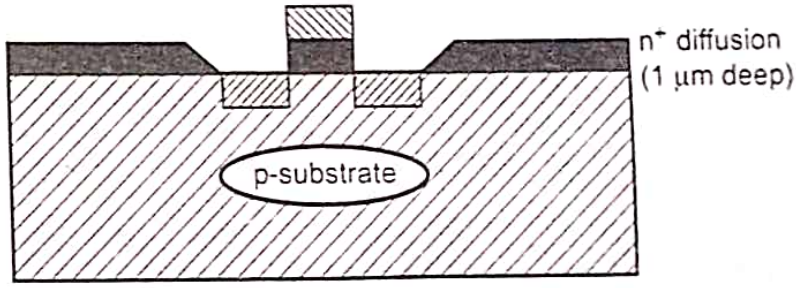


Fig. 2.2.10

Step - 8 : Thick silicon dioxide (SiO_2) is grown and then masked with photoresist and then etched to expose specific areas of polysilicon gate, drain and source areas where connections are made (contact holes).

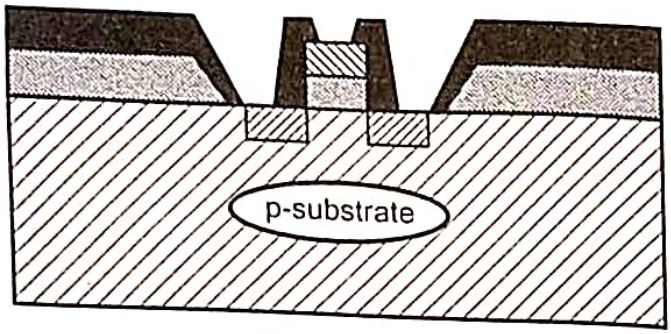


Fig. 2.2.11

Step - 9 : The wafer is then metal deposited over its surface to a thickness of typically 1 μm. The metal is usually aluminium. The metal layer is masked and etched to form the required interconnection pattern.

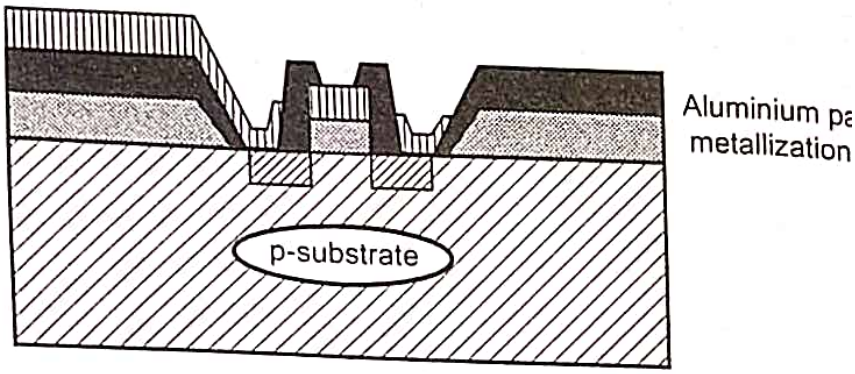


Fig. 2.2.12 nMOS fabrication process

- The fabrication process involves formation or deposition and patterning layers isolated by silicon dioxide. The layers are diffused in substrate, pol on oxide (on substrate) and metal insulated by oxide.
- In depletion mode devices it is important to introduce a mask implementation step between step 5 and step 6. The thick oxide is a masl self-aligning.

2.2.4 Fabrication of pMOS Transistor

- The pMOS fabrication process uses p-channel FETs. The basic active c p-channel.
- The electrical characteristics of p-channel MOSFETs are not easily cor compared to n-channel MOSFETs. It is bacause the mobility of p-type m considerably poorer than that of n-type material.
- The pMOS technology is rarely used today.

2.2.5 Local Oxide Silicon (LOCOS)

- Local oxidation of silicon (LOCOS) technique is based on the prir selectively growing the field oxide in certain regions, instead of selectivel away the active areas after oxide growth.
- Selective oxide growth is done by shielding the active areas with silico (Si_3N_4) during oxidation, which effectively inhibits oxide growth.

- The basic steps of the LOCOS process are illustrated in Fig. 2.2.13.

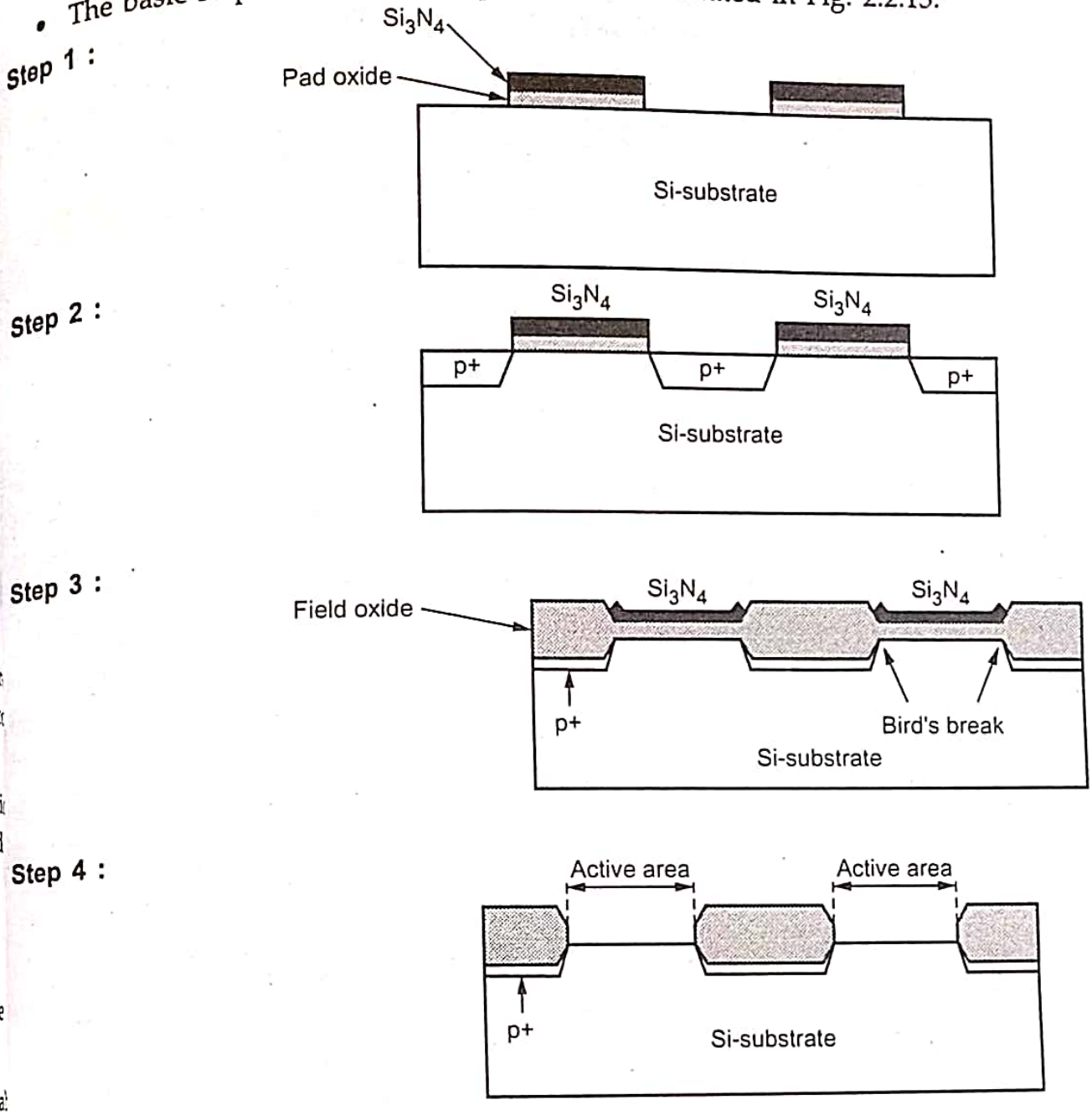


Fig. 2.2.13 LOCOS process

University Questions

1. *What are the problem associated with LOCOS technique with suitable diagrams ? Explain each of the fabrication steps involved in LOCOS technique with suitable diagrams.* **GTU : Winter-14, Marks 7**
2. *Define positive photoresist and negative photoresist.* **GTU : Summer-15, Mark 1**
3. *Describe the process of fabricatin of the NMOS transistor.* **GTU : Summer-15, Marks 7**
4. *Explain LOCOS techniques used for device isolation and state its advantages over the etched field oxidation technique.* **GTU : Summer-21, Marks 4**

5. *Compare ion implantation and diffusion methods used for impurity doping.* **GTU : Summer-21, Marks 3**
6. *Explain LOCOS method in detail.* **GTU : Winter-21, Marks 3**
7. *Describe process of photo lithography with example.* **GTU : Summer-21, Marks 3**

2.3 CMOS Fabrication Process

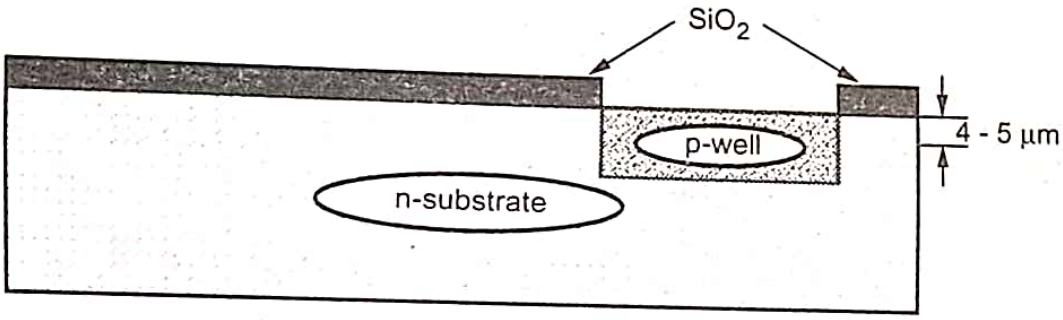
GTU : Summer-21, Winter-21

- There exists number of approaches used for fabrication of CMOS.
  1. p-well process
  2. n-well process
  3. Twin-tub process
  4. Silicon-on-insulator process (SOI).

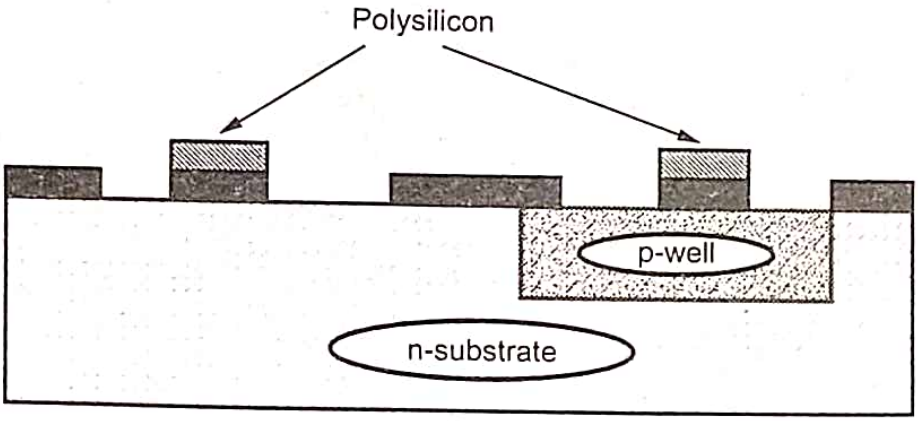
2.3.1 p-Well Process

- The structure consist of an n-type substrate in which p-devices are formed b masking and diffusion. To form n-type devices, a deep well is diffused into n-ty substrate.
- The process of diffusion is to be carried with special care because p-well dopi concentration and depth affects the threshold and breakdown voltages n-devices. For low threshold voltage (< 1.0 V), a deep well diffusion or hi resistivity is required. Deep wells require larger spacing between n-type a p-type transistors and wires due to lateral diffusion and hence larger chip area.

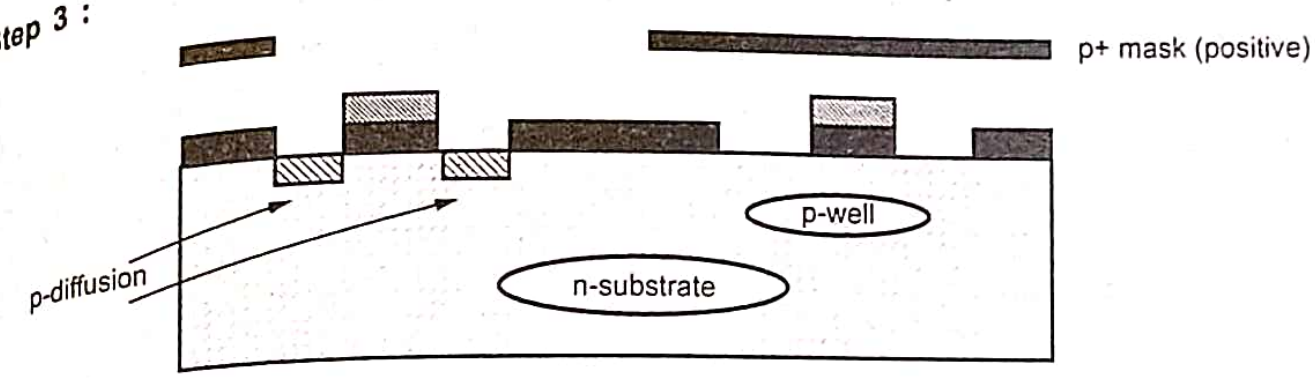
Step 1 :



Step 2 :



Step 3 :



Step 4 :

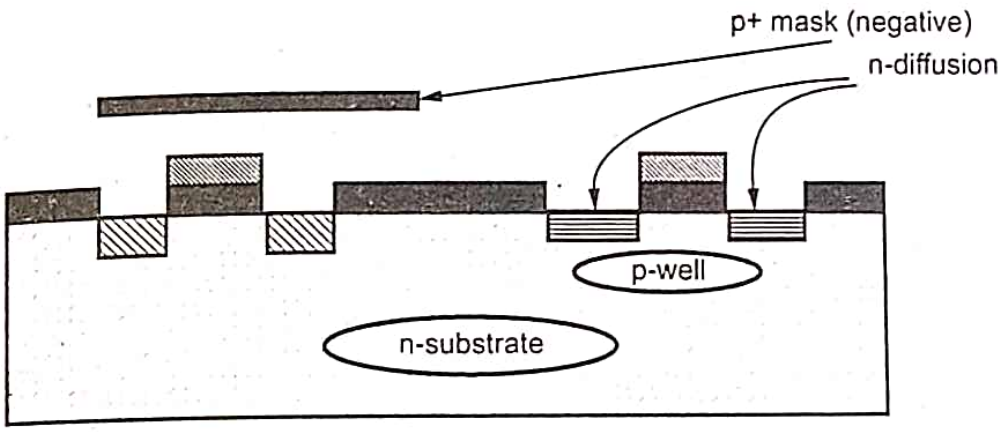


Fig. 2.3.1 p-well process steps

- The p-wells acts as substrates for the n-devices within the n-substrate provided voltage polarity restrictions are observed. These two areas are electrically isolated. Two connections V_{DD} and V_{SS} are required as there are two substrates effectively.
- Fig. 2.3.2 shows CMOS p-well inverter showing V_{DD} and V_{SS} substrate connections.

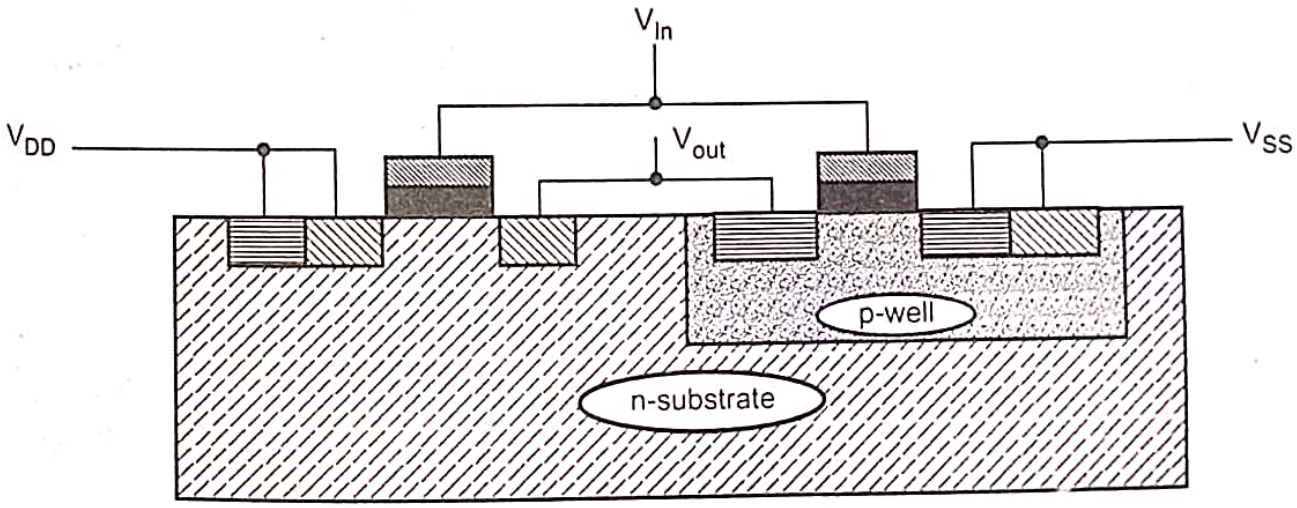


Fig. 2.3.2

- In p-well process masking, patterning and diffusion are similar process as nMOS fabrication. Typical processing steps can be summarized as :

Mask - 1	Defines the areas for deep p-well diffusions.
Mask - 2	Defines thinox regions where thick oxide is to be stripped and thin oxide grown to accommodate p and n transistors and diffusion wires.
Mask - 3	Pattern the polysilicon layer which is deposited after thin oxide.
Mask - 4	p + mask is used to define areas where p-diffusion is to take place.
Mask - 5	A negative form of p + mask with mask - 2, defines those areas where n- type diffusion is to take place.
Mask - 6	Contacts are defined.
Mask - 7	Metal layer pattern is defined by this mask.
Mask - 8	A passivation layer is applied and it defines the openings for access to bonding pads.

2.3.2 n-Well Process

- n-Well CMOS devices are superior to p-well of the lower substrate bias effects o transistor threshold also they have lower parasitic capacitances associated wi source and drain regions.
- Major steps in a typical n-well fabrication process are shown in Fig. 2.3.3. (Refer Fig. 2.3.3 on next page)
- Mask-1 defines the n-well regions, followed by a low dose phosphorous impla driven in by a high temperature diffusion step to form the n-wells. The well dep is kept such that to ensure against p-substrate to p+diffusion breakdown witho compromising the well to n+ mask separation.
- The next step includes defining the devices, diffusion paths, grow field oxid pattern the polysilicon, make contact cuts and metallization.
- An n+ mask and its complement may be used to define n and p diffusion regio respectively. These masks also include V_{DD} and V_{SS} contacts. Alternatively, p mask and its complement can be used.
- Fig. 2.3.4 shows inverter circuit fabricated by the n-well process. (Refer Fig. 2.3.4 on next page)
- n-channel devices are mainly used to form logic elements of high speed an density of elements.
- A low resistivity epitaxial p-type substrate can reduce latch up problem. Also will act as very low resistance ground plane to collect substrate currents.

- In n-well process, the performance of poorly performing p-transistor is degraded. Modern process may improve the device performance by using p-well and n-well fabrication.

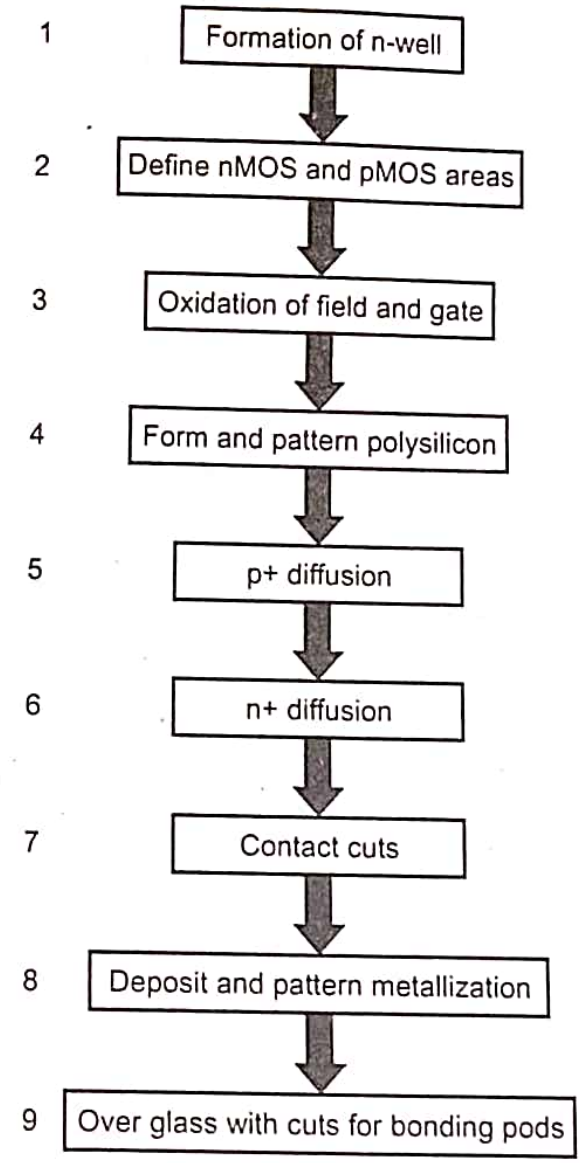


Fig. 2.3.3 Main steps in typical n-well process

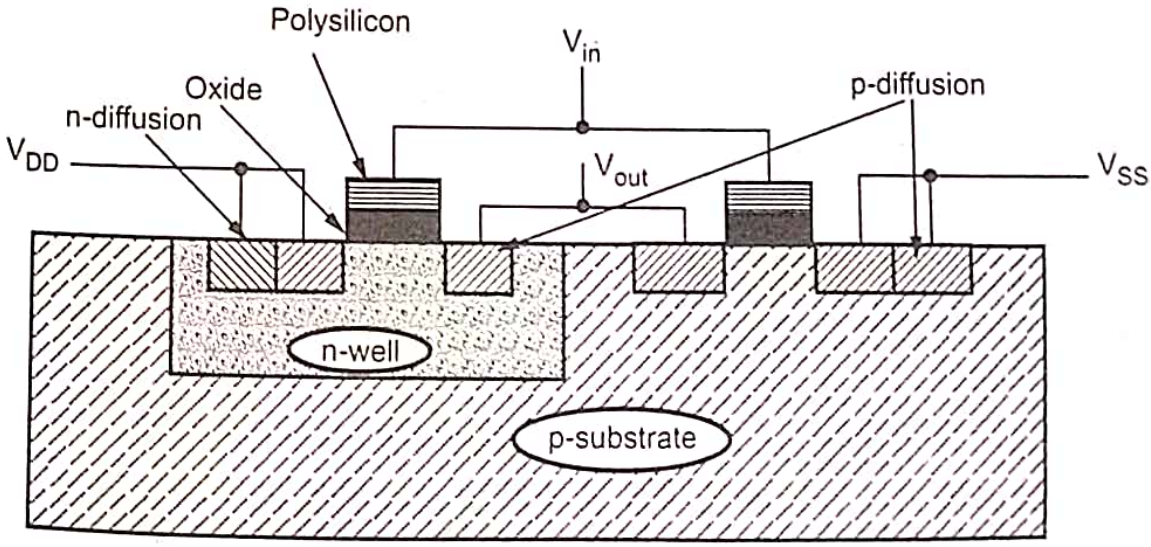


Fig. 2.3.4 n-Well CMOS inverter

2.3.3 Twin Well Process

- A twin-tub process is a logical extension of p-well and n-well approaches.
- The process starts with a substrate of high resistivity n-type material and th both p-well and n-well regions are created.
- With this process it is possible to preserve the performance of n-transisto without compromising p-transistors.
- Doping can be controlled easily and latch-up is also achieved as the twin-t process allows separate optimization of n and p-transistors. An invert arrangement with twin-tub is shown in Fig. 2.3.5.

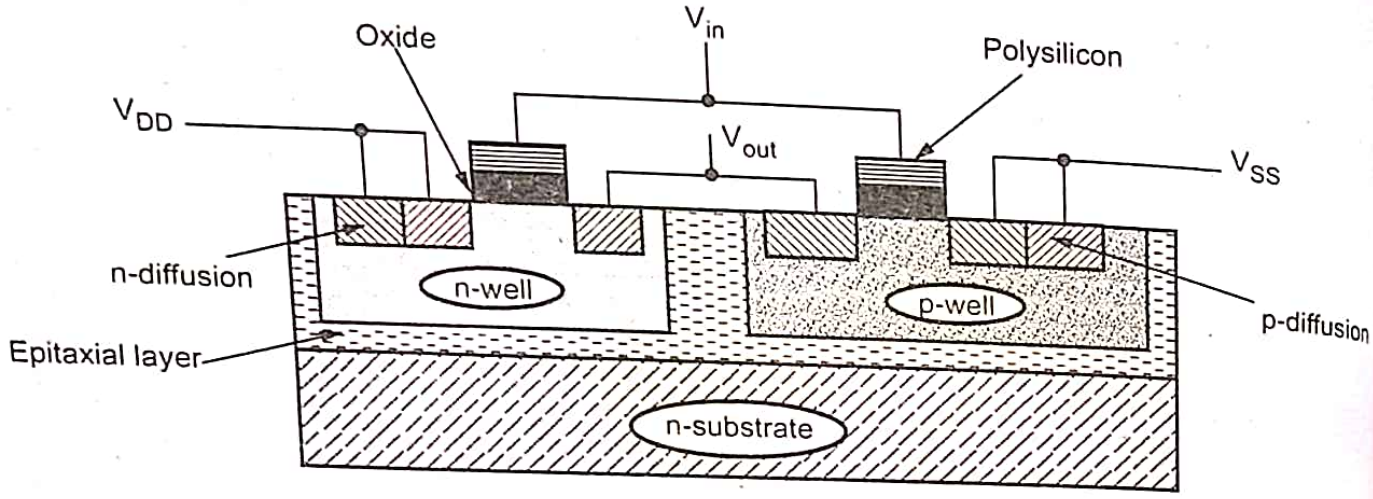


Fig. 2.3.5 Twin-tub inverter

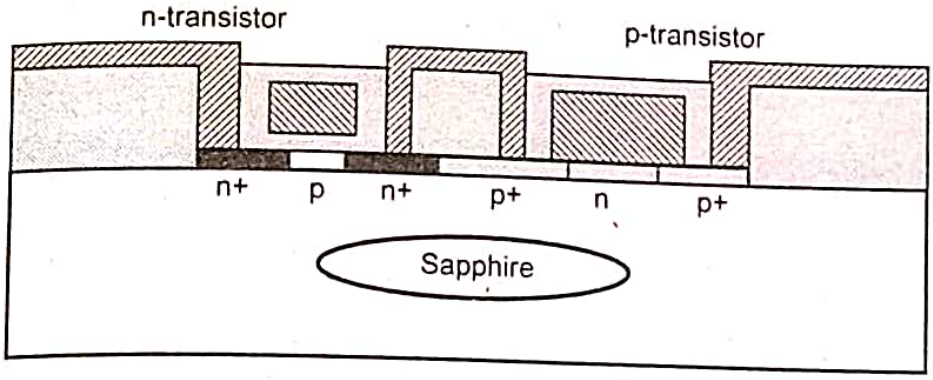
- The twin-tub CMOS technology provides the basic for separate optimization p-type and n-type transistor making it possible for threshold voltage, body effe and gain associated with n and p-devices to be independently optimized.

2.3.4 Silicon On Insulator (SOI)

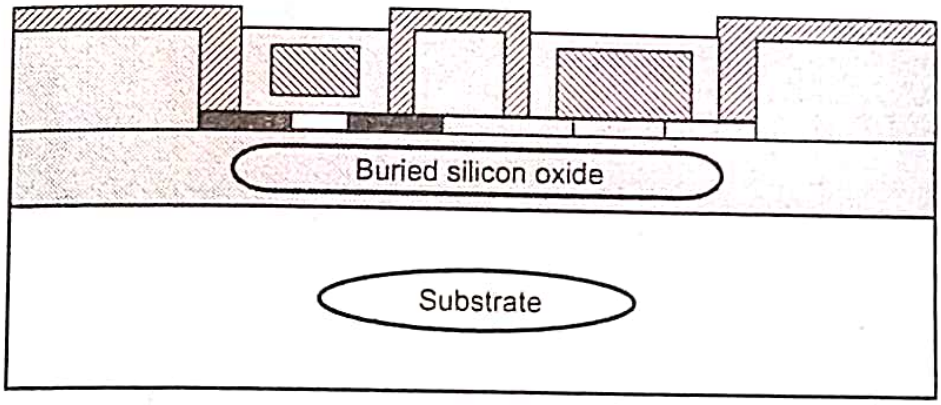
- A prominent emerging CMOS process is Silicon-On-Insulator (SOI). As the nam suggest it is a process where the transistors are fabricated on an insulator. Tw main insulators used are SiO_2 and sapphire.
- Two common types of SOI : 1. Sapphire substrate and 2. Silicon based SOI proce are shown in Fig. 2.3.6. (Refer Fig. 2.3.6 on next page)

1. Sapphire substrate

- In sapphire substrate technology, a thin layer of silicon is formed on sapphi surface. The thin layer of silicon is formed on sapphire surface. A thin layer silicon is selectively doped to define different threshold transistors.
- The gate oxide is grown at the top of thin silicon substrate and then polysilic gates are defined then nMOS and pMOS transistors are formed by implantation.



(a) Sapphire substrate



(b) Buried silicon oxide (Box)

Fig. 2.3.6 SOI types

2. Silicon based SOI process

- In silicon based SOI process, a silicon substrate is used and buried oxide (BOX) is grown on top of the silicon substrate. A thin silicon layer is then grown on the top of buried oxide and is then selectively implanted to form nMOS and pMOS transistors. The gate, source and drain regions are defined similar to bulk process.
- Sapphire is optically and RF transparent. Hence it can be used in optoelectric areas along with (group III-V based) light emitters.

SOI Advantages

- SOI process has following advantages :
  1. Lower diffusion capacitance.
  2. Smaller parasitic delay.
  3. Lower dynamic power consumption.
  4. The threshold variations are smaller. Lower nominal V_t results in faster transistors.

5. SOI is immune to latch up because the insulating oxide eliminates the parasitic bipolar devices.

2.3.5 Thermal Aspects of Processing

- The processing of fabrication of nMOS and CMOS devices involves different high temperature sequences as shown in Fig. 2.3.7.

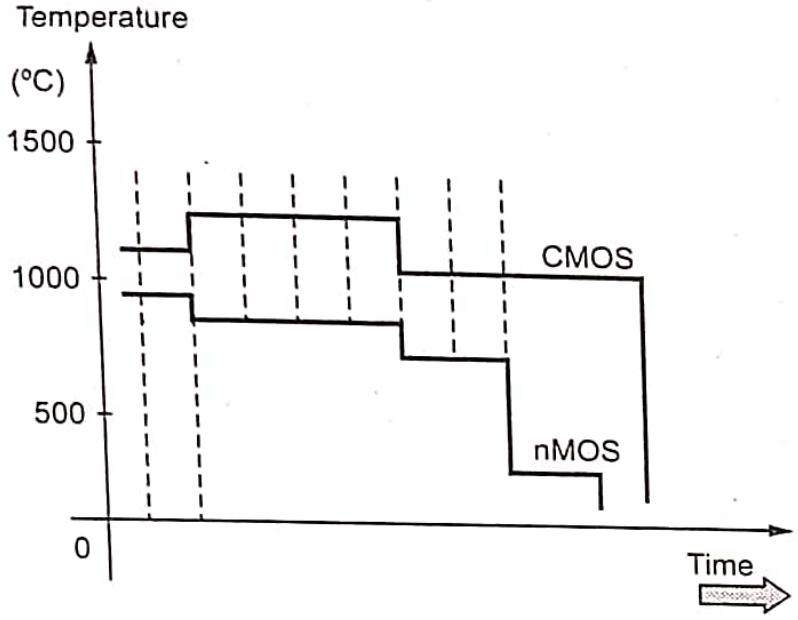


Fig. 2.3.7 Thermal sequence for nMOS and CMOS processes

- CMOS process requires temperature of 1100 to 1250 °C for diffusion process whereas nMOS process have no such requirement. Most nMOS processes require less temperature than CMOS process.
- Also the simplicity, ease of operation, high density per unit area of nMOS circuits it is more popular.

University Questions

1. *Explain the fabrication of CMOS n-well process using neat diagrams.* **GTU : Winter-21, Marks 7**
2. *Describe device isolation techniques with example.* **GTU : Summer-21, Marks 3**
3. *Explain with neat sketch diagram nMOSFET fabrication flow.* **GTU : Winter-21, Marks 7**
4. *Sketch PMOS fabrication steps.* **GTU : Summer-21, Marks 7**

2.4 Layout Design Rules

- Tlhe VLSI circuit/system designer and the fabricator may be experts in their own ways, but a smoother interface comes out very valuable because each has vested interests in making perticular task as easy as possible. Thus design rules attempt to provide a workable and reliable compromise that is friendly to both.

- On the one hand the designer wants in general smaller, tighter layouts for improved performance and decreased silicon area, whereas, the process engineer wants design rules to define a controllable and reproducible process. Thus design rules represent the best possible compromise between performance and yield.
- The design rules basically address two issues :

i) The interactions between different layers and

ii) The geometrical reproduction of features that can be reproduced by the mask-making and lithographical process.

- An important aspect regarding design rules is the achievable definition of the process line. The extent of definition is decided by process design and process line equipment. For example the use of a 10 : 1 wafer stepper instead of a 1 : 1 projection mask aligner would give better level-to-level registration. Similarly, the more mature the process line, more is the confidence of the process line capability and design rules facilitate tighter designs with fewer constraints on the designer.

Types of Design Rules :

- There are number of approaches to describe the design rules. These include 'micron' rules stated at some micron resolution and the lambda (λ)-based rules. Micron design rules are usually given as a list of minimum feature sizes and spacings for all the marks required in a given process. For example, the minimum active width might be specified as 1 μm. This is the normal style for industry.

2.5 Full-Custom Mask Layout Design

- Layout rules are also called as design rules or ground rules. The prime objective of layout rules is to build reliably functional circuit in smallest area. The design rules are also termed as compromise between performance and yield.

2.5.1 Design Rule Background

2.5.1.1 Well Rules

- The n-well is deeper implant than source/drain implants. Hence it is necessary to provide sufficient clearance between n-well edges and adjacent n^+ diffusions.
- The n-well sheet resistance can be several kW per square, therefore it is necessary to thoroughly ground the well. Also, it prevents excessive voltage drops due to well currents.

Mask Summary

- The masks are used to specify, the location of wells (n-well, p-well and deep n-well). Usually only one well is specified in a twin well process i.e. n-well, remaining area is for p-well.

2.5.1.2 Transistor Rules

- Four physical masks are used for defining CMOS transistors, These are -
  1. Active (diffusion , diff or thinox)
  2. n-select (n-implant, nim p or n plus)
  3. p-select (p-implant, pim p or p plus)
  4. Polysilicon (poly or polyg)
- The active mask defines areas where either n-type or p-type diffusion is to placed. The gates of transistors are defined by the logical AND of polysilic mask and active mask (where polysilicon crosses diffusion)
- n-select surrounds active regions where n-type diffusion is required and p-sel surrounds active regions where p-type diffusion is required.
- The poly must extend beyond the edges of active area otherwise the source a drain will be shorted by a diffusion. This extension gate is called gate extension.

Mask Summary

- The four basic masks may be called differently in some processes. Designers m use n-diffusion (ndiff) and p-diffusion (pdiff) to identify differently.

2.5.1.3 Contact Rules

- Electrical connection can be established between layers by special structures call contact cuts. There are three possible approaches for establishing contacts betwe polysilicon and diffusion in nMOS circuits. They are,
  i) Poly to metal and then metal to diffusion.
  ii) A buried contact-poly to diffusion
  iii) A butting contact-poly to diffusion using metal.
- The buried contact is a distinguishing feature of the nMOS process and is the m widely used, giving economy in space and a reliable contact. Butting contacts w in use earlier but now have been mostly superscaled by buried contacts.
- In CMOS designs, poly to diffusion contacts are almost invariably made throu metal. The process of making connections between metal and either of the oth two layers by buried contact is quite simple. Refer Fig. 2.5.1. The $2\lambda \times 2\lambda$ cont cut indicates an area from which oxide is to be removed right down upto t underlying polysilicon or diffusion surface. When deposition of metal layer tak place, the metal gets deposited through the contact cut area onto the underlyi area so that contact is established between the layers.

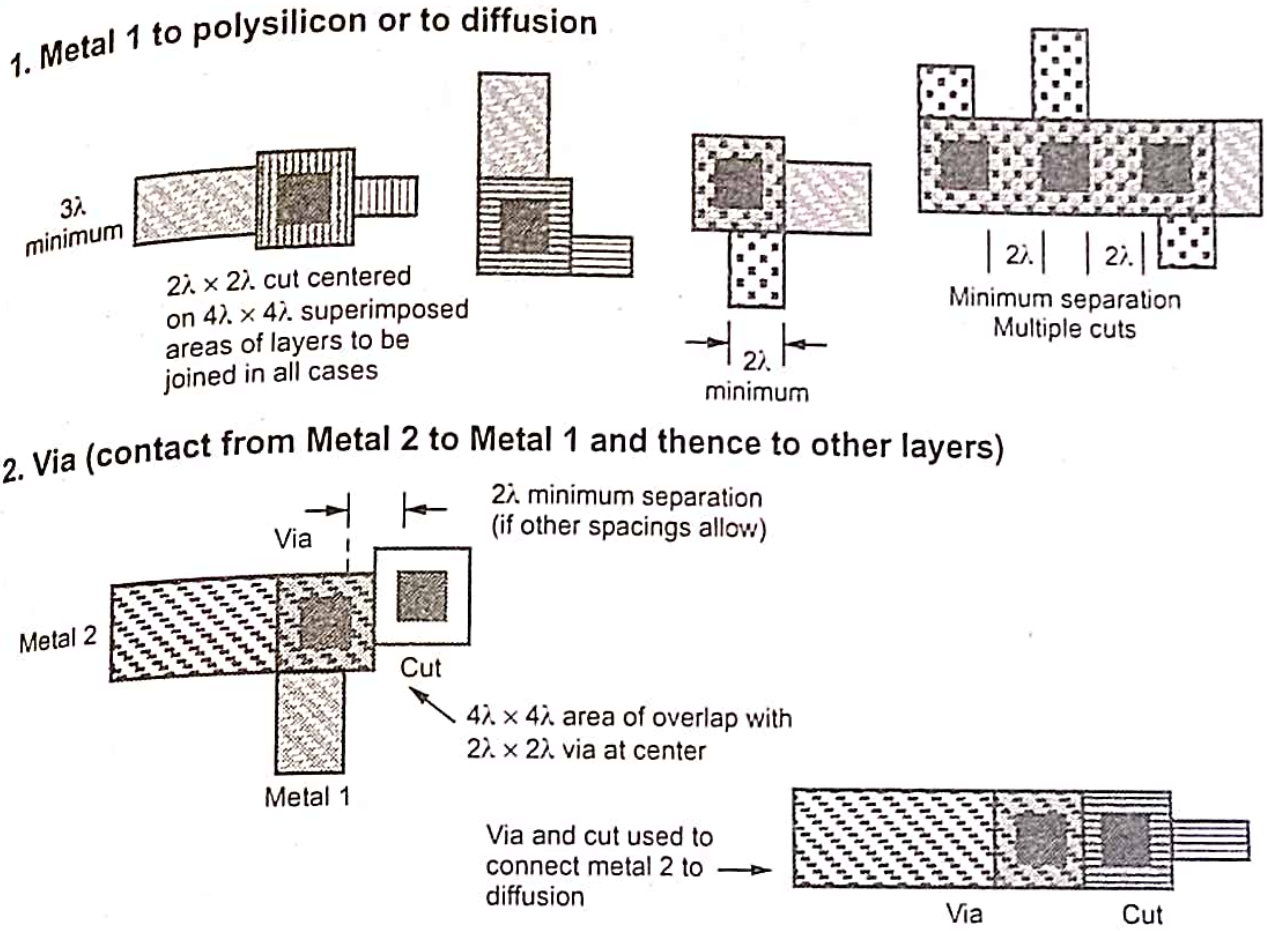


Fig. 2.5.1 Layout diagram of nMOS

- The process involved in establishing a butting contact is quite complicated. Here, a contact cut of $2\lambda \times 2\lambda$ is made down upto each of the layers to be joined. The layers are butted together is such a way that these two contact cuts become configuous. The polysilicon and diffusion outlines overlap and hence the thin oxide (thinox) under poly acts as a mark in the diffusion process. The diffusion and polylayers are also butted together.
- Finally, the contact between the two butting layers is realized by a metal overlay as shown in Fig. 2.5.3 (b) which also shows the cross-sectional view for better understanding of the nature of the contact.
- Compare this with the buried contact shown in Fig. 2.5.2 and Fig. 2.5.3. The contact is simpler. The contact cut, shown by the broken line, indicates where the thin oxide is to be removed to expose the surface of the silicon wafer before deposition of polysilicon. Thus, the polysilicon gets deposited directly on the underlying crystalline wafer. During diffusion, impurities diffuse into the poly and into the diffusion region within the contact area as well. This ensures a satisfactory contact between poly and diffusion.

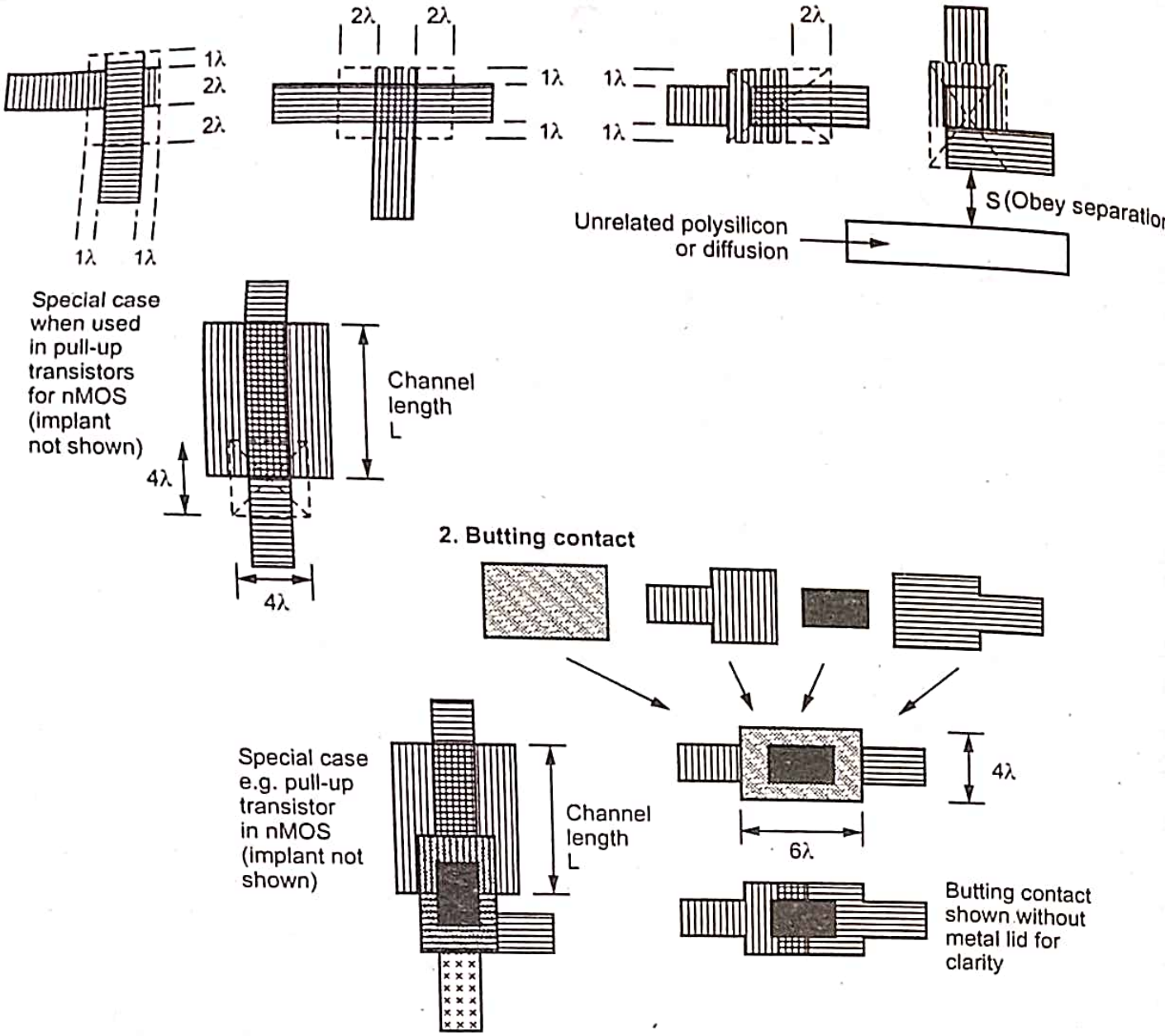
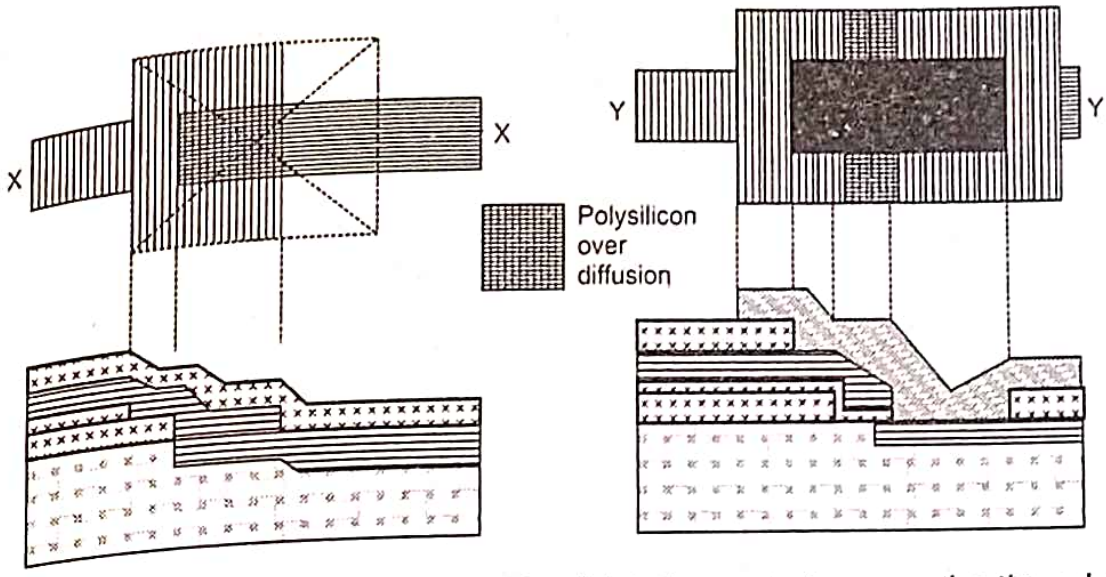


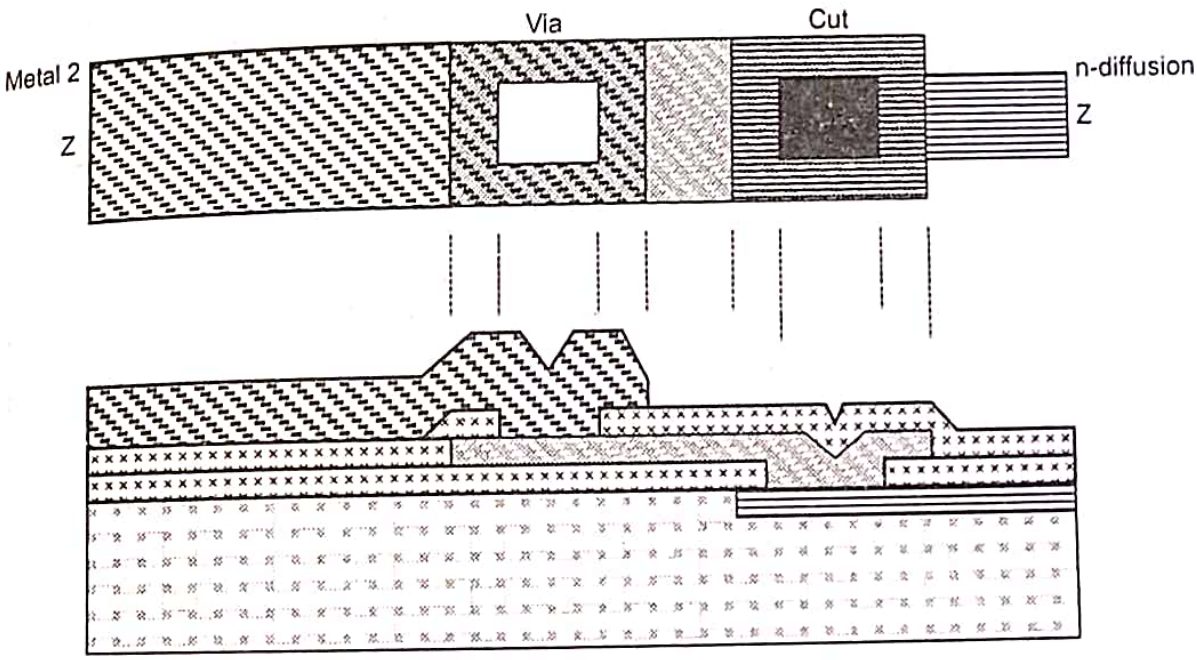
Fig. 2.5.2 Contacts polysilicon to diffusion (only nMOS)

- **Advantages of buried contact are :**
  i) It is subject to fewer design rule restrictions.
  ii) It is smaller in area (than butting contact).
- The design rules pertaining to contact cuts are formulated such that a reason contact area is established and that there will be no unintentional transi formation in series with the contact. The design rules also ensure that there is unwanted diffusion to polysilicon contacts and thus protect the gate oxide of transistors that are in the vicinity of the buried contact cut area.



(a) Buried contact ..... section through XX (b) Butting contact ..... section through YY

Contact from metal 2 to n-diffusion (not using minimum spacing via to cut)



(c) Metal 2-via-metal 1 cut-n-diifusion connection .....section through ZZ

Fig. 2.5.3 Cross-section through some contact structures

2.5.1.4 Metal Rules

- Metal spacing changes with the width of metal line.
  - Above some metal wire width, the minimum spacing may be increased (minimum metal width rule).
  - Single metal wires cannot be greater than certain width (Maximum metal width rules).
  - Spacing rules for long, closely spaced parallel metal lines.

Mask Summary

- Metal spacing dependent on width i.e. as the width increases, spacing increases.

2.5.1.5 Via Rules

- Some processes support stacked vias to be placed over polysilicon and diffu regions. Some processes allow vias to be placed within these areas, and do allow the vias to go out of boundary of polysilicon or diffusion.
- In modern planarized approach, the stacked vias reduces the area required to from a lower level metal to high-level metal.

Mask summary

- Within a layer all vias are of uniform size. The vias may increase in size tow the top of a metal stack. For example, large vias required on power buses constructed from an array of uniformly sized vias.

2.5.2 Scribe Line and Other Structures

- The completed chip is surrounded by scribe line. The scribe line construc varies from manufacturer to manufacturer. The scribe line is designed to prev ingress of contaminants from side of chip.
- Several other structures on mask are -
  1. Alignment mark
  2. Critical dimension structures
  3. Vernier structures
  4. Process check structures.
- The alignment mark is placed by boundry to align one mask to the next. T critical dimension structure can be measured after processing to check prop etching of polysilicon or metal lines.
- Vernier structures are used to judge the alignment between layers. A vernier is collection of closely spaced parallel lines on two layers. The mis-alignment of tw layers can be judged by two verniers.

2.5.3 MOSIS Scalable CMOS Design Rules

- The λ-based scalable CMOS design rules from MOSIS are simple, freely availab and allow designs to easily migrate from one process to other. The MOSIS h three sets of rules : - SCMOS, SUBM and DEEP.
- Some MOS boundary processes and associated values of λ for different rule se are summarized in Table 2.5.1.

Vendor	Feature size (μm)	Inter connect layers	Stacked vias	Sets of MOSIS		
				SCMOS (λ)	SUBM (λ)	DEEP (λ)
Orbit	2.0	2 metal	No	1.0 μm		
AMI	1.5	2 metal, 2 poly	No	0.80 μm	0.80 μm	
AMI	0.5	3 metal, 1-2 poly	Yes	0.35 μm	0.30 μm	
TSMC	0.35	4 metal, 1-2 poly	Yes		0.20 μm	
TSMC	0.25	5 metal	Yes		0.15 μm	0.12 μm
TSMC	0.18	6 metal	Yes		0.10 μm	0.09 μm

Table 2.5.1 MOS design rules options

2.5.4 Micron Design Rules

- A set of micron design rules for a hypothetical 90 nm process representating an amalgamation of several real processes is summarized in Table 2.5.2.

Layer	Description	90 nm rule (μm)
Well	Width	0.75
	Spacing to well different potential	1.5
	Spacing to well at same potential	1.0
Active (diffusion)	Width	0.15
	Spacing to active	0.20
	Substrate/well contact surround by well	0.25
	Spacing to active of opposite type	0.30
Poly	Width	0.09
	Spacing to poly over field oxide	0.15
	Spacing to poly over active	0.15
	Gate extension beyond active	0.15
	Active extension beyond poly	0.15
	Spacing poly to active	0.10

Select	Spacing from substrate/well contact to gate	0.25
	Overlap of active	0.20
	Overlap of substrate/well contact	0.10
	Spacing to select	0.30
Metal 1	Width	0.13
	Spacing to well metal 1	0.13
	Overlap of contact or via	0.01
	Spacing to metal for lines wider than 0.5 μm	0.40
Metal 2 - Metal 6	Width	0.15
	Spacing to same layer metal	0.15
	Overlap to via	0.01
	Spacing to metal for lines wider than 1.0 μm	0.40
Metal 7	Width	0.40
	Spacing to metal 7	0.40
	Overlap of via 6	0.10
	Spacing to metal 7 for lines wider than 1.0 μm	0.50

Table 2.5.2 Micron design rules (for 90 nm process)

2.6 Short Questions and Answers

Q.1 What are the advantages of BiCMOS process compare with the CMOS.

Ans. : Advantages of BiCMOS process compared to CMOS

1. Using BiCMOS process, design of analog amplifier is possible.
2. Except input circuit, the remaining stages and output drivers are realized usi bipolar transistors.
3. BiCMOS technology is used for I/O and driver circuits while CMOS technolo is used for logic only.
4. BiCMOS technology is low delay sensitive to load.
5. BiCMOS has high output drive currents.

Q.2 Compare CMOS and BiCMOS logic.

Ans. :

Sr. No.	CMOS	BiCMOS
1.	Low static power dissipation.	High power dissipation
2.	High input impedance	Low input impedance
3.	Low gm	High gm
4.	Low output drive current	High drive current
5.	High delay sensitivity to load	Low delay sensitivity to load
6.	High package density	Low package density

Q.3 What are the different MOS layers ?

Ans. : Different MOS layers -

1. n - diffusion 2. p - diffusion 3. Polysilicon 4. Metal

Q.4 Explain about the contact cuts and approaches.

Ans. : Contact cuts -

- Electrical connection can be established between layers by special structures called contact cuts.
- There are three possible approaches for establishing contacts between polysilicon and diffusion in nMOS circuits. They are :
  i) Poly to metal and then metal to diffusion.
  ii) A buried contact-poly to diff.
  iii) A butting contact-poly to diff. using metal.

Q.5 What is meant by synthesis ?

Ans. : Synthesis -

- Synthesis is the process of transforming your HDL design into a gate-level netlist, given all the specified constraints and optimization settings.
- Logic synthesis is the process of translating and mapping RTL code written in HDL (such as Verilog or VHDL) into technology specific gate level representation.

Q.6 Write about contacts and vias in layout design.

Ans. : Contacts and vias

- The connection material between diffusion and metal is called "contact". The same layer is also used to connect poly to metal, or poly2 to metal.
- The connection material between metal and metal2 is called "via". By extension, the material that connects metal2 to metal3 is "via2", metal3 to metal4 "via3", etc..

□□□

Notes

3 MOS Transistor

Syllabus

The Metal Oxide Semiconductor (MOS) structure, The MOS System underexternal bias, Structure and Operation of MOS transistor, MOSFET Current-Voltage characteristics, MOSFET scaling and small-geometry effects, MOSFETcapacitances.

Contents

3.1 MOS Structure ..... **Summer-15, Winter-21** ··· Marks 7

3.2 MOS Under External Bias ..... **Summer-21, Winter-21,** ··· Marks 7

3.3 Operation of MOS **Winter-14, 21,**

..... **Summer-16, 21,** ····· Marks 7

3.4 MOSFET Current-Voltage Characteristics **Winter-14, 21,**

..... **Summer-15, 21,** ····· Marks 7

3.5 MOSFET Scaling

3.6 MOSFET Capacitances ..... **Winter-14, Summer-14,** ··· Marks 7

3.7 Short Questions and Answers

3.1 MOS Structure

GTU : Summer-15, Winter

- Consider a simple two terminal MOS structure having three layers as :
  1. Metal gate electrode layer.
  2. Insulating oxide (SiO_2) layer.
  3. P-type bulk semiconductor (Si) or substrate.
- Under equilibrium the mobile carriers in semiconductor obeys the Mass A Law.

$$n \cdot p = ni^2$$

where, n represents electron concentration

p represents hole concentration

n_i represents intrinsic carrier concentration of silicon

(1.45×10^{10} cm^{-3} at T = 300 K)

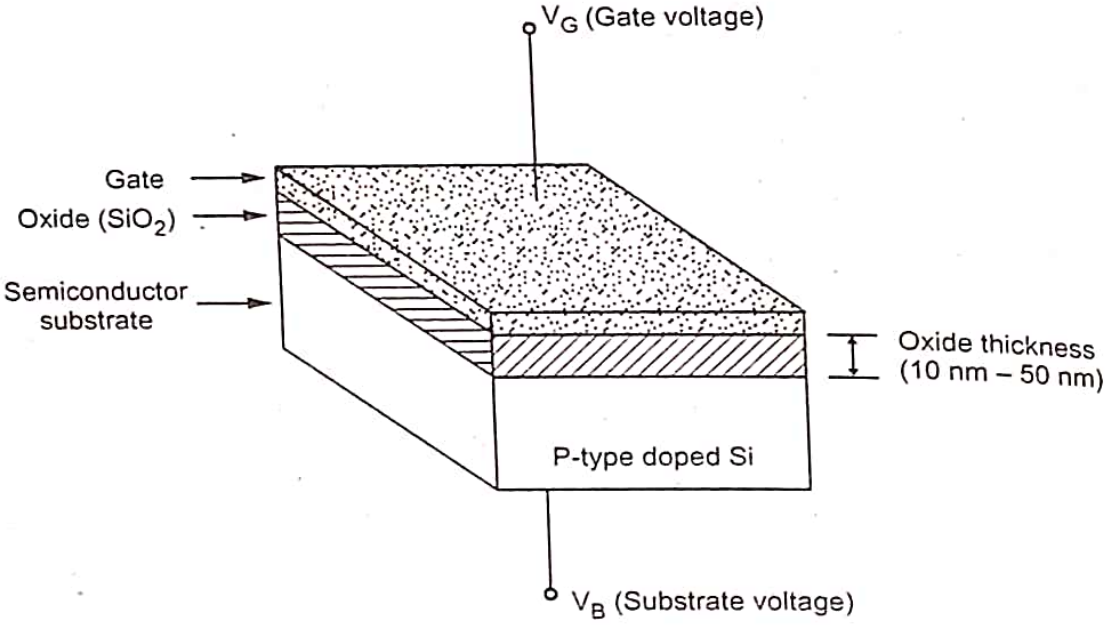


Fig. 3.1.1 Two terminal MOS structure

- The equilibrium electron and hole concentrations in P-type substrat approximated by -

$$n_{po} \cong \frac{ni^2}{N_A}$$

$$p_{po} \cong N_A$$

where, N_A represents acceptor concentration typically 10^{15} to 10^{16} cm^{-3}.

- The energy band diagram of p-type substrate is shown in Fig. 3.1.2.

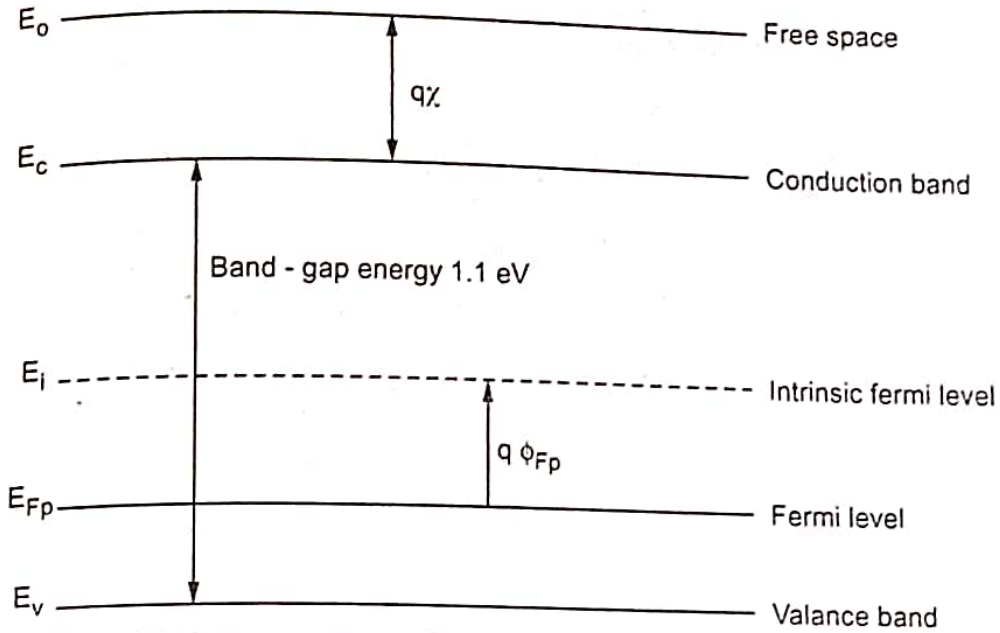


Fig. 3.1.2 Energy band diagram of p-type silicon substrate

- For silicon the band gap between conduction band and valance band is approximately 1.1 eV. The equilibrium Fermi level E_F within the band gap is determined by doping type and concentration in silicon substrate. The Fermi potential ϕ_F is given by equation -

$$\phi_F = \frac{E_F - E_i}{q}$$

For p-type semiconductor

$$\phi_{Fp} = \frac{kT}{q} ln \frac{n_i}{N_A}$$

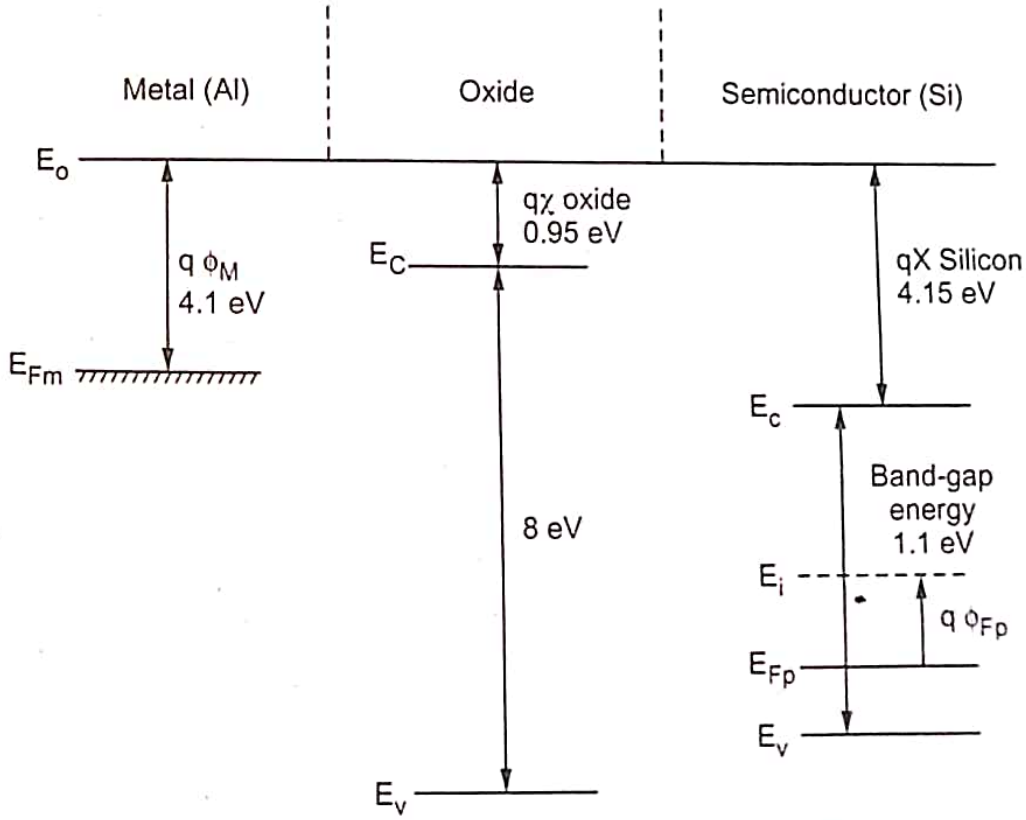


Fig. 3.1.3 Energy band diagram of MOS

where, k is Boltzmann constant

q is unit charge on electron

- The energy required for an electron to move from fermi level into free space is work function $q\phi_S$.

$$q\phi_S = q\chi + (E_C - E_F)$$

where, $q\chi$ represents electron affinity of silicon.

- The energy band diagram of metal, oxide and semiconductor layers in MOS system as three separate components is shown in Fig. 3.1.3.
- The insulating SiO_2 layer between silicon substrate and gate has a large band gap of about 8 eV and an electron affinity of 0.95 eV. The work function $q\phi_M$ of aluminium gate is 4.1 eV.
- When all three components of MOS are brought together physically. The Fermi levels of all three components are lined up to form MOS capacitor. The combined energy band diagram of MOS system is shown in Fig. 3.1.4.

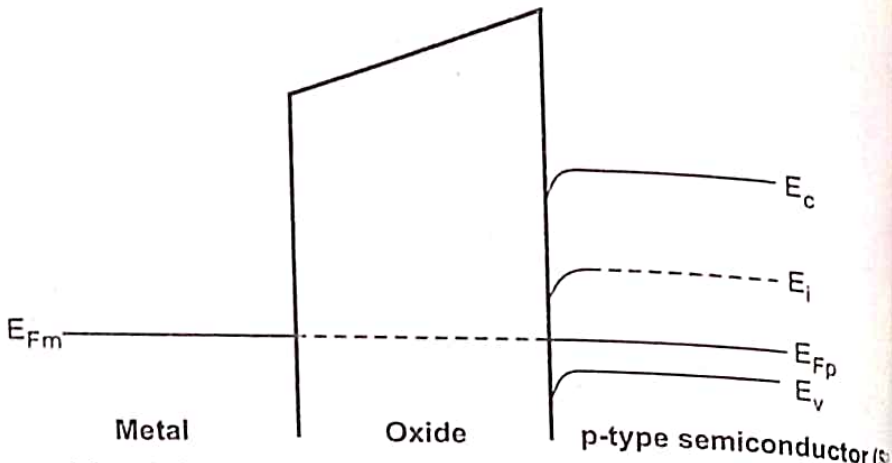


Fig. 3.1.4 Energy band diagram of combined MOS

- Due to work function difference between metal and semiconductor, a voltage drop occurs across the MOS system. The equilibrium fermi levels of semiconductor (S) and metal are almost at same potential. The fermi potential at surface is called surface potential ϕ_S which is smaller than bulk fermi potential ϕ_F.

University Questions

1. *Draw the energy band diagram of MOS structure at surface inversion and derive the expression for the maximum possible depth of the depletion region.* **GTU : Summer-15, Marks 7**
2. *Explain the band diagram of MOS structure at surface inversion and derive the expression for threshold voltage.* **GTU : Winter-21, Marks 7**

3.2 MOS Under External Bias

GTU : Summer-21, Winter-21

- Let the substrate voltage is set at $V_B = 0$; and the controlling parameter is gate voltage V_G. According to polarity and magnitude of V_G there exists three different operating regions for the MOS. These three regions are :

1. Accumulation 2. Depletion 3. Inversion

- The depth of depletion region is given by expression :

$$x_d = \sqrt{\frac{2\varepsilon_{si} \cdot |\phi_S - \phi_F|}{q \cdot N_A}} \quad \text{... (3.2.1)}$$

where, ϕ_S is surface potential

ϕ_F is bulk fermi potential

- The depletion region charge density is given by expression.

$$Q = -q \cdot N_A \cdot x_d$$

$$Q = -\sqrt{2q \cdot N_A \cdot \varepsilon_{SI} \cdot |\phi_S - \phi_F|} \quad \text{... (3.2.2)}$$

- The amount of depletion region charge is important in analyzing threshold voltage.

nversion in MOS system

If V_{ds} is increased to level greater than $V_{gs} - V_t$), in such condition IR drop = $(V_{gs} - V_t)$ akes place over less than whole length of channel. herefore near drain there is insufficient electric eld available to give rise to an **inversion layer** to reate a channel. The channel is therefore **Pinched** ff as shown Fig. 3.2.1.

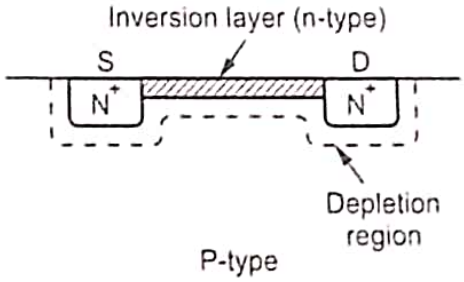


Fig. 3.2.1

apered channel in MOS transistor

When V_{DS} increases a few tenths of a volt > 0) : The depletion region near the drain idens (N+ drain is positively biased - . reverse biased with respect to the ıbstrate). The electron concentration in the version layer near the drain decreases as ey are sucked out by the drain voltage. hannel conductance decreases resulting in a op in the slope of the I_D - V_D curve.

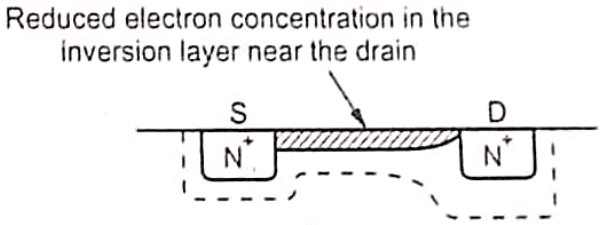


Fig. 3.2.2

The inversion layer eventually vanishes near the drain end of the channel, thus rmine a tapered channel.

2.1 NMOS Transistor

- MOS technology circuit can be realized as one of the following
  1) NMOS 2) PMOS 3) CMOS 4) BiCMOS

NMOS :

- NMOS devices are fabricated on a
source and drain regions are form
depletion region. This depletion re
p-region. The connections to source
layer and are isolated from one ano
- The current through source and dra
the two modes of NMOS transistor.
1) NMOS enhancement mode 2) NMOS depletio

Enhancement Mode :

- In enhancement mode device, a polysilicon gate is
over the region between source and drain as shown in

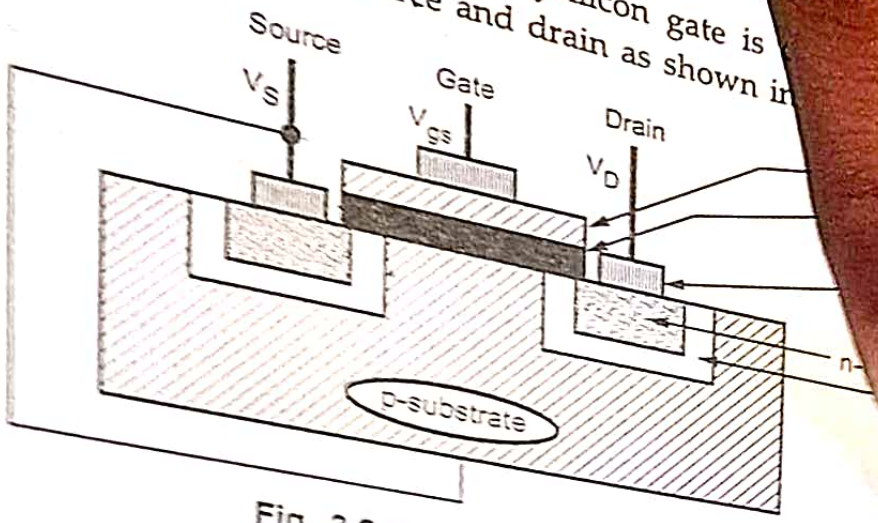


Fig. 3.2.3 NMOS enhancement mode

- When gate is not biased and $V_D = V_S = V_{gs} = 0$, channel is not established hence device is not conducting.
- When gate is connected to positive voltage with respect to source, then electric field is established between gate and substrate. Because of this, there is charge inversion in substrate under gate insulation forming a conducting path between source and drain.

Depletion Mode :

- The channel can be established even at $V_{gs} = 0$ by adding suitable impurities between drain source terminal and before depositing insulation at gate. This arrangement is shown in Fig. 3.2.4 and is called as depletion mode.
- In depletion mode, the source and drain are connected by a conducting channel the channel width can be controlled by applying suitable negative voltage to the gate terminal. The variation in gate voltage allow control of current flow between source and drain.

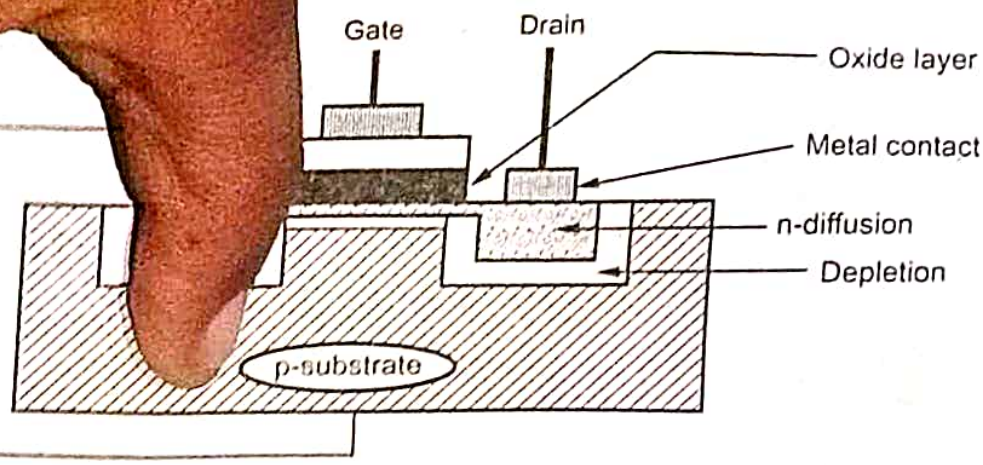


Fig. 3.2.4 NMOS depletion mode

3.2. Transistor

MOS

- B transistor structure for enhancement mode device is shown in Fi

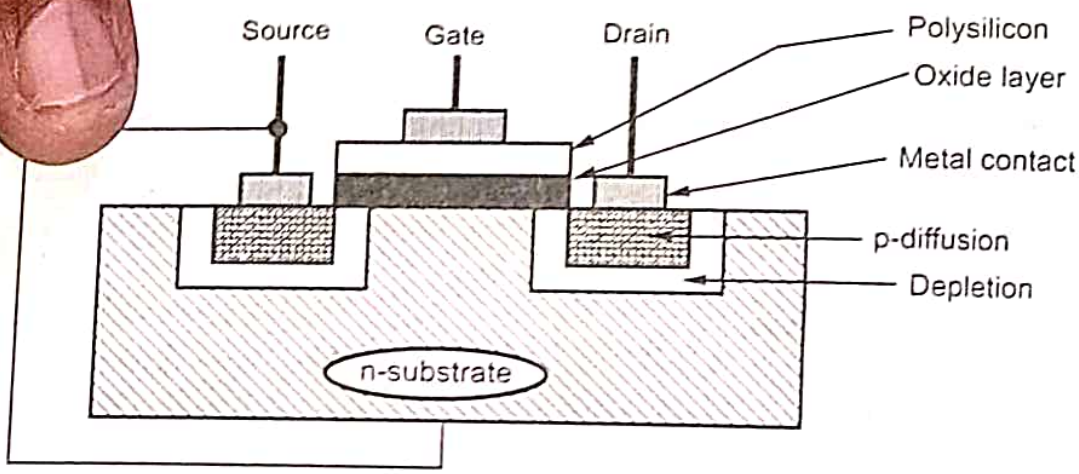


Fig. 3.2.5 PMOS enhancement mode

Application of negative voltage ($> V_{th}$) between gate and source will give rise to the formation of p-type channel between the source and drain terminal. If the drain is made negative with respect to source current can flow through channel. Here the current is carried by holes rather than electrons. The hole mobility (μ_p) compared to electron mobility (μ_n) is about 2.5 times less. Therefore, PMOS devices are inherently slower than NMOS.

3 Comparision of NMOS and PMOS

No.	NMOS	PMOS
	The mobility of electrons (carriers) in n-channel device is two times greater than holes (carriers of PMOS); thus NMOS device is faster than p-channel device.	PMOS device is slower than NMOS.

2.	High static power dissipation.	Low static power dissipation.
3.	NMOS is used for pull-down network.	PMOS is used for pull-up network.
4.	Fabrication process is short and easy.	Large steps of fabrication is required.
5.	NMOS devices have low noise immunity.	PMOS devices show high noise immunity.

University Questions

1. *When inversion is occured in MOS system ?* **GTU : Summer-17, Marks 2**
2. *Why channel is tapered in MOS transistor ?* **GTU : Summer-17, Marks 2**
3. *Explain the behaviour of MOS device under external bias with the help of energy band diagram and derive the relationship for maximum depletion width at oxide semiconductor surface.* **GTU : Summer-21, Marks 7**
4. *Explain MOS system under external bias for nMOS transistor in detail.* **GTU : Winter-21, Marks 7**

3.3 Operation of MOS

GTU : Winter-14, 21, Summer-16, 21

3.3.1 Enhancement Mode

- To understand the operation of transistor in enhancement mode, three different conditions are considered. The fundamental requirement for establishing a channel is application of threshold voltage (V_t) between gate and source.

Condition-1

- In this condition, a channel between drain and source is established but no current is flowing between source and drain i.e. $V_{ds} = 0$. Fig 3.3.1 illustrates this condition.

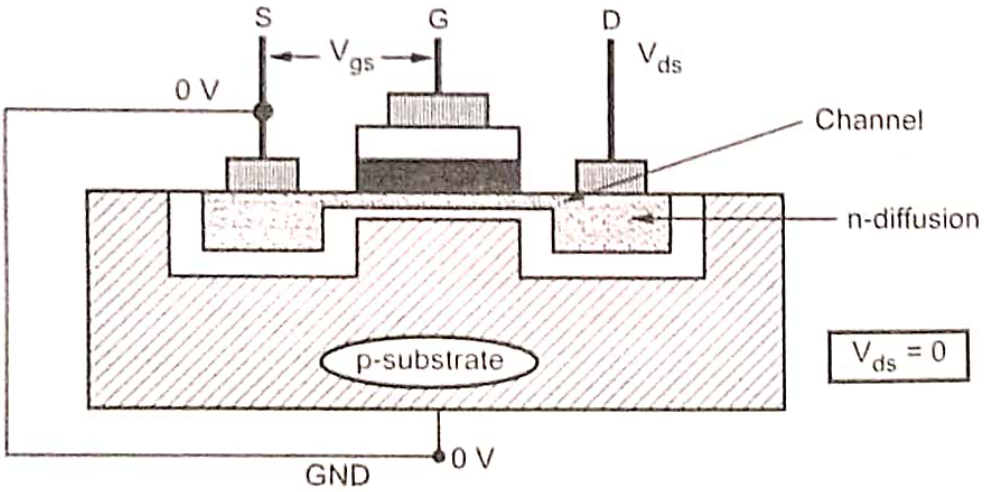


Fig. 3.3.1 Enhancement mode for $V_{ds} = 0$

Condition-2

- When voltage V_{ds} is applied between drain and source and current starts flowing through channel. The resistance of channel varies along its length this results in IR drop along the channel. The voltage is maximum at source end. Fig 3.3.2 shows the channel when $V_{ds} < V_{gs} - V_t$.

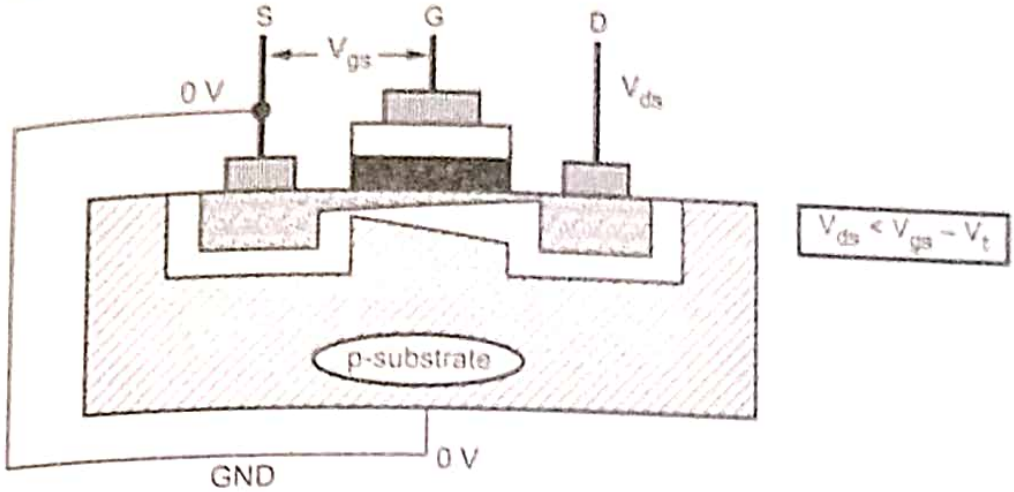


Fig. 3.3.2 Enhancement mode for $V_{ds} < V_{gs} - V_t$

- The effective gate voltage is $V_g = (V_{gs} - V_t)$ as no current flows when $(V_{gs} < V_t)$ As long as $(V_{gs} - V_t) \geq V_{ds}$, the channel will be inverted at the drain end.
- When $V_{ds} < (V_{gs} - V_t)$, a condition exists. $V_{ds} = (V_{gs}) - V_t$, the device is in non-saturated region of operation. This region is also called as triode region or ohmic region.

Condition-3

- If V_{ds} is increased to level greater than $(V_{gs} - V_t)$, in such condition IR drop = $(V_{gs} - V_t)$ takes place over less than whole length of channel. Therefore near drain there is insufficient electric field available to give rise to an inversion layer to create a channel. The channel is therefore **Pinched off** as shown in Fig. 3.3.3.

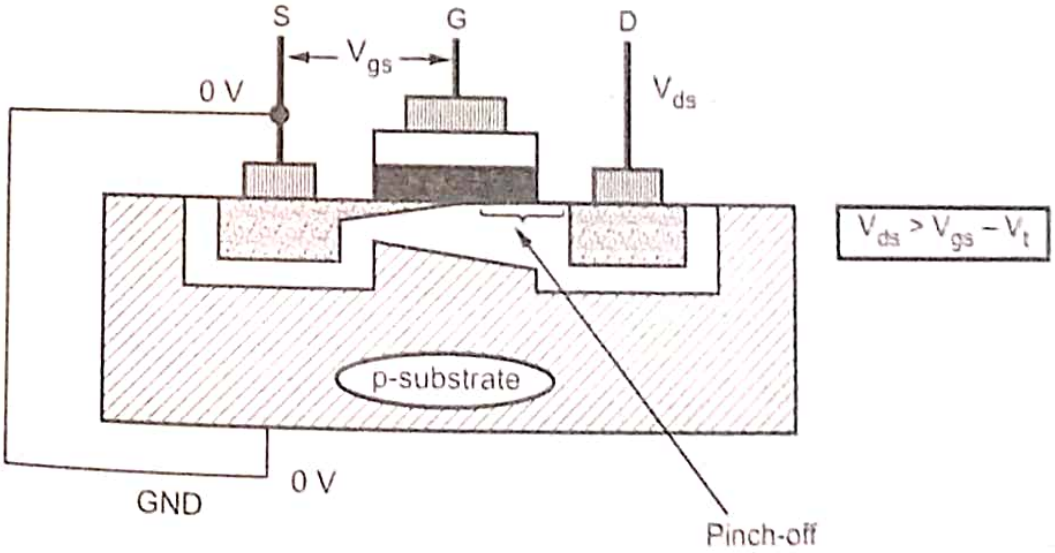


Fig. 3.3.3 Enhancement mode for $V_{ds} > V_{gs} - V_t$

- In pinch-off condition, diffusion current completes the path from source to c This causes channel to exhibit a high resistance and behaves as a constant cu source. The device operates in saturation region. The devices behaves as a cor source for increase of V_{ds} above $V_{ds} = (V_{gs} - V_t)$. Increasing V_{ds} beyond value has little effect on the channel shape and current through channel rer constant. Fig. 3.3.4 shows I_{ds} versus V_{ds} characteristics for enhancement NMOS transistor.

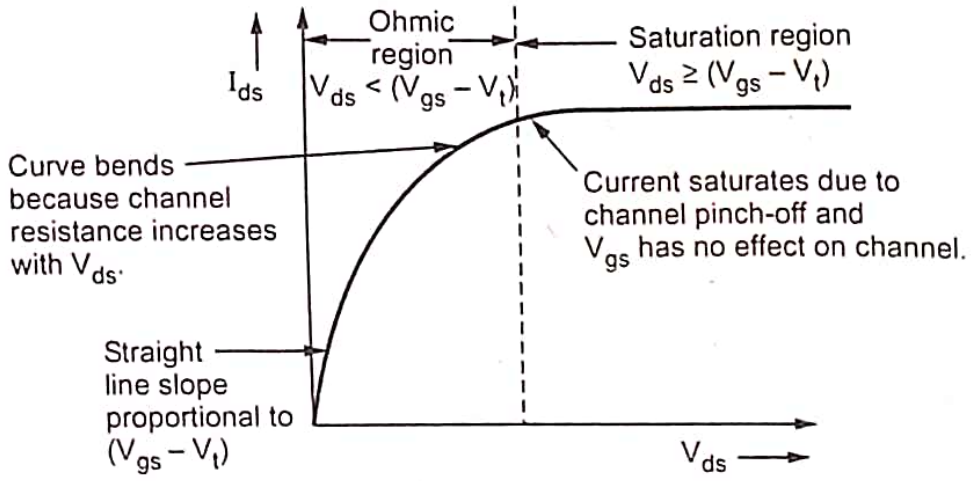


Fig. 3.3.4 I_{ds} versus V_{ds} characteristics for enhancement type NMOS transisto

Symbols of NMOS and PMOS transistors

nMOS enhancement **pMOS enhancement**

Fig. 3.3.5 Symbols of enhancement mode transistors

Drain current in saturation region in pinch-off

The drain-source voltage, V_{DS}, at which this occurs is called the saturation v V_{sat} while the current is called the saturation current, I_{Dsat}.

For $V_{DS} > V_{sat}$ the channel length (L) effectively changes by a value ΔL. The of the channel, ΔL is depleted and thus, is high resistivity. Accordingly, alm voltage increases in $V_{DS} > V_{sat}$ are dropped across this portion of the channel.

If $\Delta L << L$, the voltage at the end of the channel will be constant (V_{sat}) $V_{DS} > V_{sat}$. I_D will be constant.

3.3.2 Depletion Mode

- In depletion mode, the channel is established due to implant of impurities. The channel ceases as negative voltage V_{td} applied between gate and source even when $V_{gs} = 0$.
- V_{td} is usually less than – 0.8 V_{DD} depending on implant and substrate bias. Threshold voltage also determines V_{td}.

Symbols of PMOS and NMOS transistors

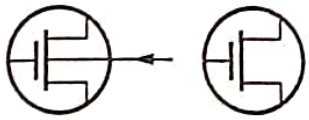


Fig. 3.3.6 NMOS depletion mode

3.3.3 Threshold Voltage

- The **threshold voltage** of a MOSFET is defined as the gate voltage where an inversion layer forms at the interface between the insulating layer (oxide) and the substrate (body) of the transistor.
- The gate structure of MOS transistor consists of charges stored in the dielectric layers, surface to surface interfaces and in the substrate itself. The threshold voltage decided by the structural details of gate-channel structure.
- For switching an enhancement mode MOS transistor from OFF to ON state necessitates applying sufficient gate voltage to neutralize these charges and enable the underlying silicon to undergo an inversion due to electric field from the gate.
- For switching an depletion mode nMOS transistor from ON to OFF state consists in applying enough voltage to the gate to add to the stored charge and invert the 'n' implant region to 'p' region.
- Threshold voltage is defined as gate voltage for which strong inversion occurs. Threshold voltage depends on several factors.
- The threshold voltage (V_t) is given by -

$$V_t = \phi_{ms} \frac{Q_B - Q_{SS}}{C_0} + 2\phi_{fN} \quad \text{... (3.3.1)}$$

where,

Q_B is the charge per unit area in the depletion layer beneath the oxide.

Q_{SS} is the charge density at Si : SiO_2 interface.

C_0 is capacitance per unit gate area.

ϕ_{ms} is work function difference between gate and Si. (– 0.9 V over p-substra and – 0.2 V over n-substrate).

ϕ_{fN} is Fermi level potential between inverted surface and bulk Si.

- The value of ϕ_{ms} is negative and negligible in case of polysilicon gate and silic substrate. Hence the magnitude and sign of V_t are decided by the balance betwe the remaining negative term $\frac{-Q_{SS}}{C_0}$ and the other two terms, both these terms a positive.
- The expression for V_t can be evaluated through the following data :

$$Q_B = \sqrt{2\varepsilon_0 \varepsilon_{si} q N(2\phi_{fN} + V_{SB})} \text{ coulomb / m}^2 \quad \text{... (3.3}$$

$$\phi_{fN} = \frac{kT}{q} \ln \frac{N}{n_i} \text{ volts} \quad \text{... (3.3}$$

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

where,

V_{SB} is substrate bias voltage (negative with respect to source for nMOS, posit for pMOS).

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

N is impurity concentration in substrate (N_A or N_D)

ε_{si} is relative permittivity of silicon (≈11.7)

n_i is intrinsic electron concentration (1.6×10^{10} /cm^3 at 300 °K)

k is Boltzmann's constant (1.4×10^{-23} Joule/°K)

- Sometimes a potential may exist between source and substrate as a result of body effects. Then the threshold voltage may get altered. All MOS devices are fabricated on a common substrate. Substrate voltage of all devices is equal. However, in practice several devices may be required to be connected in series for realizing gating functions as shown in Fig. 3.3.7.

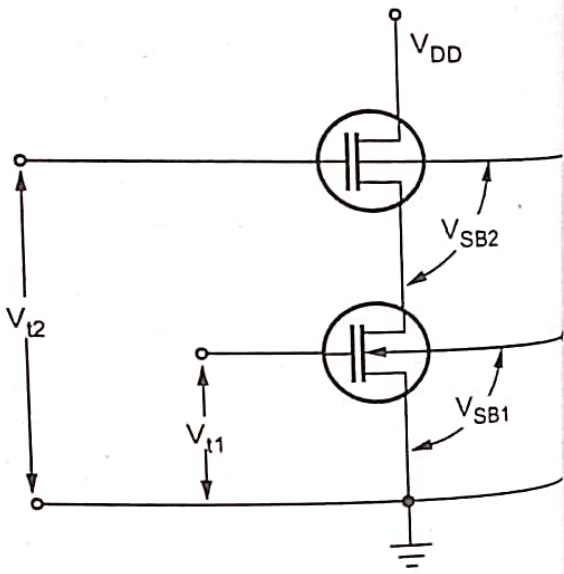


Fig. 3.3.7 Substrate bias on series connected transistors

- Under normal circumstances ($V_{GS} > V_t$), the depletion layer width remains constant and charge carriers get pulled into the channel from the

source. Increasing V_{SB} causes the channel to be depleted of charge carriers and thus the threshold voltage is raised. The density of trapped carriers increases within depletion layer. In order to maintain charge neutrality, the channel charge decreases. The ultimate effect is substrate voltage V_{SB} adds to the junction potential of channel-substrate junction. This increases gate-channel potential drop and hence threshold voltage.

- The change in threshold voltage is given by -

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

where,

γ is a constant which depends on substrate doping so that more lightly doped substrate smaller will be body effect.

V_t can be written as,

$$V_t = V_t(0) + \left[\frac{D}{\varepsilon_{ins}\varepsilon_0}\right]\sqrt{2\varepsilon_0\varepsilon_{si}QN}\,(V_{SB})^{\frac{1}{2}} \quad \text{... (3.3.4)}$$

where,

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

- Threshold voltage V_t can also be written as,

$$V_t = V_t(0) + \gamma\left[\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|}\right] \quad \text{... (3.3.5)}$$

where, $\gamma = \dfrac{\sqrt{2\,q\varepsilon_{si}\,N}}{C_{ox}}$

γ is called as body effect coefficient or body factor.

- Expression shows that threshold voltage is affected by material parameters and substrate voltage. The threshold voltage of an n-channel enhancement type MOSFET is positive and p-channel enhancement type MOSFET is negative.
- The magnitude of body effects can be understood by typical values of V_t, which are as under -

For nMOS enhancement mode transistors :

$$\left.\begin{array}{ll} V_{SB} = 0; & V_t = 0.2V_{DD} \\ V_{SB} = 5\,V; & V_t = 0.3V_{DD} \end{array}\right\} \text{Values for pMOS are similar but negative}$$

For nMOS depletion mode transistors :

$V_{SB} = 0\,V;\quad V_{td} = -0.7\,V_{DD}$

$V_{SB} = 5\,V;\quad V_{td} = -0.6\,V_{DD}$

Example 3.3.1 *If $V_t(0) = -0.4$ V under $V_{SB} = 0$ then calculate V_T when $V_{SB} = -2.5$ V an $2\phi_F = 0.6$ V for a body effect coefficient $= -0.4$.*

Solution : Given

$$V_t(0) = -0.4 \text{ V} \qquad V_{SB} = -2.5 \text{ V}$$

$$\gamma = -0.4 \qquad 2\phi_F = 0.6$$

Threshold voltage is given by,

$$V_t = V_t(0) + \gamma\left[\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|}\right]$$

$$V_t = -0.4 - 0.4\left[\sqrt{0.6 + |-2.5|} - \sqrt{0.6}\right]$$

$$V_t = -0.4 - 0.4\left[\sqrt{0.6 + 2.5} - \sqrt{0.6}\right]$$

$$V_t = -0.4 - 0.4\left[\sqrt{3.1} - \sqrt{0.6}\right]$$

$$V_t = -0.4 - 0.3944$$

$$\mathbf{V_t = -0.7944 \text{ V}}$$

Example 3.3.2 *Calculate the threshold voltage V_{to} at $V_{sb} = 0$, for a polysilicon gate n-chan MOS transistor, with the following parameters: substrate doping density $Na = 10^{16}$ cm plysilicon gate doping density $N_D = 10^{20}$ cm^{-3}, gate oxide thickness $t_{ox} = 500$ A° oxide interface fixed charge density $N_{ox} = 2\times10^{10}$ cm^{-2}, Thermal voltage $= KT/q$ $= 0.026$ V, energy gap of silicon (Eg) $= 1.12$ eV, $ni = 1.45\times10^{10}$ cm^{-3}, $\varepsilon_o = 8.85\times10^{-14}$ F/cm, $\varepsilon_{ox} = 3.97\varepsilon_o$ and $\varepsilon_{si} = 11.7\varepsilon_o$.* **GTU : Summer-16, Mark**

Solution : Fermi potential of polysilicon $\phi_F = 0.55$ V

$$\phi_{Gc} = \phi_F(\text{substrate}) - \phi_F(\text{gate}) = -0.35 \text{ V} - 0.55 \text{ V} = -0.9 \text{ V}$$

Depletion region charge density at $V_{SB} = 0$

$$\phi_{B0} = -\sqrt{2\cdot q - N_A \cdot \varepsilon_{si}|-2\phi_F(\text{substrate})|}$$

$$= -\sqrt{2(1.6\times10^{-19})(10^{16})(11.7)(8.85\times10^{-14})|-2\times0.35|}$$

$$= -4.82\times10^{-8} \text{ C/cm}^2$$

Oxide free charge

$$Q_{ox} = q\cdot N_{ox} = 1.6\times10^{-19}\text{C}\times4\times10^{10}\text{ cm}^{-2} = 6.4\times10^{-9} \text{ C/cm}^2$$

Gate oxide capacitance

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \frac{3.97\times8.85\times10^{-14}\text{ F/cm}}{500\times10^{-8}\text{ cm}} = 7.03\times10^{-8} \text{ F/cm}^2$$

Threshold voltage

$$V_{T0} = \phi_{Gc} - 2\phi F(\text{substrate}) - \frac{QB_0}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} = -0.90 - (-0.70) - (-0.69) - 0.09$$

$$= 0.40 \text{ V}$$

Example 3.3.3 *Consider a MOS system with the following parameters :*
$t_{ox} = 20$ *nm*, $\Phi_{GC} = -0.85$ *V*, $N_A = 2\times10^5$ cm^{-3}, $Q_{ox} = q\times 2\times 10^{11}$ C/cm^2
i) Determine the threshold voltage, V_{T0} *under zero bias at room temperature (T = 300 °K)*
Consider $\varepsilon_{ox} = 3.97\,\varepsilon_0$ *and* $\varepsilon_{si} = 11.7\,\varepsilon_0$
ii) Determine the type (p-type or n-type) and amount of channel implant (N_1/cm^2) *required to change the threshold voltage to 0.8 V.* **GTU : Summer-21, Marks 7**

Solution : Fermi potentials for the p-type substrate and for the n-type polysilicon gate :

$$\phi_F(\text{substrate}) = \frac{kT}{q}\ln\left(\frac{n_i}{N_A}\right)$$

$$\phi_F = 0.026\text{ V}\times\ln\left(\frac{1.45\times10^{10}}{2\times10^5}\right) = 0.2909 \text{ V}$$

Depletion region charge density at $V_{SB} = 0$ is

$$\phi_{Bo} = \sqrt{2\cdot q\cdot N_A\cdot\epsilon_{si}\,|-2Q_{F(\text{substrate})}|}$$

$$= \sqrt{2\times(1.6\times10^{-19})(2\times10^5)(11.7\times8.85\times10^{-14})|-2\times0.29|}$$

$$= -1.96\times10^{-13}\text{ C / cm}^{-2}$$

Oxide interface charge is

$$Q_{ox} = q\cdot N_{ox} = 1.6\times10^{-19}\text{ C}\times(4\times10^{10})\text{ cm}^{-2}$$

$$= 6.4\times10^{-9}\text{ C / cm}^2$$

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{3.97\times8.85\times10^{-14}\text{ F / cm}}{20\times10^{-7}\text{ cm}} = 1.75\times10^{-7}\text{ F / cm}^2$$

$$V_{T0} = \phi_{GC} - 2\phi_{F(\text{substrate})} - \frac{Q_{Bo}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}}$$

$$= -0.85-(-2\times0.2909)-\left(\frac{-1.96\times10^{-13}}{1.75\times10^{-7}}\right)-\left(\frac{6.4\times10^{-9}}{1.75\times10^{-7}}\right)$$

$$= -0.85-(-2\times0.2909)-(-1.12\times10^{-6})-(0.036)$$

$$= \mathbf{0.304\ V}$$

...Ans.

$$\Delta V = 0.6 - V_{To} = 0.6 - (-0.304) = 0.904$$

$$N_I = \frac{1.04\, C_{ox}}{q} = \frac{1.04\,(1.75\times10^{-7})}{1.6\times10^{-19}} = 1.13\times 10^{12}\ cm^{-2}$$

University Questions

1. *Define threshold voltage (V_T) of MOSFET device. Which are the four physical component which V_T depends ? Derive expression for V_T* **GTU : Winter-14, Mark**
2. *Derive the threshold voltage expression for CMOS two input NAND gate.* **GTU : Winter-21, Mark**
3. *Derive the threshold voltage equation for nMOS transistor.* **GTU : Winter-21, Mark**
4. *Derive equation for depletion depth for inversion region of two terminal MOS structure.* **GTU : Summer-21, Mar**

3.4 MOSFET Current-Voltage Characteristics

GTU : Winter-14, 21, Summer-15

Nonideal I-V Effects

- MOS transistors have three regions of operation :
  1. Cut-off or subthreshold region
  2. Linear or nonsaturation region
  3. Saturation region.
- In cut-off or subthreshold region, where $V_{GS} < V_t$

$$I_{DS} = 0$$

- **Subthreshold drain current** is the current between the source and drai MOSFET when the transistor is in subthreshold region, i.e. for gate-to voltages below the threshold voltage.
- In Linear or nonsaturation region, where $V_{DS} < V_{dsat}$

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

- In saturation region, where $V_{DS} > V_{dsat}$

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

- In all above equations many practical effects are neglected. Practically i and saturation region, the saturation current increases less than quadratic increasing V_{GS} because of velocity saturation and mobility degradation.

Example 3.4.1 *For a CMOS inverter with following specification compute V_{IL}, V_{IH} and find out inverter is symmetric or not ?*

$V_{DD} = 1.8\ V,\quad V_{tn} = 0.5\ V,\quad V_{tp} = -0.5\ V,\quad \mu_n = 580\ cm^2/VS$

$\mu_p = 290\ cm^2/VS,\quad (W/L)p = 2(W/L)n.$ **GTU : Winter-16, Marks 7**

Solution : Critical values of CMOS inverter are computed as :

$$V_{IL} = \frac{1}{8}\cdot(3\,V_{DD} + 2\,V_{T0,n})$$

$$V_{IL} = 1/8\ (3 \times 1.8 + 2 \times 0.5) = 0.8\ V$$

$$V_{IH} = \frac{1}{8}\cdot(5\,V_{DD} - 2\,V_{T0,n})$$

$$V_{IH} = 1/8\ (5 \times 1.8 - 2 * 0.5) = 1\ V$$

For a symmetric inverter, the sum of V_{IL} and V_{IH} is always equal to V_{DD}.

$$V_{IL} + V_{IH} = V_{DD}$$

$$0.8 + 1 = 1.8\ V$$

Therefore, inverter is symmetric.

3.4.1 Velocity Saturation

- Ideally carrier drift velocity and hence current increases linearly with the lateral electric field (E_{lat}) between source and drain. But in practice drift velocity rolls off due to carrier scattering and eventually saturates of v_{sat}. A plot of carrier velocity versus electric field is shown in Fig. 3.4.1.

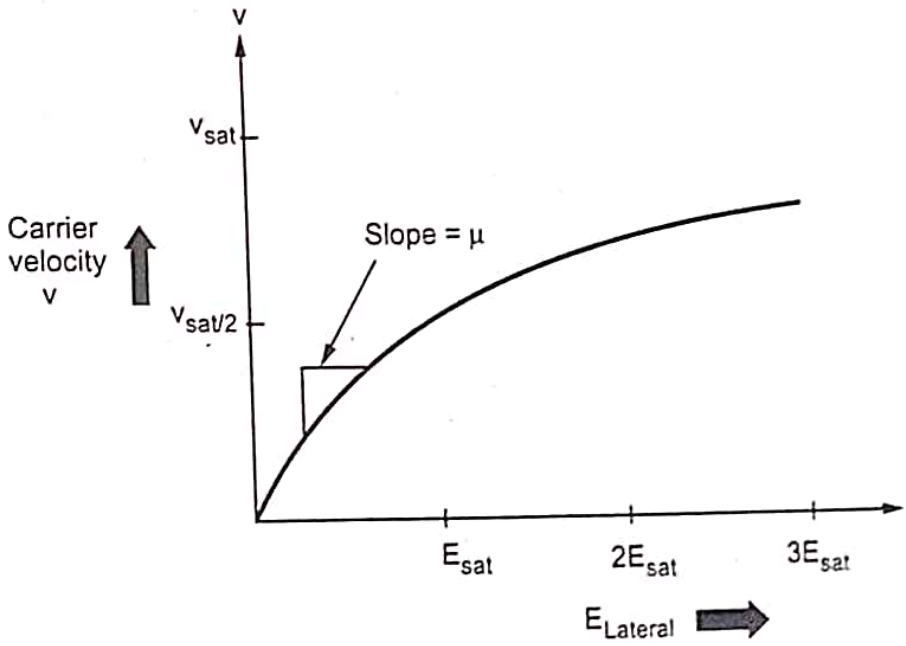


Fig. 3.4.1 Carrier velocity versus electric field

- E_{sat} is determined by expression,

$$v_{sat} = \mu\ E_{sat}$$

For electrons

$$v_{sat} = 6-10\times10^6\ cm/sec$$

For holes $v_{sat} = 4-8\times10^6$ cm / sec

This results in saturation field for nMOS transistors, $E_{sat} = 2 \times 10^4$ V/cm.

- If transistor is completely velocity saturated i.e. $v = v_{sat}$ then saturation curr expression is modified to -

$$I_{DS} = C_0 W (V_{GS} - V_t) v_{sat}$$

Here the drain current is quadratically dependent on voltage without veloc saturation and linearly dependent when fully velocity saturated.

- For moderate supply voltages, transistors operate in region where :
  - Velocity no longer increases linearly with field.
  - Not completely saturated.
- This behavior is approximated by α-power law model. α is called veloc **modulation index** determined by curve fitting measured I-V data. The model based on three parameters : α, β P_c and P_v.

$$I_{DS} = \begin{cases} 0 & V_{GS} < V_t & \text{Cut-off} \\ I_{dsat}\dfrac{V_{DS}}{V_{dsat}} & V_{DS} < V_{dsat} & \text{Linear} \\ I_{dsat} & V_{DS} > V_{dsat} & \text{Saturation} \end{cases}$$

where,

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

$$V_{dsat} = P_v (V_{GS} - V_t)^{\frac{\alpha}{2}}$$

- The low-field mobility of holes is much lower than that of electrons, so PM transistors experience less velocity saturation than NMOS for given V_{DD}. There value of α is larger for PMOS than NMOS transistors.

3.4.2 Mobility Degradation

- Strong vertical electrical fields due to large V_{GS} cause the carriers to scatter reduces the carrier mobility μ. This phenomena is called **mobility degradati** The mobility degradation can be modeled by replacing μ with smaller μ_{eff}.

3.4.3 Channel Length Modulation

- For a transistor in saturation, ideally I_{DS} is independent of V_{DS}, this ma transistor a perfect current source.

- The reverse biased p-n junction between drain and body forms a depletion region with a width L_d that increases with reverse biased voltage. The depletion region effectively shortens the channel length to -

$$L_{eff} = L - L_d$$

- The shorter channel length results in, higher current I_{DS} increases with V_{DS} in saturation region as shown in Fig. 3.4.2.

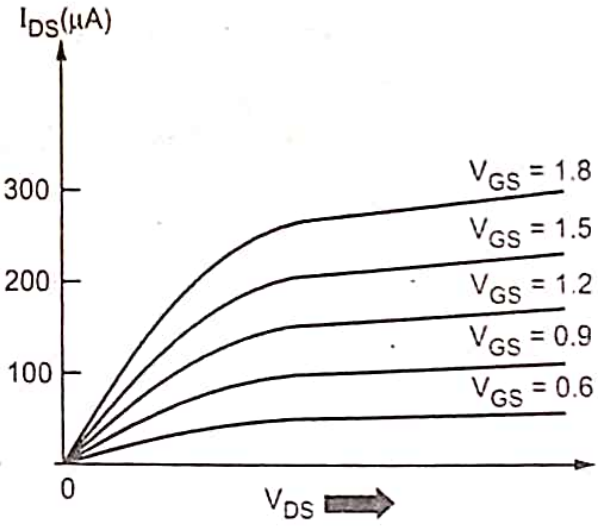


Fig. 3.4.2 I - V characteristics (for NMOS) with channel length modulation

- I_{DS} in saturation region is given by expression :

$$I_{DS} = \beta \frac{(V_{GS} - V_t)^2}{2} \cdot (1 + \lambda V_{DS})$$

where,

λ is channel length modulation factor.

- As the channel length gets shorter, the effect of channel length modulation becomes more significant. The λ is inversely proportional to channel length. The channel length modulation model is useful for conceptual understanding of device.
- The channel length modulation is an important parameter for analog designer as it greatly affects gain of amplifier.

3.4.4 Body Effect / Substrate Bias Effect

- CMOS devices are consists of common substrate. The substrate bias level is normally equal.

Effect of substrate bias :

- Consider two nMOS transistors connected vertically in series as shown in Fig. 3.4.3.
- The source to substrate voltage V_{sb} as observed vertically upward i.e. $V_{sb1} = 0$ and $V_{sb2} \neq 0$. It means that V_{sb2} and V_{sb1} are not same.

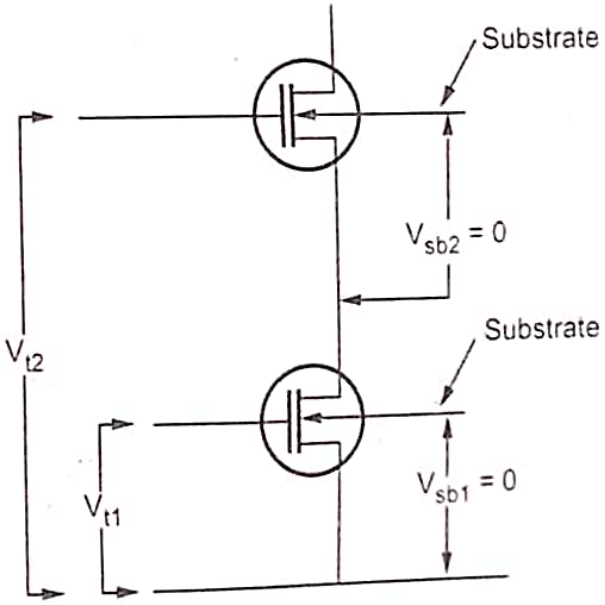


Fig. 3.4.3

This variation of V_{sb} causes variation in threshold voltages i.e. V_{t2} and V_{t1}. Since $V_{sb2} > V_{sb1}$ therefore, $V_{t2} > V_{t1}$. This variation of threshold voltage due to source to substrate voltage is referred as **body effect.** In complex logic circuit body effect is significantly important.

- The body effect (change in threshold voltage) is because of variation in depletion charge region under oxide as a result of source substrate voltage as shown in Fig. 3.4.4.

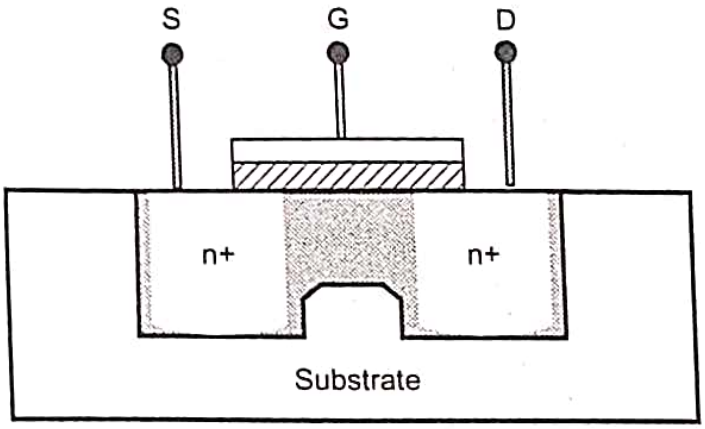


Fig. 3.4.4 N-channel MOSFET

- Change in threshold voltage ΔV_{th} is given by -

$$\Delta V_{th} = \frac{\sqrt{2q\ N_A\ \varepsilon_{si}}}{C_{ox}} \left[\sqrt{2\phi_s + V_{sb}} - \sqrt{2\phi_s}\right]$$

$$\Delta V_{th} = \gamma_n \left[\sqrt{2\phi_s + V_{sb}} - \sqrt{2\phi_s}\right]$$

where,

γ_n is body effect coefficient depends on gate oxide thickness and substrate doping. Typical value for γ_n is in the range of 0.4 to 1.2.

3.4.5 Drain Punch-Through

- When a high voltage is applied to drain, the deplection region around drain may extend to source, causing the current to flow irrespective of gate voltage. This condition is called as **drain punch-through.**

Causes of punch-through

1. Higher voltage to gate causing breakdown.
2. Arcing across the thin oxide.

University Questions

1. *How can we measure following parameters of n-channel MOSFET device ? Explain with necessary test set-ups.*
   1. K_n 2. V_{T0} 3. Substrate bias coefficient (λ) 4. CLM parameter (λ)
   GTU : Winter-14, Marks 7
2. *Is PMOS is subjected to substrate-bias effect (Body effect) in CMOS Inverter ? Justify your answer.* **GTU : Summer-15, Marks 7**
3. *Explain the substrate bias effect in nMOS and pMOS devices.* **GTU : Summer-21, Marks 4**
4. *Explain the experimental methods of measuring following parameters of nMOS : Threshold voltage (V_T), channel length modulation coefficient (λ), substrate-bias coefficient (γ), transconductance parameter (k_n).* **GTU : Summer-21, Marks 7**
5. *Explain Gradual Channel Approximation (GCA), with help of that derive current voltage equations for nMOS transistor.* **GTU : Winter-21, Marks 7**
6. *Explain the substrate bias effect.* **GTU : Winter-21, Marks 3**
7. *Explain the channel length modulation effect in detail.* **GTU : Winter-21, Marks 4**
8. *Describe gradual channel approximation and derive equation for drain current.* **GTU : Summer-21, Marks 7**
9. *Derive the drain current equation for MOSFET using Gradual Channel Approximation (GCA).* **GTU : Winter-21, Marks 7**

3.5 MOSFET Scaling

GTU : Summer-21, Winter-21

- VLSI fabrication processes are being aimed at achieving smaller line widths and size (cheap area) for higher packing density. This scaling-down must also improve circuit performance. Several indicators of micro circuit technology are,
  i) Count of gates on chip ii) Minimum size of device
  iii) Power dissipation iv) Maximum frequency of operation v) Die size
- Scaling improves the figure of merit by shrinking the dimensions of transistors and interconnection between them.

3.5.1 Types of Scaling

- There are three types of scaling -
  1. Constant field (constant E) scaling
  2. Constant voltage (constant V) scaling
  3. Lateral scaling.
- Constant field scaling says that the characteristics of devices can be maintained and the basic operational characteristics preserved if critical parameters of a device are scaled in accordance to a given criteria. The scaled device is obtained by

applying a dimensionless factor x to dimensions, device voltages concentratio densities.

- In constant voltage scaling, voltage V_{DD} is kept constant and the process is scale When device is scaled down by dimension $\frac{1}{x}$ its current density increases by x whereas gate shrink increases the current density by x^2.
- In lateral scaling only gate length is scaled called gate shrink. Lateral scaling easy to implement. The package density increases many fold. Two scaling facto $1/\alpha$ and $1/\beta$ are used in Fig. 3.5.1 in order to cater for all the three scalir models.

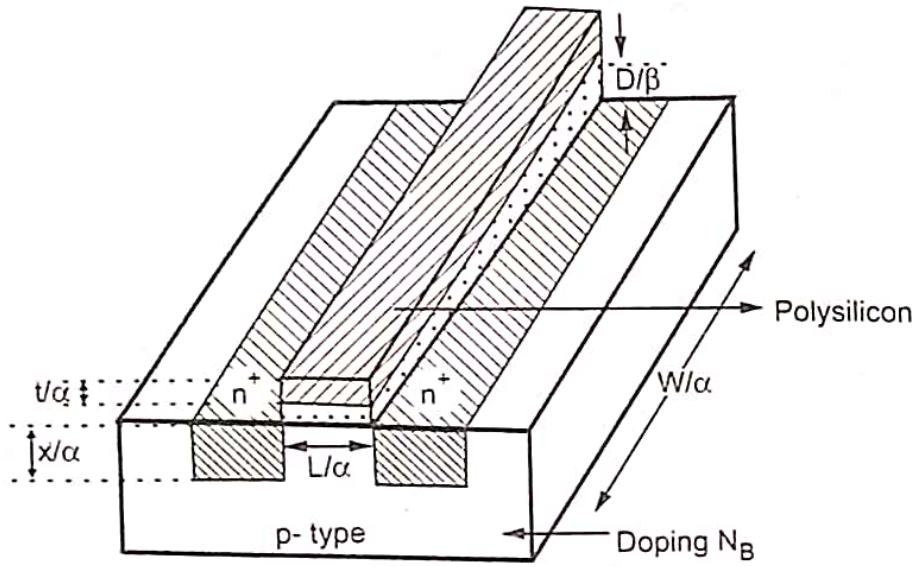


Fig. 3.5.1

3.5.2 Scaling Factors for Device Parameters

1. Gate capacitance per unit area (C_{ox}) :

$$C_{ox} = \frac{\varepsilon_{ox}}{D}$$

where, D is gate oxide thickness.

$$\therefore \quad \text{Scaling factor } (C_{ox}) = \frac{1}{1/\beta} = \beta$$

2. Gate area (A_g) :

$$A_g = WL$$

where,

W is channel width.

L is channel length.

$$\therefore \quad \text{Scaling factor } (A_g) = \frac{1}{\alpha^2}$$

3. Gate capacitance (C_g) :

$$C_g = WL\,C_{ox}$$

$$\therefore \quad \text{Scaling factor } (C_g) = \frac{\beta^2}{\alpha^2}$$

4. Parasitic capacitance (C_x) :

$$C_x \propto \frac{A_x}{d}$$

where,

A_x is area of depletion region.

d is depletion width.

$$\therefore \quad \text{Scaling factor} \quad (C_x) = \frac{1}{\alpha^2} \cdot \frac{1}{1/\alpha} = \frac{1}{\alpha}$$

5. Carrier density in channel (Q_{on}) :

$$Q_{on} = V_{gs}\,C_o$$

Q_{on} is average charge per unit area in channel in ON state.

$$\therefore \quad \text{Scaling factor} \quad (Q_{on}) = \frac{1}{\beta} \cdot \beta = 1$$

6. Channel resistance (R_{on}) :

$$R_{on} = \frac{1}{Q_{on}\,\mu} \frac{L}{W}$$

where,

μ is carrier mobility in channel.

$$\therefore \quad \text{Scaling factor} \quad (R_{on}) = \frac{1}{1} \cdot \frac{1/\alpha}{1/\alpha} = 1$$

7. Maximum frequency of operation (f_o) :

$$f_o = \frac{\mu\,C_o\,V_{DD}}{C_g} \cdot \frac{W}{L}$$

$$\text{Scaling factor for} \quad f_o = \frac{1}{\beta/\alpha^2} = \frac{\alpha^2}{\beta}$$

8. Gate delay (T_d) :

$T_d \propto C_g \; R_{on}$

Scaling factor for $T_d = \beta \cdot \frac{1}{\alpha^2} = \frac{\beta}{\alpha^2}$

9. Saturation current (I_{dss}) :

$$I_{dss} = \frac{W}{L} \cdot \frac{C_{ox}\,\mu}{2} (V_{gs} - V_t)^2$$

$\therefore$ Scaling factor for $I_{dss} = \left(\frac{1}{\beta}\right)^2 \cdot \beta = \frac{1}{\beta}$

10. Current density (J) :

$$J = \frac{I_{dss}}{A}$$

Scaling factor for $J = \frac{1/\beta}{1/\alpha^2} = \frac{\alpha^2}{\beta}$

11. Switching energy per gate (E_g) :

$$E_g = (V_{DD})^2 \frac{C_g}{2}$$

Scaling factor for $E_g = \frac{1}{\beta^2} \frac{\beta}{\alpha^2} = \frac{1}{\alpha^2 \beta}$

12. Power dissipation per gate (P_g) :

- Power dissipation can be of two types i.e. static and dynamic. Static po dissipation takes place when device holds a particular state. Dynamic po dissipation takes place when device changes its state.

$$P_g \text{ (static)} = \frac{(V_{DD})^2}{R_{on}}$$

$$P_g \text{ (dynamic)} = E_g \cdot f_o$$

Both power dissipation are scaled by $1/\beta^2$.

$\therefore$ Scaling factor for P_g is $1/\beta^2$.

13. Power dissipation per unit area (P_a) :

$$P_a = \frac{P_g}{A_g}$$

$\therefore$ Scaling factor for $P_a = \dfrac{1/\beta^2}{1/\alpha^2} = \dfrac{\alpha^2}{\beta^2}$

14. Speed power product (P_T) :

$$P_T = T_d \; P_g$$

$\therefore$ Scaling factor for P_T is $\dfrac{\beta}{\alpha^2} \cdot \dfrac{1}{\beta^2} = \dfrac{1}{\alpha^2 \beta}$

- Scaling factor for different devices are summarized here.

Parameter	Constant field (E)	Constant voltage (V)	Lateral scaling
Supply voltage V_{DD}	$\frac{1}{\alpha}$	1	$\frac{1}{\beta}$
Channel width W	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$
Channel length L	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$
Gate oxide thickness D	$\frac{1}{\alpha}$	1	$\frac{1}{\beta}$
Gate area A_g	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha^2}$
Parasitic capacitance C_x	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$
Gate capacitance C_g	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$\frac{\beta}{\alpha^2}$
Gate capacitance per unit area C_{ox}	α	1	β
Saturation current I_{dss}	$\frac{1}{\alpha}$	1	$\frac{1}{\beta}$
Conductor cross-section A_c	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha^2}$
Carrier density Q_{on}	1	1	1
Channel resistance R_{on}	1	1	1
Current density J	α	α^2	$\frac{\alpha^2}{\beta}$
Switching energy E_g	$\frac{1}{\alpha^3}$	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha^2} \times \beta$
Logic 1 level V_g	$\frac{1}{\alpha}$	1	$\frac{1}{\beta}$

Power dissipation per gate P_g	$\frac{1}{\alpha^2}$	1	$\frac{1}{\beta^2}$
Power dissipation per unit area P_a	1	α^2	$\frac{\alpha^2}{\beta^2}$
Gates per unit area N	α^2	α^2	α^2
Power speed product P_T	$\frac{1}{\alpha^3}$	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha^2\beta}$
Gate delay T_d	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$\frac{\beta}{\alpha^2}$
Maximum operating frequency f_o	α	α^2	$\frac{\alpha^2}{\beta}$

3.5.3 Effects of Scaling on Threshold Voltage (V_t)

- Because of voltage scaling, the power dissipation of transistor is reduced b factor of $1/s^2$. But there is limitation for reduced threshold voltage to a minir of 0.2 V at operating temperature.
- The off state leakage current increases by about ten times for every 0 reduction of V_t. For a chip with an integration level of 100 million transistors average leakage current of turned off device should not exceed a few 10^{-8} A. This limits the threshold voltage to 0.2 V.
- The saturation of threshold voltage leads to the saturation of power supp well, at the minimum value of 0.1 V.

3.5.4 Merits of Scaling

1. Increased chip density i.e. more number of gate counts on single chip.
2. Improved chip performance i.e. increased speed and reduced consumption.
3. Improved device characteristics.
4. Reduced parasitic capacitance.
5. Reduced interconnect delays between devices.
6. Chip cost reduces considerably.

3.5.5 Demerits of Scaling

1. Eventhough overall power consumption reduces but power consumpt unit area increases due to scaling. Hence device gets heated up du operation.

2. Because of scaling down, carrier mobility reduces which inturn reduces gain of device.
3. Reduced conductor size decreases current carrying capacity.
4. High package density increases heat generated which is to be dissipated by forced cooling.
5. Poor I-V characteristics due to narrow channels.
6. Electromigration.
7. Tunneling current through thin oxides.
8. Interconnect delay.

University Questions

1. *What is scaling ? Describe voltage scaling and derive necessary parameter.*

GTU ; Summer-21, Marks 7

2. *What is the need of scaling ? Mention the merits and demerits of constant field scaling.*

GTU : Winter-21, Marks 3

3.6 MOSFET Capacitances

GTU : Winter-14, Summer-14

- The onchip capacitances are functions of layout geometries and manufacturing process.
- The source and drain also have capacitances. These capacitances does not affect the operation of device but they affects the circuit performance and are called as **parasitic capacitances.**
- The parasitic capacitance arise from reverse bised p-n junction and are called as diffusion capacitance.The size of these junction are dependent on -
  - Area and perimeter of source and drain diffusion.
  - The depth of diffusion.
  - The doping levels.
  - Voltage applied.

Parasitic Capacitance

- The parasitic capacitances associated with a MOSFET is shown in Fig. 3.6.1.

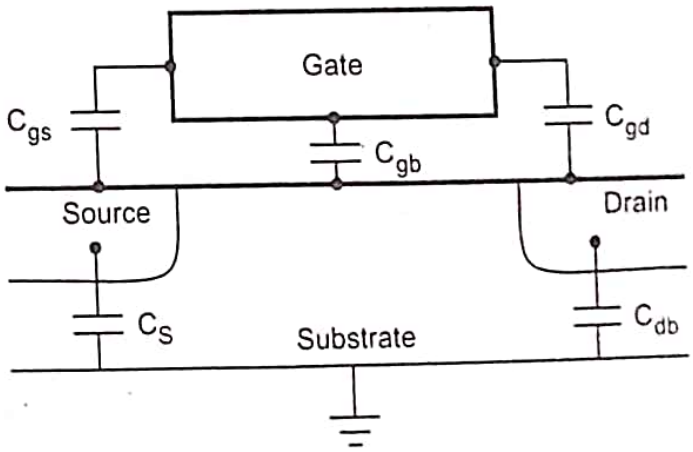


Fig. 3.6.1

- Various parasitic capacitances are -

C_{gs} represents gate to source capacitance.

C_{gb} represents gate to substrate capacitance.

C_{gd} represents gate to drain capacitance.

C_{sb} represents source to substrate capacitance.

C_{db} represents drain to substrate capacitance.

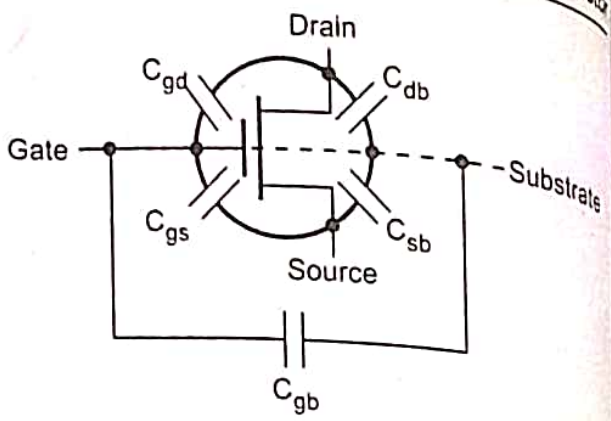


Fig. 3.6.2

Fig. 3.6.2 shows these capacitances in circuit symbol.

- C_{sb} and C_{db} are known as junction or diffusion capacitance arises from the depletion charge between the source and oppositely doped substrate. Similarly between drain and substrate. The capacitance varies with source or drain voltage.
- The total capacitance seen from gate terminal of CMOS transistor is -

$$C_g = C_{gs} + C_{gb} + C_{gd}$$

- The parasitic MOSFET capacitances can be classified into two types :
  1. Oxide related capacitances
  2. Junction capacitances

3.6.1 Oxide Related Capacitances

- The MOS gate sits above the channel overlapping, the source and drain diffusion areas. Hence gate capacitance has two components :
  i) Intrinsic capacitance (over the channel)
  ii) Overlap capacitance (to the source, drain and body).
- The intrinsic capacitance is approximated as simple parallel plate capacitance and overlap (bottom plate) capacitance depends on mode of operation (cut-off, linear and saturation) of transistor.

1. Cut-off

- When transistor is OFF i.e. $V_{GS} = 0$, the charge on the gate is matched with opposite charge from the body (as channel is not inverted), this is called gate-to-body capacitance (C_{gb}). When V_{GS} is below threshold, a depletion region forms at the surface. Because of this bottom plate moves downward from oxide and reduces the capacitance.

2. Linear

- When $V_{GS} > V_t$, the channel inverts and again serves as a good conductive bottom plate. But the channel is connected to the source and drain rather than the

body. For low values of the channel charge is shared between source and drain i.e. $C_{gs} = C_{gd} = \frac{C_0}{2}$.

When V_{GS} is increased, the region near drain becomes less inverted, therefore more fraction of capacitance is attributed to source compared to drain.

turation

When $V_{DS} > (V_{GS} - V_t)$, the transistor saturates and channel pinches off. All capacitance is attributed to source. Due to pinch off, capacitance reduces to $C_{gs} = \frac{2}{3}C_0$.

The intrinsic MOS gate capacitances in these three regions are summarized in Table 3.6.1.

Parameter	Cut-off region	Linear	Saturation
C_{gb}	C_0	0	0
C_{gs}	0	$\frac{C_0}{2}$	$\frac{2}{3}C_0$
C_{gd}	0	$\frac{C_0}{2}$	0
$C_g = C_{gs}+C_{gd}+C_{gb}$	C_0	C_0	$\frac{2}{3}C_0$

Table 3.6.1 Intrinsic MOS gate capacitance

rlap capacitance

The gate overlaps the source and drain partially. This causes added overlap capacitance and is proportional to the width of transistor. Fig. 3.6.3 shows this overlap capacitance.

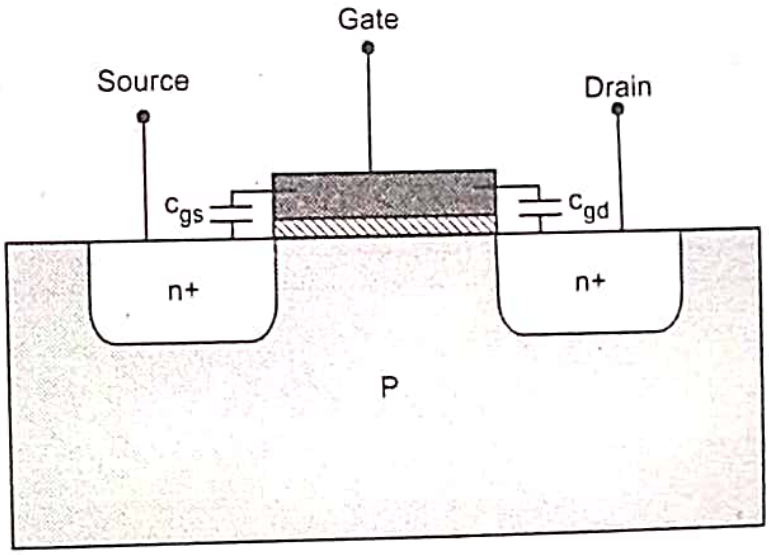


Fig. 3.6.3 Overlap capacitance

The overlap capacitance becomes relatively more important for shorter channel transistor. The effective gate capacitance varies with switching activity of drain and source. The overlap capacitance is dependent on voltage. For delay calculation, the overlap capacitance can be approximated by following expression.

$$C_g = C_{gs} + C_{gd} + C_{gb}$$

3.6.2 Junction Capacitances

The parasitic capacitance exists in the reverse biased p-n junction between source diffusion and the body. The capacitance is dependent on area AS and sidewall perimeter PS of the source diffusion region. Fig. 3.6.4 shows diffusion region geometry.

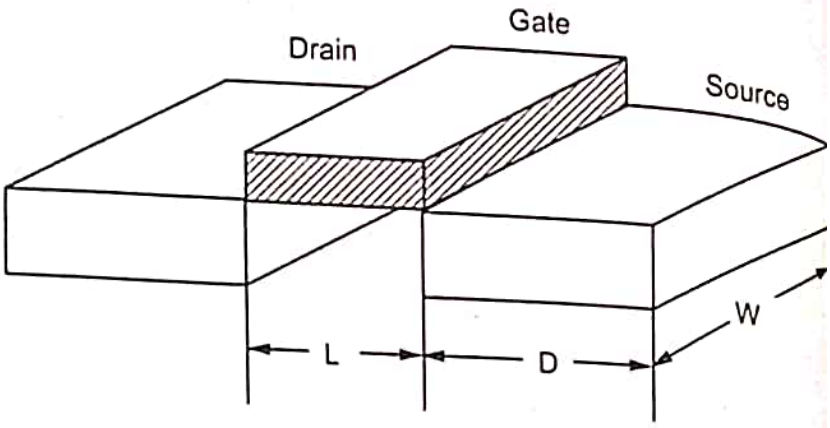


Fig. 3.6.4 Diffusion region geometry

Area $AS = W \cdot D$

Perimeter $PS = 2.W + 2\,D$

- Total source parasitic capacitance is given by -

$$C_{sb} = AS \cdot C_{jbs} + PS \cdot C_{jbssw}$$

where,

C_{jbs} has units of capacitance/area and C_{jbssw} has units of capacitance/length.

- Since the depletion region thickness depends on reverse bias, parasitics nonlinear,

$$C_{jbs} = C_J\left(1+\frac{V_{Sb}}{\psi_0}\right)^{-M_J} \quad ...(3.6$$

where,

C_J is junction capacitance at zero bias.

M_J is junction grading coefficient between (0.5 to 0.33)

ψ_0 is built in potential

$$\psi_0 = v_T \cdot ln\,\frac{N_A N_D}{n_i^2} \quad ...(3.6$$

where,

v_T (thermal voltage) $= \frac{kT}{q}$ (26 mV at room temperature)

N_A, N_D are doping levels of body and source diffusion region.

n_i is intrinsic carrier concentration in undoped silicon.

(usually 1.45×10^{10} /cm^3 at 300° K)

- The sidewall capacitance term is of a similar form but uses different coefficients.

$$C_{jbssw} = C_{Jsw}\left(1+\frac{V_{sb}}{\psi_0}\right)^{-M_{Jsw}} \quad ...(3.6.3)$$

- Considering all the facts on MOS transistor can be viewed as a four-terminal device with capacitances between each terminal pair as shown in Fig. 3.6.5.

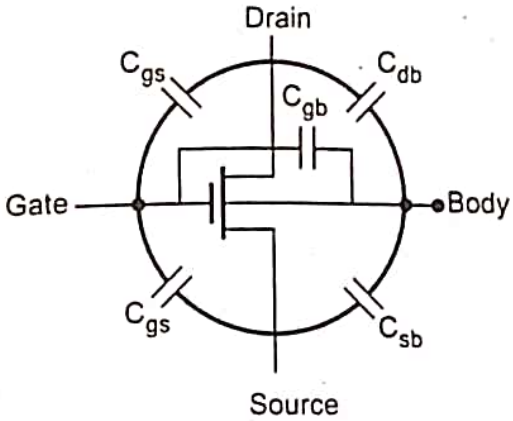


Fig. 3.6.5 Capacitances of MOS transistor

University Questions

1. *Draw the structure of MOS capacitor and derive the expression for the maximum depth of the depletion region for the same.* **GTU : Winter-14, Marks 7**
2. *Discuss MOSFET capacitances in brief.* **GTU : Summer-14, Marks 7**
3. *Explain the MOSFET capacitances in detail.* **GTU : Summer-21, Marks 7**

3.7 Short Questions and Answers

Q.1 Define g_m of MOS transistor.

Ans. : Transconductance :

- Transconductance is an expression of the performance of a Field-Effect Transistor (FET).
- For an FET, transconductance is the ratio of the change in drain current to the change in gate voltage over a defined, arbitrarily small interval on the drain - current - versus - gate - voltage curve.
- The symbol for transconductance is g_m. The unit is the siemens, the same unit that is used for Direct - Current (DC) conductance.
- If δI_{DS} represents a change in collector or drain current caused by a small change in base or gate voltage δV_{GS}, then the transconductance is approximately :

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

- In general, the larger the transconductance figure for a device, the greater gain(amplification) it is capable of delivering, when all other factors are h constant.

Q.2 Define threshold voltage of a MOS device.

Ans. : Threshold voltage :

- Threshold voltage is the voltage applied between gate and source of a MOS that is needed to turn the device on for linear and saturation regions operation.
- Threshold Voltage (V_{TH}) is the value of V_G that will cause the inter potential to be equal in magnitude and opposite in sign to the substrate poten ϕ_p.
- Physically this mean that there would now be a mobile electron concentratio the surface that is equal in magnitude to the mobile hole concentration is p-substrate. When this happens, we say that the surface is INVERTED and electron channel at the surface is called the inversion layer. So define thresh voltage as $V_G = T_{TH}$ when $\phi_s - \phi_p$.

Q.3 List the fabrication procedures for IC technologies.

OR List the basis process for IC fabrication.

Ans. : Various processes in fabrication are :

1. Masking
2. Oxidation
3. Lithography
4. Diffusion
5. Ion implantation
6. Metallization
7. Passivation

Q.4 Define figure of merit of MOS transistor.

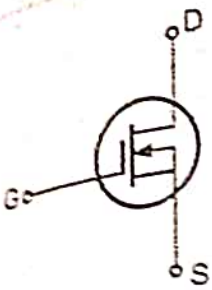
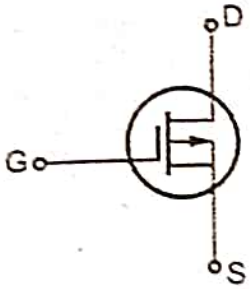
Ans. : Figure of merit

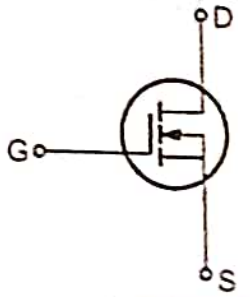
- Figure of merit is a measure for frequency response and switching perform of a MOS transistor.
- Figure of merit is defined as : $\omega = g_m / C_g$

Where, gm is transconductance. C_g is gate capacitance.

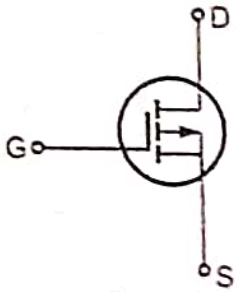
Q.5 Distinguish between enhancement and depletion mode transistor actio NMOS.

Ans. :

DMOSFET	EMOSFET
 n-channel	 p-channel
• An insulting oxide layer (SiO_2) is present between gate and channel.	• The insulting oxide layer is present between gate and substrate.
• n or p type channel is present.	• Channel is not present. At the time of operation, an induced channel gets created.
• (For n channel MOSFET) V_{GS} = –ve (for depletion mode) V_{GS} = +ve (for enhancement mode)	V_{GS} only +ve (for n-channel MOSFET)
• (For n channel MOSFET) If V_{GS} = more –ve $I_D \downarrow\downarrow$	V_{GS} = more +ve $I_D \uparrow\uparrow$



n-channel



p-channel

Q.5 Define scaling and explain it.

Ans. : Scaling

- The design of high - density chips in MOS VLSI (Very Large Scale Integration) technology requires that the packing density of MOSFETs used in the circuits is as high as possible and consequently, that the sizes of the transistors are as small as possible. The reduction of the size, i.e., the dimensions of MOSFETs, is commonly referred to as scaling.
- Several indicators of micro circuit technology are,
  i) Count of gates on chip
  ii) Minimum size of device
  iii) Power dissipation
  iv) Maximum frequency of operation
  v) Die size
- Scaling improves the figure of merit by shrinking the dimensions of transistors and interconnection between them.

Q.7 What is meant parasitic capacitance ?

Ans. : Parasitic capacitance

- The parasitic capacitance arise from reverse biased p-n junction and are called as diffusion capacitance. The size of these junction are dependent on :
- Area and primeter of source and drain diffusion.
- The depth of diffusion
- The doping levels.
- Voltage applied.

Parasitic capacitance

- The parasitic capacitance associated with a MOSFET is shown in Fig. 3.7.1.
- Various parastic capacitances are :
- C_{gs} represents gate to source capacitance.
- C_{gb} represents gate to substrate capacitance.
- C_{gd} represents gate to drain capacitance.
- C_{sb} represents source to substrate capacitance.
- C_{db} represents drain to subtrate capacitance.

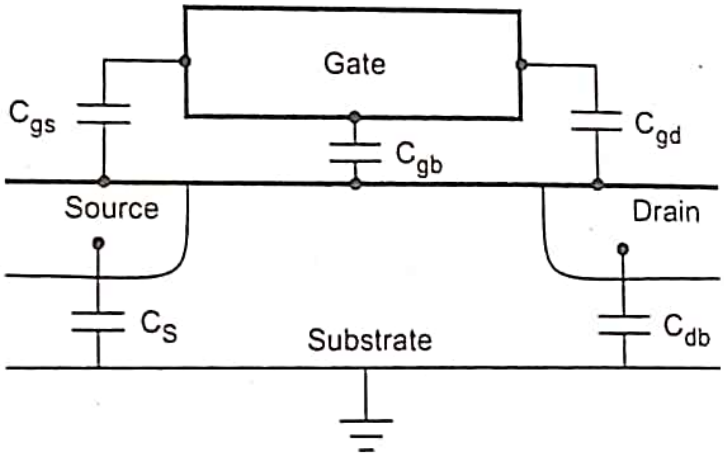


Fig. 3.7.1

4 MOS Inverters : Static Characteristics

Syllabus

Introduction, Resistive load Inverter, Inverter with n-type MOSFET load(Enhancement and Depletion type MOSFET load), CMOS Inverter.

Contents

4.1 Introduction ..... **Summer-21, Winter-21,** ... Marks 4

4.2 Resistive Load Inverter **Summer-15, 16, 21,**
..... **Winter-21,** ..... Marks 7

4.3 CMOS Inverter **Summer-15, Winter-21,**

4.4 CMOS Inverter DC Characteristics..... **Winter-14, 16, 21,**
..... **Summer-15, 21** ..... Marks 7

4.5 Short Questions and Answers

4.1 Introduction

GTU : Summer-21, Winter-21

4.1.1 nMOS Inverter

- The simplest form of an nMOS logic circuit is an inverter as shown in Fig. 4.1.1. It consists of a load resistance R called the pull-up resistor and a pull-down transistor T connected in series between supply voltage V_{DD} and ground.
- The value of the resistor is so decided as to limit the pull-up current to some fraction of maximum pull-down current provided by the transistor. For V_{DD} equal to +5 V, value of R would typically be 40 kΩ which limits the pull-up current to about 125 μA.

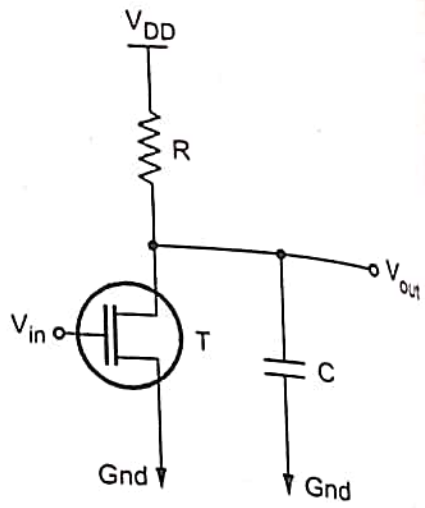


Fig. 4.1.1 The nMOS Inverter

- The input voltage is applied to the gate of transistor T. The transistor offers a ve high input impedance through the gate capacitance C_g. In a typical applicati the output drives a similar load capacitance C.
- If V_{in} is less than V_{th}, the threshold voltage of T, transistor T will be turned o and the load capacitance C will be charged to V_{DD}, which will be the outp voltage V_{out}. However, due to a small leakage current flowing through T, t output voltage V_{out} never equals V_{DD} but is slightly less than V_{th}. If V_{in} increased beyond V_{th}, transistor T will initially conduct in the saturation regi the capacitance C will start discharging, pulling down V_{out} to a lower voltage a the current I_{DS} will increase due to the larger value of V_{GS} which equals V_{in}. T continues until the transistor goes into the linear region of the $I_{DS} - V_{DS}$ cur When V_{in} equals V_{DD}, T will conduct heavily and the output voltage will be giv by,

$$V_{o1} = V_{DD} \frac{R_{ch}}{R + R_{ch}}$$

- Where R_{ch} is the channel resistance with $V_{GS} = V_{DD}$. Since R_{ch} is very sm compared to R, the output voltage will be almost the ground potential. In a typ application, the output of the inverter is used to drive another inverter and must be less than V_{th} to ensure that the pull-down transistor of the driven inve is turned off. Since $V_{th} = 0.2 V_{DD}$, it is necessary that

$$R \geq 4R_{ch}$$

- This means that the output capacitance C discharges through R_{ch} at least four times faster during the pull-down phase than during its pull-up phase through R. This basic asymmetry of switching, times is a fundamental limitation of 'ratioed' logic which this inverter represents.
- To have a higher speed of operation, R must be made small. This not only requires a smaller value of R_{ch} (requiring a larger silicon area for the pull-down transistor) to satisfy the ratio criteria, but also involves more power dissipation in the load resistor $[(V_{DD} - V_{o1})^2 / R$ watts] due to increased current. An effect same as that due to lowering R can be achieved by lowering the value V_{th}, which results in a larger source-to-drain current. This would certainly give higher speed, but to produce a lower value of the output voltage V_{o1}, R_{ch} should be accordingly lower. For example, if $V_{th} = 0.1\,V_{DD}$, the ratio becomes $R \geq 9\,R_{ch}$. Lower channel resistance implies larger silicon area for the pull-down.
- Fabricating a large resistive load on the silicon surface takes up quite a bit of area. A better solution is to use a depletion mode transistor as a pull-up load with source and gate connected together so that $V_{GS} = 0$. Such an arrangement is shown in Fig. 4.1.2.

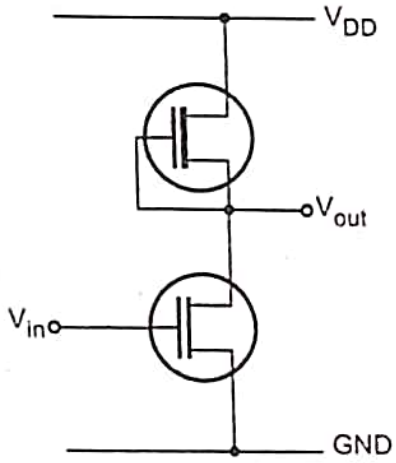
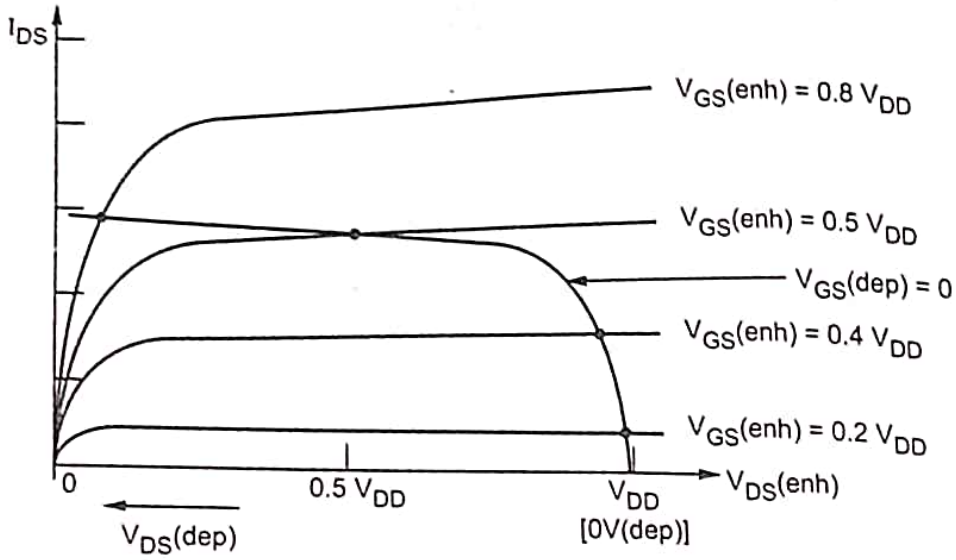


Fig. 4.1.2 nMOS inverter

- With such a configuration, with no current drawn from the output, the currents I_{DS} for both transistors are equal. For the depletion mode transistor T_2, the gate is connected to the source so it is always on and only the characteristic curve $V_{GS} = 0$ is relevant. In such a configuration the depletion mode device is called the pull-up (P.U.) and the enhancement mode device the pull-down (P.D.) transistor.
- The transfer characteristic of such an inverter can be obtained by superimposing the $V_{GS} = 0$ depletion mode characteristic curve on the family of curves for the enhancement mode device with the condition that the maximum voltage across the enhancement mode device corresponds to minimum voltage across the depletion mode device. The points where these curves intersect are the points on the transfer characteristics of the nMOS inverter in Fig. 4.1.2

- The superimposition of the two characteristics is shown in Fig. 4.1.3.



$V_{DS}(enh) = V_{DD} - V_{DS}(dep) = V_{out}$
$V_{GS}(enh) = V_{in}$ . . . intersection points give transfer characteristic

Fig. 4.1.3 Derivation of nMOS inverter transfer characteristic

4.1.2 Transfer Characteristics

- The intersection results into the transfer characteristic as shown in Fig. 4.1.4.

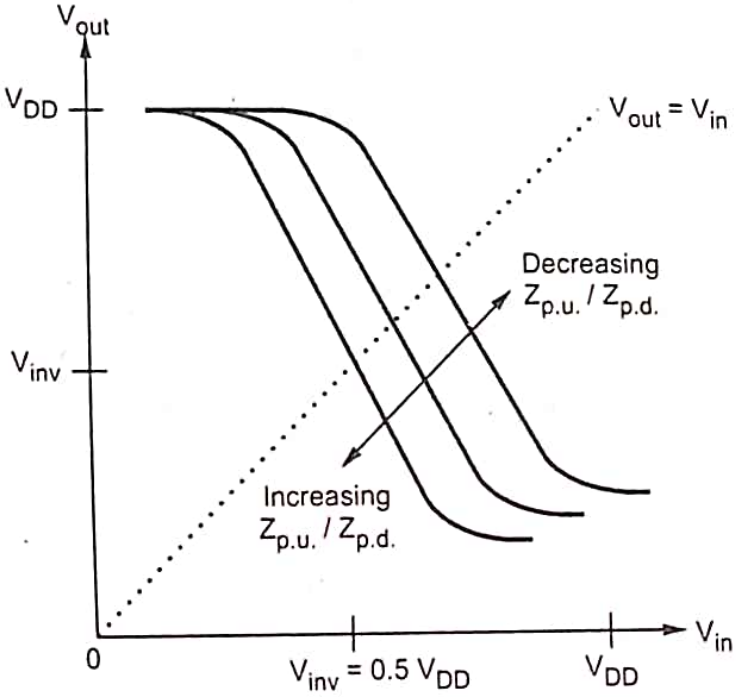


Fig. 4.1.4 nMOS inverter transfer characteristic

- Now the input voltage V_{in} is the gate to source voltage V_{GS} for the pull-d transistor. As this exceeds the threshold voltage of the p.d. transistor, cu begins to flow. This causes the output voltage V_{out} to decrease and fu increase in V_{in} causes the p.d. transistor to come out of saturation and bec resistive. Initially, the pull-up transistor is resistive till the p.d. device turns on

- The gain is defined as the slope of the transfer characteristic during transition.

$$\text{Gain} = \frac{\delta V_{out}}{\delta V_{in}}$$

- The point on the transfer characteristic at which $V_{out} = V_{in}$ is denoted as V_{inv}.

4.1.3 Noise Margin

- **Noise margin** is the amount of noise that a CMOS circuit could withstand without compromising the operation of circuit.
- Noise margin does make sure that any signal which is logic '1' with finite noise added to it, is still recognised as logic '1' and not logic '0'. It is basically the difference between signal value and the noise value.
- Noise margin is a parameter closely related to the input-output voltage characteristics.
- Noise margin or noise immunity is allowable input gate voltage so that output will not be corrupted. Two parameters are used to specify noise margin or noise immunity, these are - LOW noise margin (NM_L) and HIGH noise margin (NM_H).
- The value of LOW noise margin (NM_L) is the difference in maximum LOW input voltage recognized by input gate and maximum LOW output voltage of driver gate.

$$NM_L = V_{IL} - V_{OL} \quad \text{...(4.1.1)}$$

- The HIGH noise margin (NM_H) is difference between minimum HIGH output voltage of driving gate and minimum HIGH input voltage recognized by the receiving gate.

$$NM_H = V_{OH} - V_{IH} \quad \text{...(4.1.2)}$$

where,

V_{IH} is minimum HIGH input voltage

V_{IL} is maximum LOW input voltage

V_{OH} is minimum HIGH output voltage

V_{OH} is maximum LOW output voltage.

- Noise margin definitions are shown in Fig. 4.1.5.
- The input range between V_{IL} and V_{IH} are called as intermediate region or forbidden zone. It do not represent any valid logic levels. It is desirable to have V_{IH} and V_{IL} should be close to each other i.e. to switch transfer characteristics abruptly.

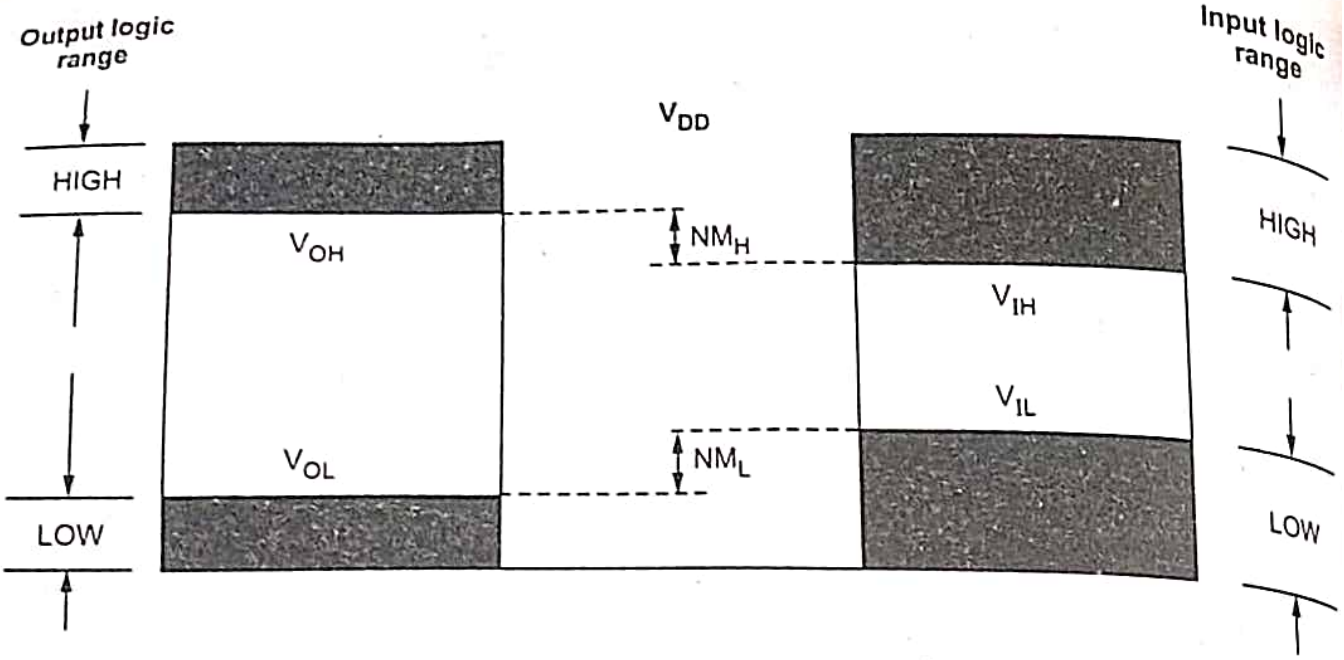


Fig. 4.1.5 Noise margins

- The voltage levels on transfer characteristics are shown in Fig. 4.1.6. The logi levels are at unity gain point (slope = – 1).

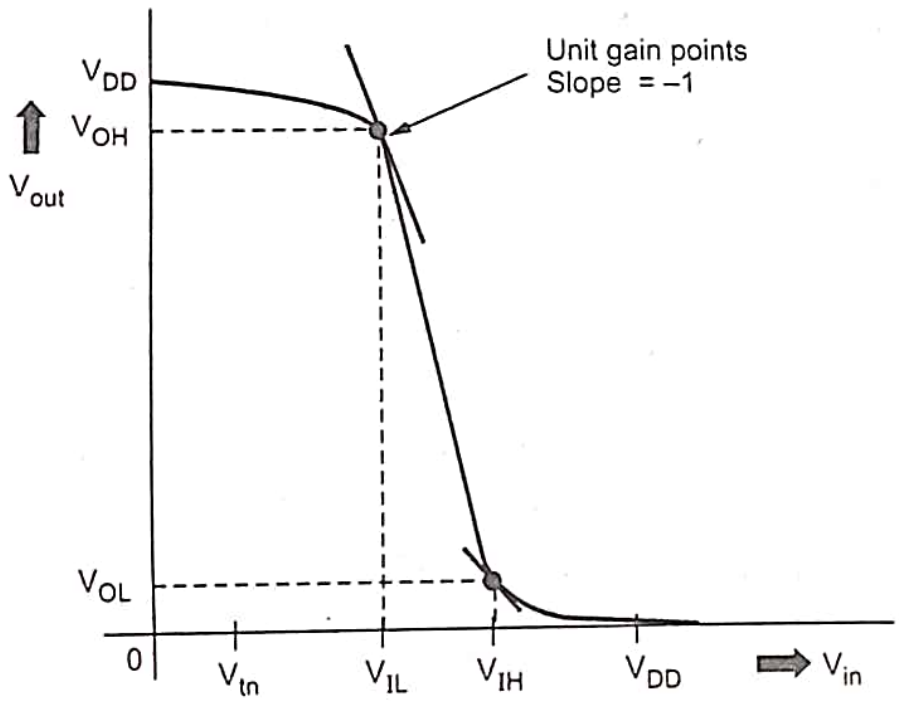


Fig. 4.1.6 CMOS inverter noise margin

- Typical values of noise margins are,

$$NM_L = 0.46\ V_{DD}$$

$$NM_H = 0.34\ V_{DD}$$

- When input is at its worst legal value, the output is slightly degraded, thi called as **noise feed through** or **propagated noise.**

4.1.3.1 Rise Time

- The delay associated with the CMOS inverter can be more precisely estimated by splitting the output transitions into full time τ_f and rise-time τ_r corresponding to the charging and discharging of the capacitive load C_L. These two phenomenon can be independently estimated more precisely to arrive at the overall delay.

Estimation of Rise-time :

- The pull-up p-device drives the capacitive load and can be assumed to be in saturation for the entire charging period of the load capacitor C_L. The equivalent circuit for this condition is shown in Fig. 4.1.7. Now the saturation current for the p-transistor is given by,

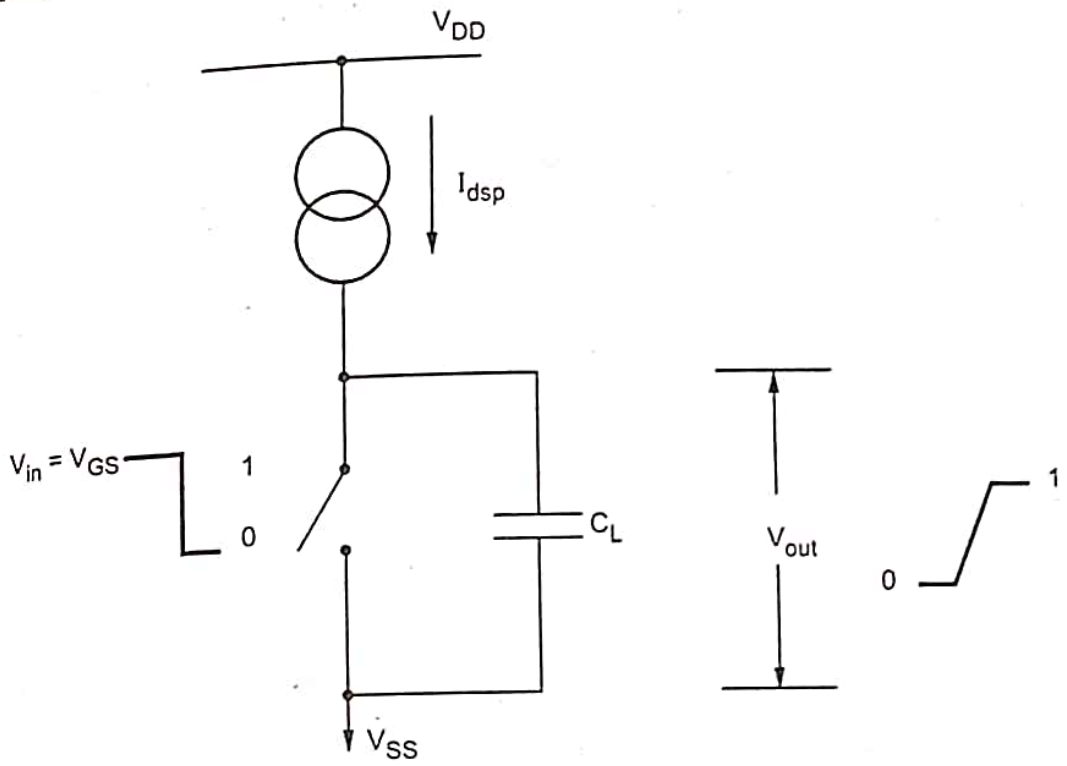


Fig. 4.1.7 Rise-time model

$$I_{dsp} = \frac{(V_{GS} - |V_{tp}|)^2 \beta_p}{2}$$

- This current being constant, output voltage,

$$V_{out} = \frac{t I_{dsp}}{C_L}$$

Putting expression for I_{dsp}, we get,

$$t = \frac{2 V_{out} C_L}{(V_{GS} - |V_{tp}|)^2 \beta_p}$$

- The rise time τ_r corresponds to time taken by V_{out} to reach V_{DD} (approximately), so that,

$$\tau_r = \frac{2C_L V_{DD}}{(V_{DD} - |V_{tp}|)^2 \beta_p}$$

But $|V_{tp}| = 0.2 V_{DD}$ from which,

$$\tau_r \approx \frac{3C_L}{V_{DD} \beta_p} \quad \text{... (4.1.3)}$$

4.1.3.2 Fall-Time

- Fall-time is associated with the discharging of C_L through the pull-down n-type device. The equivalent circuit model for fall-time estimation is shown in Fig. 4.1.8 which shown a constant discharge current.

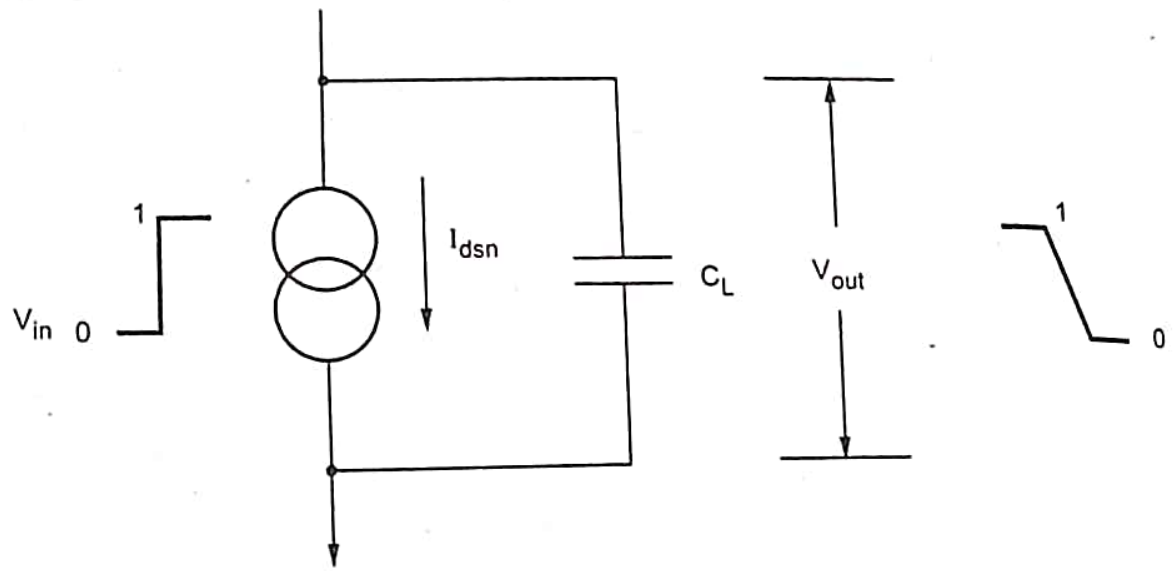


Fig. 4.1.8 Fall-time model

Hence fall-time comes to,

$$\tau_f \approx \frac{3C_L}{V_{DD} \beta_n} \quad \text{... (}$$

From equations (4.1.3) and (4.1.4) and

$$\frac{\tau_r}{\tau_f} \approx \frac{\beta_n}{\beta_p}$$

- However, μ_n and μ_p are not same and $\mu_n = 2.5\mu_p$ because of which $\beta_n \approx 2.5\beta$ shows that rise time is slower by a factor of 2.5 when both the n and p-t are minimum size devices.

- Keeping the channel length minimum, symmetrical operation can be achieved by making $W_p = 2.5 W_n$ where W_p is channel width of p-device and W_n that of n-device. However, with the other geometries as per the minimum size lambda-based rules, this would result in the inverter having an input capacitance of 2.5 $\square C_g$ for p-device plus 1 $\square C_g$ for n-device giving a total capacitance of 3.5 $\square C_g$.
- The analytical models used above for the estimation of rise and fall-times are adequate enough to get optimistic results. However, they do not consider certain factors affecting the rise and fall-times such as,
  i) τ_f and τ_r are proportional to C_L.
  ii) τ_f and τ_r are proportional to $1 / V_{DD}$.
  iii) For equal n and p-transistor geometries, $\tau_r = 2.5 \tau_f$.
- Whenever output signals are required to propagate from the chip to off-chip destinations, the output experiences a comparatively large capacitive load. Such off-chip capacitances may be several orders higher than on chip $\square C_g$ values and typically the off-chip load $C_L \geq 10^4$ $\square C_g$. Such capacitances must clearly be driven through low resistances to avoid excessively long delays.

University Questions

1. *Explain the effect of noise on the performance of a digital system and define noise marigns (NM_L and NM_H)* **GTU : Summer-21, Marks 3**
2. *Draw and explain the inverter circuit of depletion type nMOS load.* **GTU : Winter-21, Marks 4**
3. *Sketch resistive load NMOS inverter circuit and its VTC. Derive equation of V_{OL}.* **GTU : Summer-21, Marks 4**
4. *Define : τ_{PHL}, τ_{PLH}, τ_{rise} with necessary diagram.* **GTU : Summer-21, Marks 3**
5. *Draw voltage transfer characteristics of inverter and define V_{IL}, V_{IH}, V_{OL}, V_{OH}, NM_L and NM_H.* **GTU : Winter-21, Marks 4**
6. *Draw CMOS inverter with leads name of pMOS and nMOS transistors. Derive V_{IL} critical voltage equation of CMOS inverter.* **GTU : Winter-21, Marks 4**

4.2 Resistive Load Inverter

GTU : Summer-15, 16, 21, Winter-21

- Fig. 4.2.1 shows resistive load inverter.
- Since V_{DS} (V_{out}) is larger than ($V_{in} - V_{T0}$); the driver MOSFET is in saturation.
- The current flowing through R is given by :

$$I_R = \frac{K_n}{2}(V_{in} - V_{T0})^2$$

- For larger input voltages, the transistor remains in linear mode. The output voltage is given by :

$$I_R = \frac{K_n}{2}[2(V_{in} - V_{T0})V_{out} - V_{out}^2]$$

- Input voltages and corresponding operating mode is summarized in Table 4.2.1.

Input voltage	Operating mode
$V_{in} < V_{T0}$	Cut-off
$V_{T0} \le V_{in} < V_{out} + V_{T0}$	Saturation
$V_{in} \ge V_{out} + V_{T0}$	Linear

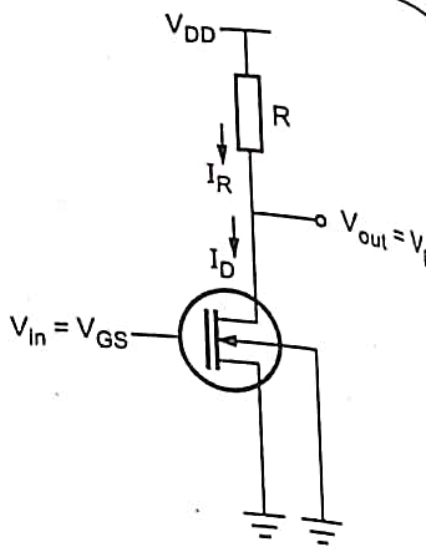


Fig. 4.2.1 Resistive load inve

- The voltage transfer characteristic (VTC) of typical resistive-load inverter ci with different operating mode is shown in Fig. 4.2.2.

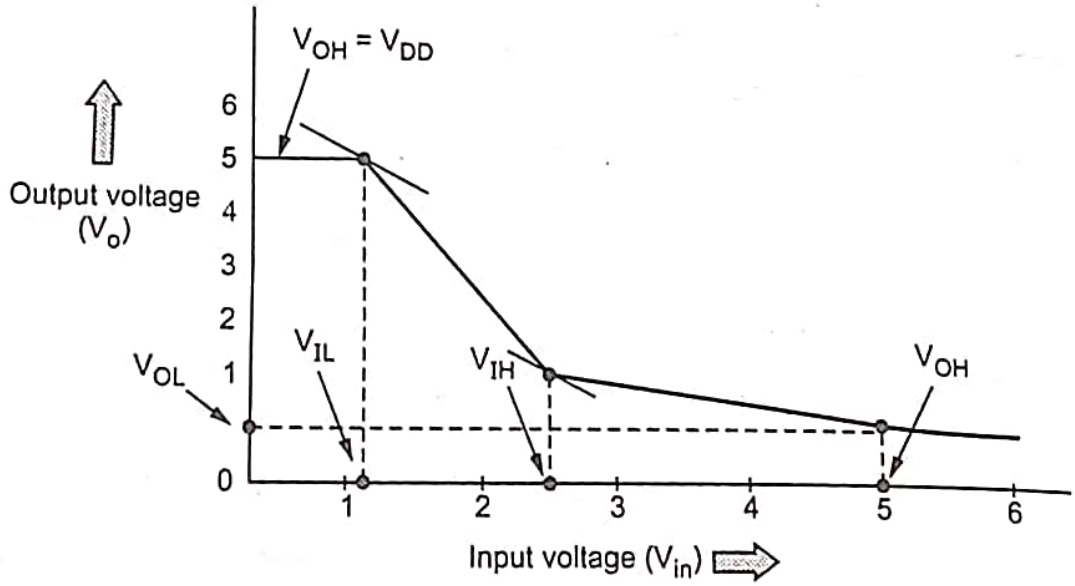


Fig. 4.2.2 VTC of resistive load inverter circuit

Device parameters

$$V_{DD} = 5\text{ V};\ V_{T0} = 1\text{ V}$$

$$K_n = 40 \times 10^{-5}\ \text{A/V}^2$$

$$R_L = 10\ \text{k}\Omega$$

Calculation of V_{OH}

- Output voltage $V_{out} = V_{DD} - R_L I_R$
- When input is low (smaller than threshold), MOSFET is cutoff therefore I_R =

- The output voltage of the inverter is :

$$V_{OH} = V_{DD}$$

Calculation of V_{OL}

- The output low voltage is :

$$V_{OL} = V_{DD} - V_{TU} + \frac{1}{K_n R_L} - \sqrt{\left(V_{DD} - V_{T0} + \frac{1}{K_n R_L}\right)^2 - \frac{2V_{DD}}{K_n R_L}}$$

Calculation of V_{IL}

- V_{IL} is the smaller voltage of the two input voltages at which the slope of VTC becomes $-1 \left(\frac{V_{out}}{dV_{in}} = -1\right)$

$$V_{IL} = V_{T0} + \frac{1}{K_n R_L}$$

Calculation of V_{IH}

- V_{IH} is the larger voltage of the two voltage points on VTC at which the slope is equal to – 1.

$$V_{IH} = V_{T0} + \sqrt{\frac{8}{3} \cdot \frac{V_{DD}}{K_n R_L} - \frac{1}{K_n R_L}}$$

Solved Examples

Example 4.2.1 *Consider a resistive-load inverter circuit with V_{DD} = 5 V, K_n = 10 μA/V², V_{TO} = 0.8 V, R_L = 400 kΩ, and W/L = 2. Calculate the critical voltages (V_{OH}, V_{OL}, V_{IL} and V_{IH}) on the VTC and find the noise margins of the circuit.*

GTU : Summer-15, Marks 7

Solution :

V_{OH} : When input is low; the driver nMOS is cutoff. The output voltage is -

$$V_{OH} = V_{DD} = 5 \text{ V}$$ ...Ans.

V_{OL} : For resistive load inverter, the transconductance of the driver transistor is

$$K_n = K'_n\left(\frac{W}{L}\right) = 10\ (2) = 20\ \mu A/V^2$$

$$K_n \cdot R_L = 20\ \mu A/V^2 \times 400\ k\Omega = 8\ V^{-1}$$

$$V_{OL} = V_{DD} - V_{TU} + \frac{1}{K_n R_L} - \sqrt{\left(V_{DD} - V_{TO} + \frac{1}{K_n R_L}\right)^2 - \frac{2V_{DD}}{K_n R_L}}$$

$$V_{OL} = 5 - 0.8 + \frac{1}{8} - \sqrt{\left(5 - 0.8 + \frac{1}{8}\right)^2 - 2\left(\frac{5}{8}\right)}$$

$$V_{OL} = 0.147 \text{ V}$$

V_{IL} :

$$V_{IL} = V_{T0} + \frac{1}{K_n R_L}$$

$$V_{IL} = 0.8 + \frac{1}{8}$$

$$V_{IL} = 0.925 \text{ V}$$

V_{IH} :

$$V_{IH} = V_{T0} + \sqrt{\frac{8}{3} \cdot \frac{V_{DD}}{K_n R_L}} - \frac{1}{K_n R_L}$$

$$V_{IH} = 0.8 + \sqrt{\frac{8}{3} \cdot \frac{5}{8}} - \frac{1}{8}$$

$$V_{IH} = 1.97 \text{ V}$$

Noise margins :

$$NM_L = V_{IL} - V_{OL} = 0.93 - 0.15 = 0.78 \text{ V}$$

$$NM_H = V_{OH} - V_{IH} = 5.0 - 1.97 = 3.03 \text{ V}$$

Example 4.2.2 *Consider a CMOS inverter with the following parameters :*
$V_{Ton} = 0.6$ V, $V_{Top} = -0.7$ V, $K'_n = 50\ \mu A/V^2$, $K'_p = 16\ \mu A/V^2$, $(W/L)_n = (W/L)_p = 5$. *Calculate the noise margins of this circuit. The power supply voltage* $V_{DD} = 3.3$ V.

GTU : Summer-16, Marks

Solution : $K_R = \frac{K_n}{K_p} = \frac{50}{16} = 3.125$

$V_{OL} = 0$; $V_{OH} = 3.3$ and $V_{Out} = 3.14$ and 0.27

$$V_{IL} = \frac{2V_{Out} + V_{T0,p} - V_{DD} + K_R V_{T0,n}}{1 + K_R} = \frac{2(3.14) + (-0.7) - 3.3 + (3.12)(0.6)}{1 + (3.12)} = 1.007$$

$$V_{IH} = \frac{V_{DD} + V_{T0,p} + K_R(2V_{out} + V_{T0,n})}{1 + K_R}$$

$$V_{IH} = \frac{3.3+(-0.7)+3.12(2\times 0.27+0.6)}{1+3.12} = 1.494 \text{ V}$$

$$NM_L = V_{IL} - V_{OL} = 1.007 - 0 = 1.007 \text{ V} \quad \text{...Ans.}$$

$$NM_H = V_{OH} - V_{IH} = 3.3 - 1.494 = 1.806 \text{ V} \quad \text{...Ans.}$$

Example 4.2.3 *Design a resistive load inverter with $R_L = 1\,k\Omega$, such that $V_{OL} = 0.6$ V. The enhancement type driver transistor has the following parameters :*

$V_{DD} = 5$ V, $V_{TO} = 1$ V, $\gamma = 0.2\,V^{1/2}$, $\lambda = 0$, $\mu_n C_{ox} = 22\,\mu A/V^2$.

Determine i) Require aspect ratio W/L ii) V_{IL} and V_{IH} and iii) noise margin NM_L and NM_H.

GTU : Summer-16, Winter-21, Marks 7

Solution : $V_{OH} = V_{DD}$; $V_{OL} = 0.6$

$$\frac{V_{DD} - V_{OL}}{R_L} = \frac{K'_n}{2}\cdot\frac{W}{L}\left[2(V_{OH} - V_{TO})\cdot V_{OL} - V_{OL}^2\right]$$

$$\frac{5-0.6}{1\times 10^3} = \left(\frac{22\times 10^{-6}}{2}\right)\frac{W}{L}\cdot\frac{W}{L}\left[2(5-1)0.6 - 0.6^2\right] \qquad \because K_n = K'_n\frac{W}{L}$$

$$\left(\frac{W}{L}\right)^2 = 90.09$$

$$\therefore \frac{W}{L} = 9.49$$

University Questions

1. *Draw the inverter circuit with resistive load inverter. Derive critical voltage points V_{OH}, V_{OL}, V_{IL}, V_{IH} for resistive load inverter circuit.* **GTU : Summer-15, Marks 7**
2. *Explain resistive load inverter in brief and derive V_{IL} and V_{IH} critical voltage equations for this inverter.* **GTU : Summer-21, Marks 4**
3. *Draw resistive load inverter. Derive V_{IL} and V_{IH} critical voltage equation of resistive load inverter.* **GTU : Winter-21, Marks 4**

4.3 CMOS Inverter

GTU : Summer-15, Winter-21

- In CMOS, both p and n-channel transistors are used. The circuit is fabricated on an n-type silicon substrate in which a p-type "well" or "tub" is created by diffusion. The n-channel transistors are created in the p-well region. The p-channel devices are made in the n-substrate under an ion-implanted layer called the p^+ layer. Similar to the depletion-mode implantation used for nMOS.
- In this p-well process, special "p-plugs" are used to connect the p-well substrate and the source of the n-channel transistor to ground. Similarly, an "n-plug" connects the n-substrate and the source of the p-type transistor to V_{DD}, the positive supply voltage.
- A schematic circuit representation of the CMOS inverter is shown in Fig. 4.3.1 alongwith its transfer characteristics.

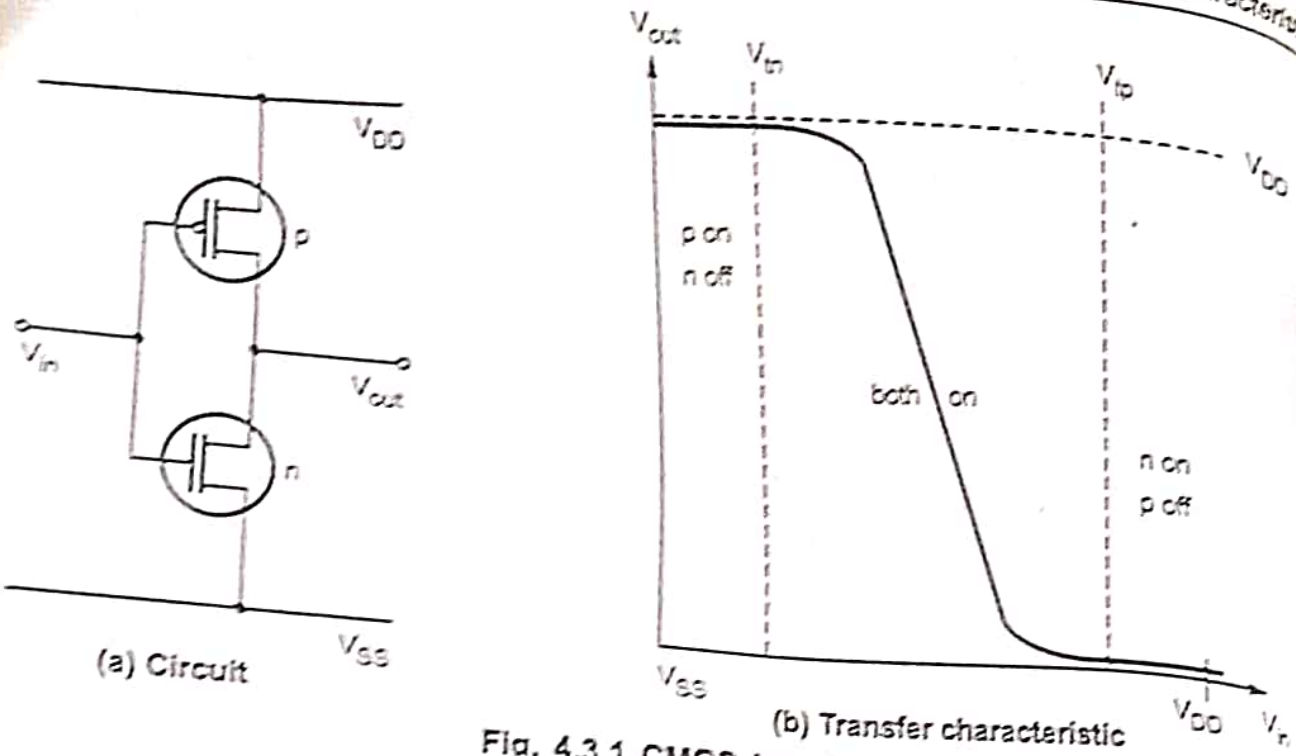


Fig. 4.3.1 CMOS Inverter

- The operation of the circuit on an inverter can be explained as follows. All voltages are referenced with respect to V_{SS}, the ground potential. When the input voltage V_{in} is zero, the gate of the p-channel transistor is at V_{DD} below the source potential, that is, $V_{GS} = V_{DD}$. This turns on this transistor, offering a low-resistance path to load capacitance C which will be charged up to V_{DD}. No current flows through the n-channel transistor, which is turned off since $V_{GS} = 0$ for this transistor. Now if the input voltage is raised to the threshold voltage level of the n-channel transistor and then to V_{DD}, the n-channel transistor will conduct while the p-channel transistor gets turned off, discharging the load capacitance C to ground potential. Note that the current flows until the output node reaches V_{DD} (when charging) or ground (when discharging), the transistors providing either the charge or the discharge current.
- A detailed analysis of the circuit now follows on the basis of its transfer characteristics shown in Fig. 4.3.1 (b). The current/voltage relationships for the MOS transistor is given by the expression for I_{DS}.

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

in the resistive region and for saturation region by,

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

The factor K depends on the geometry of the technology involved since

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

The factor W/L is also attributed to the geometry and it is common practice to write

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

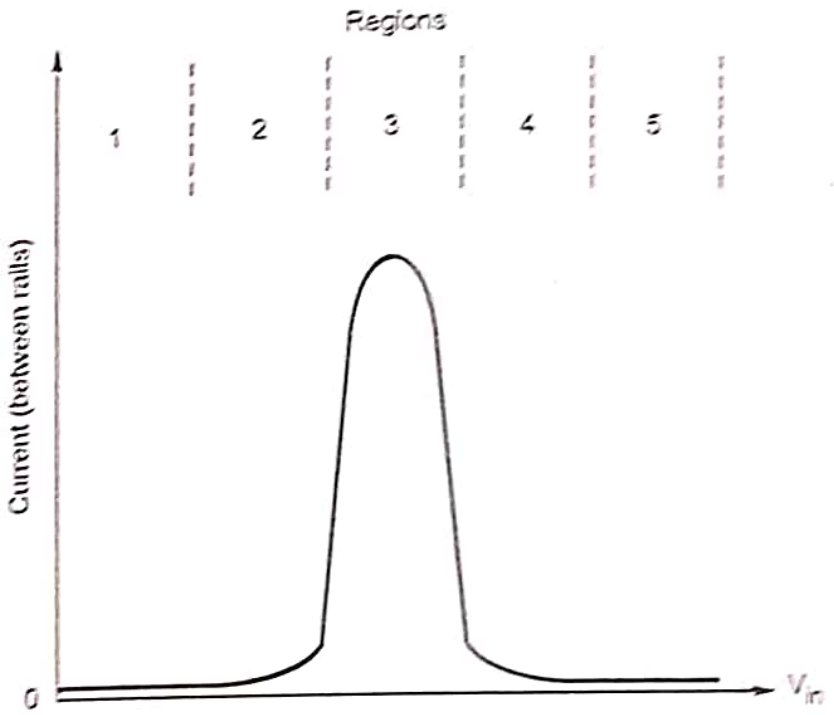
which gives for example,

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

in saturation. The factor β is applicable to both nMOS and pMOS transistors as follows.

$$\beta_p = \frac{\varepsilon_0 \varepsilon_{ins} \mu_p}{D}\frac{W_p}{L_p}$$

$$\beta_n = \frac{\varepsilon_0 \varepsilon_{ins} \mu_n}{D}\frac{W_n}{L_n}$$



(c) CMOS inverter current versus V_{in}

Fig. 4.3.1

here W_p and L_n, W_n and L_n are the p and n-transistor dimensions respectively nd μ_p and μ_n are the hole and electron mobilities respectively. From Fig. 4.3.1 (b) nd (c), we find that the CMOS inverter has five distinct regions of operation.

he region 1 corresponds to operation when V_{in} = logic 0, the p-transistor turned lly on while the n-transistor is fully turned off. Hence no current flows through

the inverter circuit and the output is directly connected to V_{DD} through the pull-up p-transistor. The output has a good logic 1 level. The inverted output corresponds to region 5 when V_{in} = logic 1, the n-transistor turned fully on while the p-device is fully off. In this region, again, no current flows through the circuit. These two regions viz. region 1 and 5 are the static conditions.

- Now consider region 2 of operation. The input voltage has increased to a level that is just above the threshold voltage of the n-transistor. The n-transistor conducts and has a large voltage difference between drain and source and is in saturation. The p-transistor is also conducting, but with only a small voltage difference between its drain and source and hence it operates in the unsaturated resistive region. The inverter circuit draws a small current from the V_{DD} supply.
- In the region 4, conditions are same as those in region 2 but with the roles of the p and n-transistors reversed. That is, p-transistor has a large voltage across it while the n-transistor has a small drop across it. The current drain in both the regions 2 and 4 is small.
- Most of the energy consumed in switching from one state to the other is attributed to the large current flow in region 3. This is the region of operation in which the inverter exhibits gain and in which both the devices are in saturation. Since the two transistors are in series, the current through them is same and we can write,

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

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

is the drain to source current for the p-transistor.

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

is the drain-to-source current for the n-transistor. Writing for V_{in} in terms of the β ratio and other circuit parameters, we get,

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

- In region 3, both transistors are in saturation. Here they act as sources which can be shown on the equivalent circuit as in Fig. 4.3.2.
- The region 3 is very unstable and changeover from one logic level to the other is rapid.

If $V_{in} = -V_{tp}$ and $\beta_p = \beta_n$, equation reduces to

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

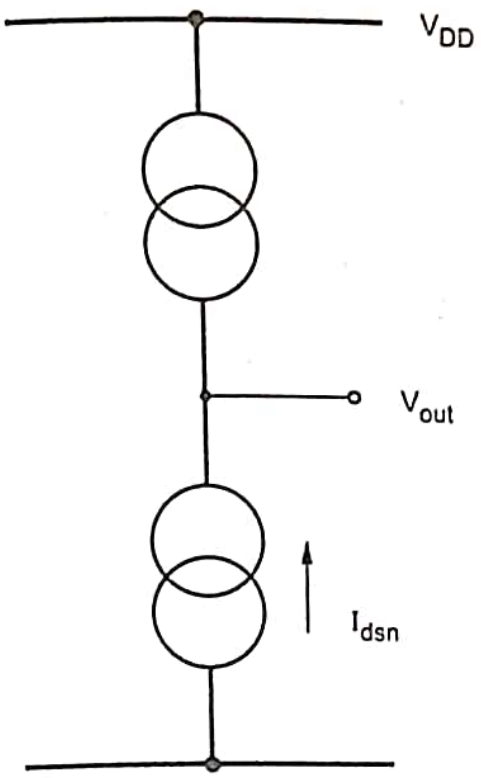


Fig. 4.3.2 Equivalent circuit for region 3

This indicates that the changeover between logic levels is symmetrically set about the point corresponding to $V_{in} = V_{out} = 0.5 V_{DD}$ because only at this point the two β factors will be equal. However, in order that $\beta_n = \beta_p$, the device geometries should satisfy the condition that

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

The mobilities of electrons and holes are inherently unequal.

Thus it is necessary that the width to length ratio (W/L) of the p-device be two to three times that of the n-device i.e.

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

- However mobility μ is not constant and is affected by the transfer electric field in the channel which is a function of V_{GS}. Thus mobility is dependent on V_{in}. The actual mobility is given by the empirical relation.

$$\mu = \frac{\mu_z}{[1 - \phi (V_{GS} - V_t)]}$$

where μ_z is the mobility with zero transverse field, ϕ is a constant approximately equal to 0.5. Thus equal β values exist only around the point of symmetry when $V_{in} = V_{out} = 0.5\ V_{DD}$.

- By keeping minimum size geometries for both p - and n-devices, effect β ratio is minimized. Variation of β causes transfer characteristic of the inverter to change as indicated in Fig. 4.3.3. However, the changes indicated are highly exaggerated and would apply for quite large β ratio variations (e.g. upto 10 : 1).

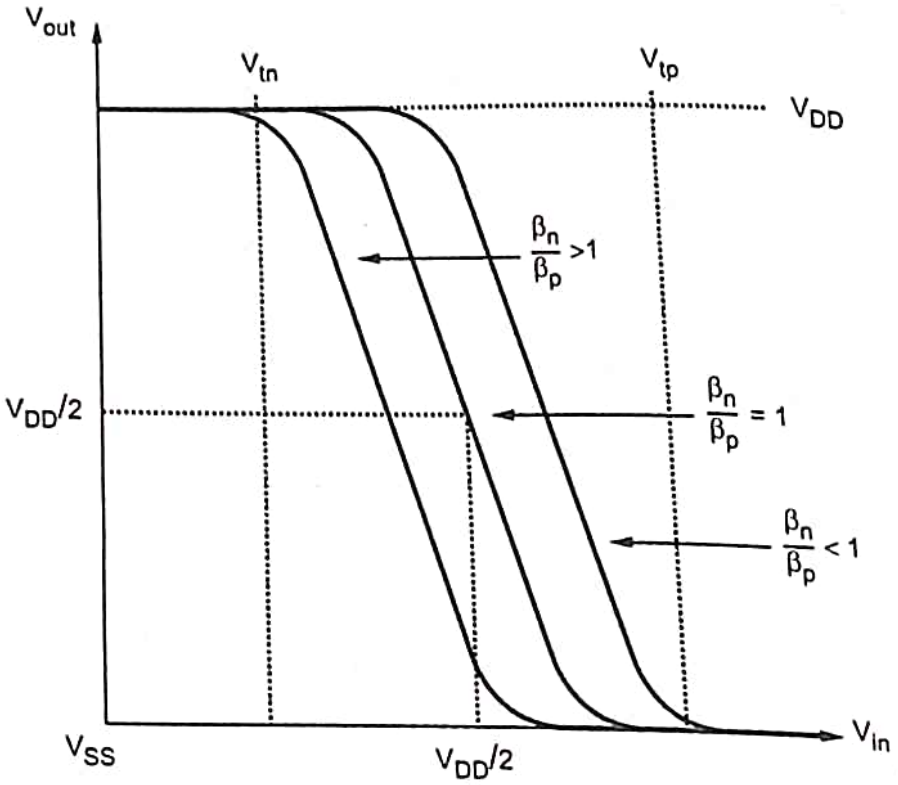


Fig. 4.3.3 Trends in transfer characteristic with β ratio

Mid point voltage V_M on DC characteristic

- The mid point voltage is the point where the Voltage Transfer Characterstic (VT intersects the unity gain line that is defined by,

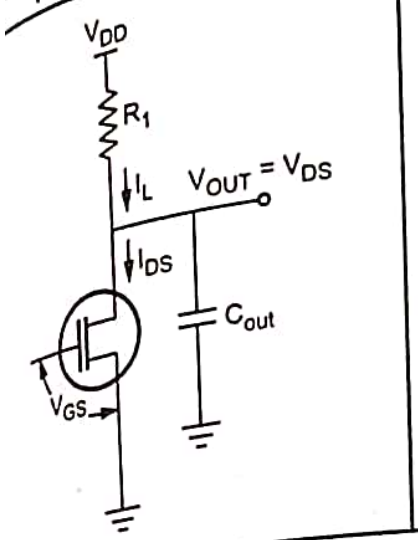
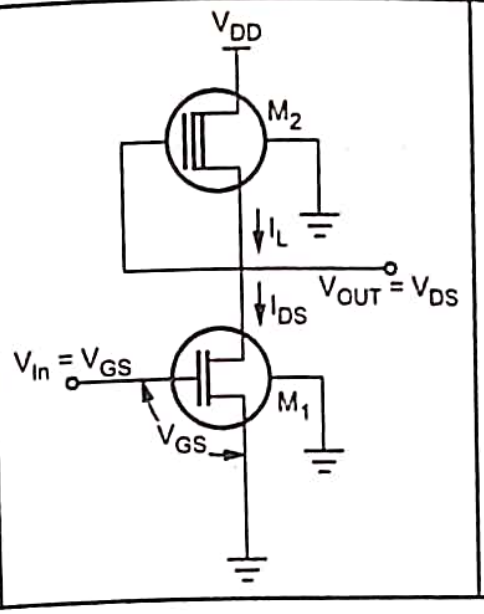
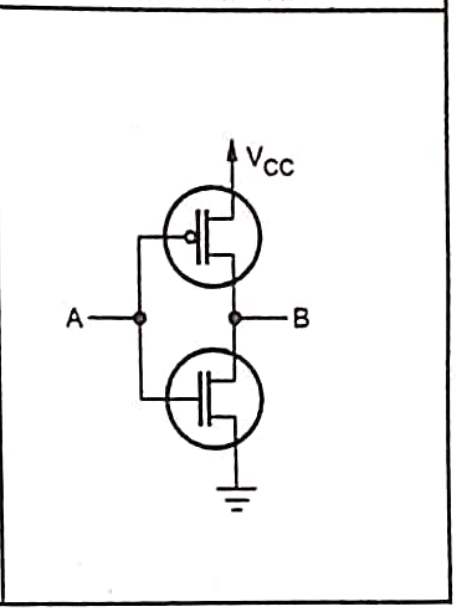
$$V_{out} = V_{in} = V_M$$

- The mid point voltage is given as,

$$V_M = \frac{V_{DD} - |V_{tp}| + \sqrt{\frac{\beta_n}{\beta_p}} \cdot V_{tn}}{1 + \sqrt{\frac{\beta_n}{\beta_p}}}$$

Design

4.1 Comparison of Resistive Load, Depletion Load and CMOS Inverters

Resistive load inverter	Depletion load inverter	CMOS inverter
		

Solved Examples for Understanding

Example 4.3.1 *For a CMOS inverter calculate the shift in transfer characteristic curve when β_n/β_p ratio is varied from $\frac{1}{1}$ to $\frac{10}{1}$.*

Solution : Assuming $V_{DD} = 5$ V

$$V_{th} = -V_{tp}$$

$$V_{tn} = 0.2$$

The shift in transfer characteristics of CMOS inverter when the ratio $\frac{\beta_n}{\beta_p}$ varies from 1 to 10, the gate threshold voltage V_{inv}, where $V_{in} = V_{out}$ is given by expression.

$$V_{in} = \frac{V_{tp} + V_{DD} + V_{tn}\sqrt{\frac{\beta_n}{\beta_p}}}{1 + \sqrt{\frac{\beta_n}{\beta_p}}}$$

When $\sqrt{\frac{\beta_n}{\beta_p}} = 1$; Let the gate threshold voltage be V_{in1}

$$V_{in1} = \frac{V_{tp} + V_{DD} + V_{tn}(1)}{1+1}$$

$$V_{in1} = \frac{V_{tp} + V_{DD} + V_{tn}}{2} \quad \text{but } V_{tn} = -V_{tp}$$

$$V_{in1} = \frac{V_{DD}}{2} = \frac{5}{2} = 2.5 \text{ V}$$

When $\sqrt{\frac{\beta_n}{\beta_p}} = 10$, let the gate threshold voltage be V_{in2}

$$V_{in2} = \frac{V_{tp} + V_{DD} + V_{tn}\sqrt{10}}{1+\sqrt{10}}$$

$$V_{in2} = \frac{-V_{tp} + 5 + V_{tn}\sqrt{10}}{1+\sqrt{10}}$$

$$V_{in2} = \frac{5 - V_{tn} + V_{tn}\sqrt{10}}{1+\sqrt{10}} \qquad \text{but } V_{tn} = 0.2$$

$$V_{in2} = \frac{5 - V_{tn}\,(1-\sqrt{10})}{1+\sqrt{10}}$$

$$V_{in2} = \frac{5-(0.2)(-2.16)}{4.16} = \frac{5+0.432}{4.16}$$

$$V_{in2} = 1.30 \text{ V}$$

Shift in transition region is $V_{in1} - V_{in2} = 2.5 - 1.30 = \mathbf{1.2\ V}$

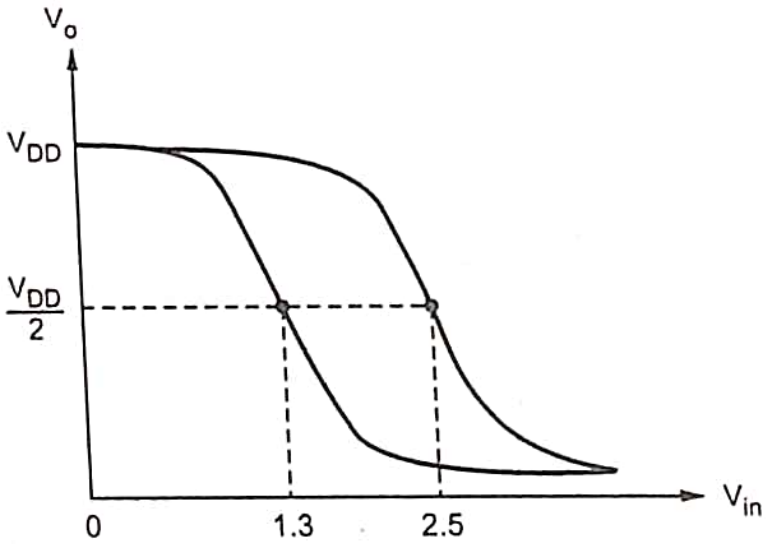


Fig. 4.3.4 Transfer curve

Therefore the transition region shifts from right to left by **1.2 volts**

Design

Example 4.3.2 *Calculate mid-point voltage V_m values on the D.C. characteristics for the following β ratios.*

$V_{tn} = |V_{tp}| = 0.2\ V_{DD}$

i) $\frac{\beta_n}{\beta_p} = 1$ ii) $\frac{\beta_n}{\beta_p} = 10$ iii) $\frac{\beta_n}{\beta_p} = 0.1$

Solution : $V_{tn} = |V_{tp}| = 0.2 \times V_{DD} = 0.2 \times 5\ V = 1.0\ V$

Mid-point voltage is given by,

$$V_M = \frac{V_{DD} - |V_{tp}| + \sqrt{\frac{\beta_n}{\beta_p}} \cdot V_{tn}}{1 + \sqrt{\frac{\beta_n}{\beta_p}}}$$

i) For $\frac{\beta_n}{\beta_p} = 1$

$$V_M = \frac{5 - 1 + \sqrt{1} \cdot 1}{1 + \sqrt{1}}$$

$$V_M = \frac{5}{2} = 2.5\ V$$ ... Ans.

ii) For $\frac{\beta_n}{\beta_p} = 10$

$$V_M = \frac{5 - 1 + \sqrt{10} \times 1}{1 + \sqrt{10}}$$

$$V_M = \frac{5 - 1 + 3.162}{1 + 3.162} = \frac{7.162}{4.162} = 1.72\ V$$ ... Ans.

iii) For $\frac{\beta_n}{\beta_p} = 0.1$

$$V_M = \frac{5 - 1 + \sqrt{0.1} \times 1}{1 + \sqrt{0.1}}$$

$$V_M = \frac{3.683}{1.316} = 2.8\ V$$ ... Ans.

University Questions

1. *Draw voltage transfer characteristics of enhancement mode and depletion mode MOSFETs.* **GTU : Summer-15, Mark 1**
2. *As we reduce V_{DD} of CMOS inverter, write $V_{DD\ (Min)}$ for which CMOS inverter operates correctly.* **GTU : Summer-15, Mark 1**
3. *Why in symmetrical and ideal CMOS inverter required (W/L)p ≈ 2.5 (W/L)n. Justify.* **GTU : Winter-21, Marks 3**

4.4 CMOS Inverter DC Characteristics GTU : Winter-14,16,21, Summer-15,21

- The D.C. transfer-characteristics relate the output voltage to the input voltage, assuming that input voltage changes slowly so that internal capacitance can charge or discharge fully.
- A complementary CMOS inverter is shown in Fig. 4.4.1. Its D.C. transfer function is a curve between V_{out} Vs V_{in}.

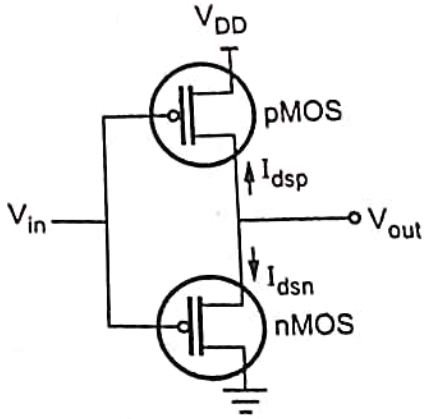


Fig. 4.4.1 Complementary CMOS Inverter

- Let V_{tn} is threshold voltage of n-channel device and V_{tp} is threshold voltage of p-channel device (negative).
- Following Table 4.4.1 summarizes relationship between voltages for the three regions of operation of a CMOS inverter.

	Cut-off	Linear	Saturated
nMOS	$V_{gsn} < V_{tn}$	$V_{gsn} > V_{tn}$	$V_{gsn} > V_{tn}$
	$V_{in} < V_{tn}$	$V_{in} > V_{tn}$	$V_{in} > V_{tn}$
		$V_{dsn} < V_{gsn} - V_{tn}$	$V_{dsn} > V_{gsn} - V_{tn}$
		$V_{out} < V_{in} - V_{tn}$	$V_{out} > V_{in} - V_{tn}$
pMOS	$V_{gsp} > V_{tp}$	$V_{gsp} < V_{tp}$	$V_{gsp} < V_{tp}$
	$V_{in} > V_{tp} + V_{DD}$	$V_{in} < V_{tp} + V_{DD}$	$V_{in} < V_{tp} + V_{DD}$
		$V_{dsp} > V_{gsp} - V_{tp}$	$V_{dsp} < V_{gsp} - V_{tp}$
		$V_{out} > V_{in} - V_{tp}$	$V_{out} < V_{in} - V_{tp}$

Table 4.4.1 Regions of operations of p and n transistors

Design

Since the source of nMOS transistor is grounded i.e. $V_{gsn} = V_{in}$ and $V_{dsn} = V_{out}$. As the source of pMOS transistor is connected to V_{DD} i.e. $V_{gsp} = V_{in} - V_{DD}$ and $V_{dsp} = V_{out} - V_{DD}$.

Transfer characteristics (V_{out} as a function of V_{in}) can be found graphically, analytically or by simulation. Important assumptions are,

$$I_{dsn} = |I_{dsp}|$$

$$V_{tp} = -V_{tn}$$

$$\beta_n = \beta_p$$

Two important equations for MOS transistors are,

$$I_{DS} = \mu C_0 \frac{W}{L}\left(V_{GS} - V_t - \frac{V_{DS}}{2}\right)V_{DS}$$

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

- A plot of I_{dsn} and I_{dsp} interms of V_{dsn} and V_{dsp} for various values of V_{gsn} and V_{gsp} is shown in Fig. 4.4.2.

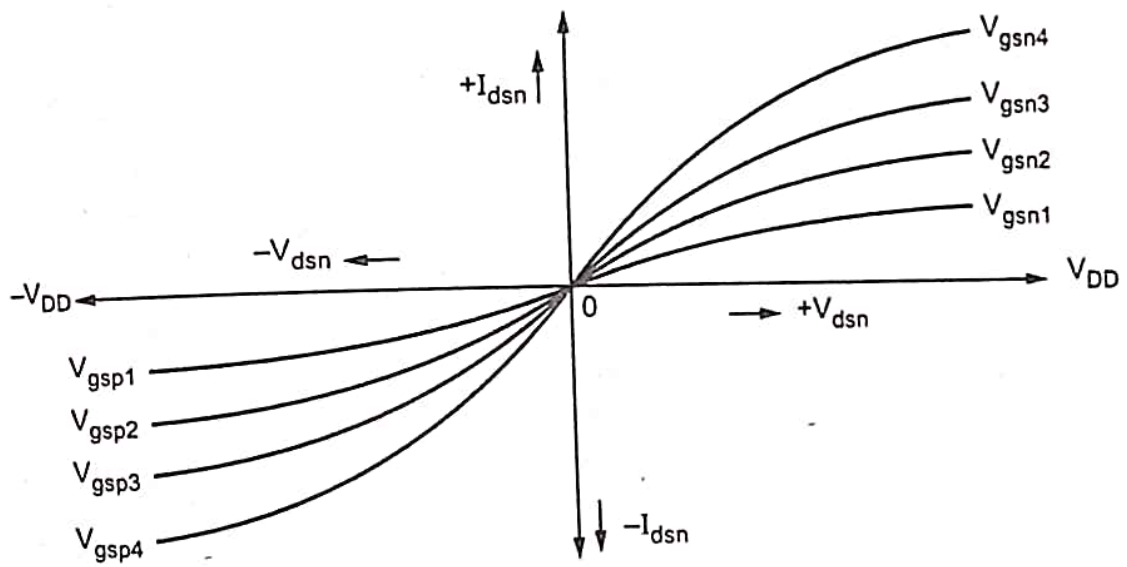


Fig. 4.4.2 Variation of I_{DS} Vs V_{DS}

- A same plot of I_{dsn} and $|I_{dsp}|$ interms of V_{out} for various values of V_{in} is shown in Fig. 4.4.3.

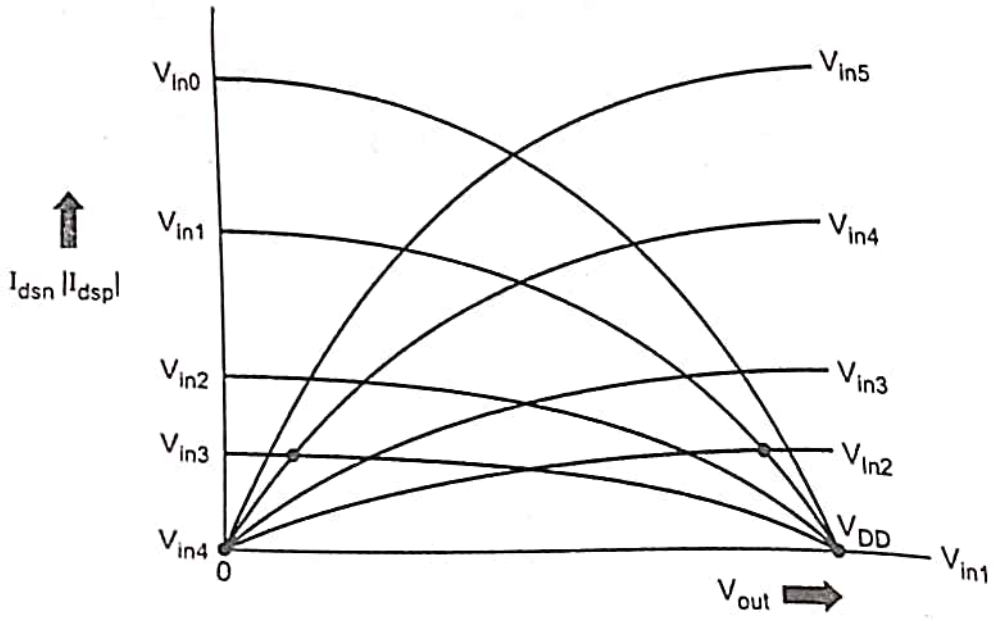


Fig. 4.4.3 I_{dsn} and $|I_{dsp}|$ interms of V_{out}

- The possible operating point are marked with dots. These are values of V_{out} where $I_{dsn} = |I_{dsp}|$ for various values of V_{in}. The inverter D.C. tra characteristics (V_{out} Vs V_{in}) is shown in Fig. 4.4.4 with possible operating poi

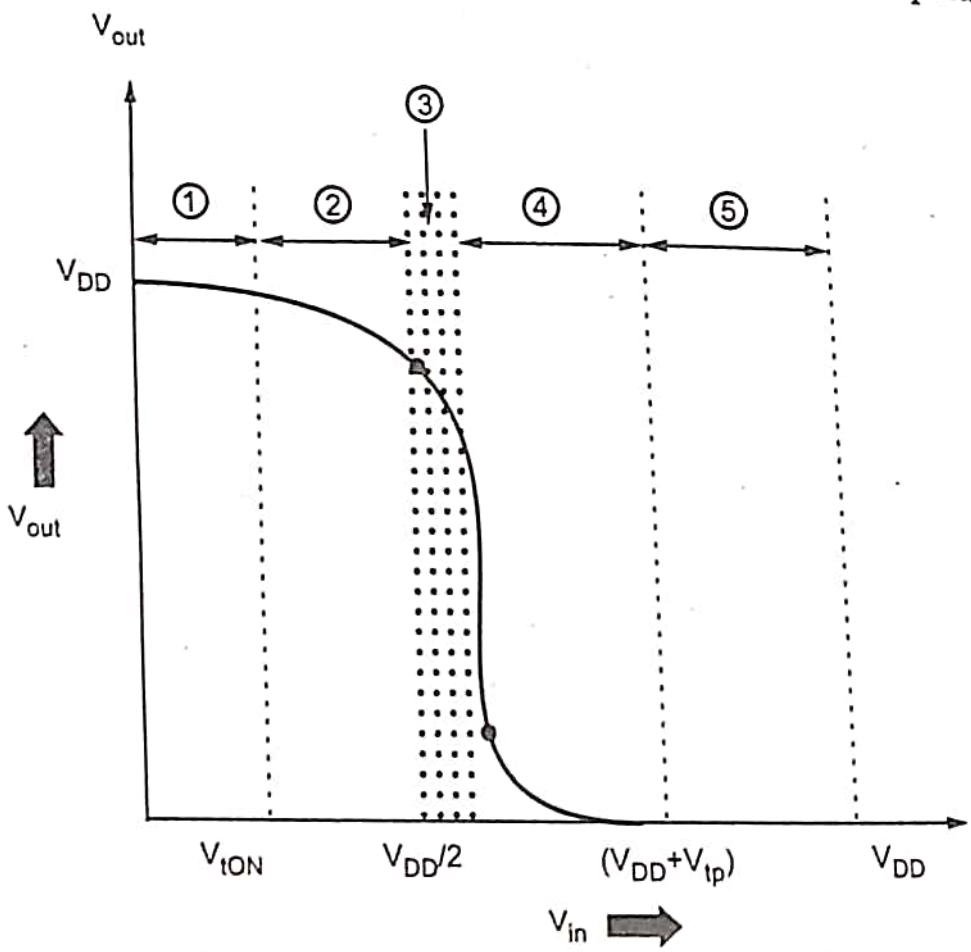


Fig. 4.4.4 (a) CMOS voltage transfer characteristics

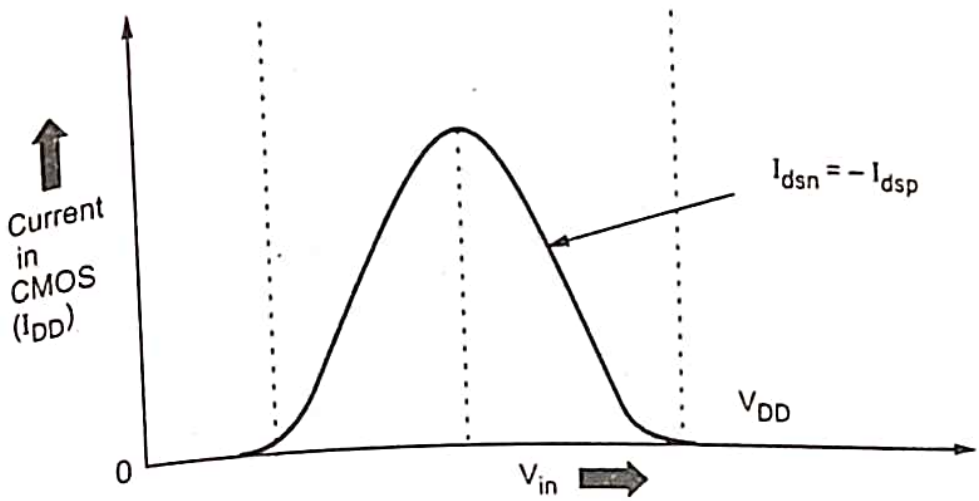


Fig. 4.4.4 (b) CMOS current

4.4.1 Voltage Transfer Curve (VTC)

- The CMOS voltage transfer characteristics is divided into five regions. The switching point is at $V_{DD}/2$.

Region ① : $V_{in} < V_{tON}$

- In this region nMOS transistor is in cut-off ($I_{dsn} = 0$) and pMOS transistor is in linear region. The drain to source current I_{dsp} for p-device is zero.

$$V_{dsp} = V_{out} - V_{DD}$$

The output voltage is given by,

$$V_{out} = V_{DD} \qquad (V_{dsp} = 0)$$

Region ② : $V_{tON} < V_{in} < V_{DD}/2$

- In this region pMOS transistor operates in non-saturated region while nMOS transistor operates in saturation. The output voltage in this region is given by,

$$V_{out} = (V_{in} - V_{tp}) + \sqrt{(V_{in} - V_{tp})^2 - 2\left(V_{in} - \frac{V_{DD}}{2} - V_{tp}\right)V_{DD} - \frac{\beta_n}{\beta_p}(V_{in} - V_{tON})^2}$$

$$V_{out} \approx V_{OH}$$

where,

β_n and β_p are the gain factors for nMOS and pMOS respectively. (β_n and β_p are also known as K_n and K_p).

Region ③ : $V_{in} < V_{DD}/2$

- In this region both pMOS and nMOS transistors are in saturation region. Output voltage is,

$$V_{out} \approx V_{in}$$

- This region defines the gain of CMOS inverter when used as small signal amplifier.

Region ④ : $V_{DD}/2 < V_{in} \leq (V_{DD} + V_{tp})$

- In this region pMOS transistor is in saturation and nMOS transistor is in non-saturation region. Output voltage is given by,

$$V_{out} = (V_{in} - V_{tON}) - \left[(V_{in} - V_{tON})^2 - \frac{\beta_p}{\beta_n}(V_{in} - V_{DD} - V_{tP})^2\right]^{1/2}$$

$$V_{out} = V_{OL}$$

Region ⑤ : $V_{in} \geq (V_{DD} + V_{tp})$

- In this region pMOS transistor is in cut-off region and nMOS transistor is in linear mode. Output voltage is,

$$V_{out} = 0$$

- Table 4.4.2 summarizes operating region of pMOS and nMOS transistors corresponding V_{in} and V_{out}.

Regions	V_{In}	Operating Regions		V_{out}
		pMOS	nMOS	
1.	$< V_{tON}$	Linear	Cut-off	$V_{OH} = \frac{V_{DD}}{2}$
2.	$V_{tON} \leq V_{in} < \frac{V_{DD}}{2}$	Linear	Saturation	$V_{OH} > \frac{V_{DD}}{2}$
3.	$V_{th} = \frac{V_{DD}}{2}$	Saturation	Saturation	0 (Drops sharply)
4.	$\frac{V_{DD}}{2} < V_{in} \leq \frac{V_{DD}}{2} - \lvert V_{tp} \rvert$	Saturation	Linear	$= V_{OL}\left(< \frac{V_{DD}}{2}\right)$
5.	$> \frac{V_{DD}}{2} + V_{tp}$	Cut-off	Linear	0

Table 4.4.2

4.4.2 β_n/β_p Ratio

The effect of $\frac{\beta_n}{\beta_p}$ ratio on transfer characteristics is shown in Fig. 4.4.5.

Design

- For changing $\frac{\beta_n}{\beta_p}$ there is need to change the channel dimensions i.e. channel length L and channel width W. As the ratio $\frac{\beta_n}{\beta_p}$ is decreased, the transition region shifts from left to right but output voltage transition remains sharp.
- For CMOS inverter ratio $\frac{\beta_n}{\beta_p} = 1$ is desirable as it allows a capacitive load to charge and discharge in equal times by providing equal current sources and sink capabilities.

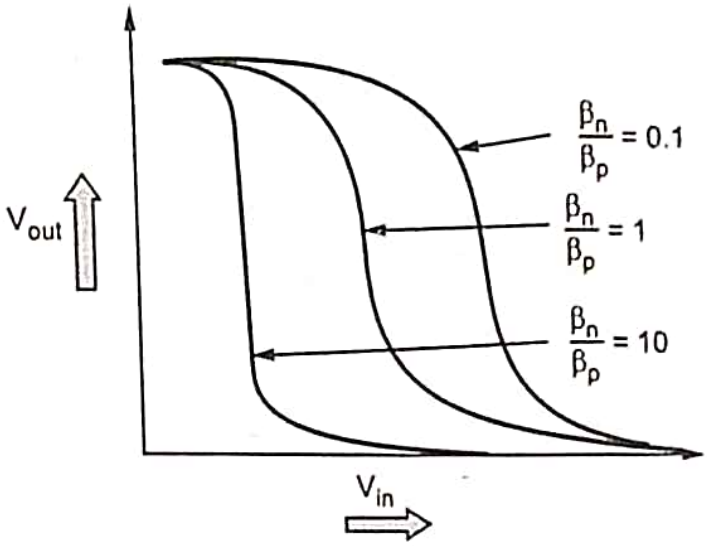


Fig. 4.4.5 Effect of $\frac{\beta_n}{\beta_p}$ on inverter D.C. transfer characteristics

4.4.3 Effects of Temperature on Transfer Curve

- The effective carrier mobility (μ) decreases with increase in temperature of MOS device. This results in decrease in β it is related by,

$$\beta \propto T^{-1.5}$$

hence $$I_{ds} \propto T^{-1.5}$$

- The overall transfer characteristics is shifted to the left with increase in temperature.

4.4.4 Critical Voltage Points

- For this inverter $V_{OH} = V_{DD}$ and $V_{OL} = 0$.
- The critical voltage V_{IL} as a function of output voltage is expressed as -

$$V_{IL} = \frac{2V_{out} + V_{T0,p} - V_{DD} + K_R V_{T0,n}}{1 + K_R}$$

where,

$$K_R = \frac{K_n}{K_p}$$

- The critical voltage V_{IH} as a function of V_{out}

$$V_{IH} = \frac{V_{DD} + V_{T0,p} + K_R(2V_{out} + V_{T0,n})}{1 + K_R}$$

- The inverter threshold voltage V_{th} is found as -

$$V_{th} = \frac{V_{T0,n} + \sqrt{\frac{1}{K_R}}(V_{DD} + V_{T0,p})}{1 + \sqrt{\frac{1}{K_R}}}$$

Solved Examples

Example 4.4.1 *Consider a CMOS inverter circuit with the following parameters : $V_{DD} = 3.3$ V, $V_{T0,n} = 0.6$ V, $V_{T0,p} = -0.7$ V, $K_n = 400$ microAmp/V^2, $K_p = 160$ microAmp/V^2. Calculate the noise margin of the circuit.*

GTU : Summer-15,21, Marks 7

Solution : $K_R = \frac{K_n}{K_p} = \frac{400}{160} = 2.5$ and

$V_{T0,n} \neq |V_{T0,p}|$ hence it is not a symmetrical inverter.

i) $V_{OH} = V_{DD} = V_{out} = 3.3$ V

ii) $V_{OL} = V_{out} = 0$

iii)
$$V_{IL} = \frac{2V_{out} + V_{T0,p} + K_R V_{T0,n}}{1 + K_R}$$
$$= \frac{2V_{out} - 0.7 - 3.3 + (2.5 \times 0.6)}{1 + 2.5}$$
$$= 0.57\ V_{out} - 0.71$$

KCL equation for CMOS inverter :

$$K_n (V_{in} - V_{T0,n})^2 = K_p\ [2(V_{in} - V_{DD} - V_{T0,p}) \cdot (V_{out} - V_{DD}) - (V_{out} - V_{DD})^2]$$

Substituting ($V_{in} = 0.57\ V_{out} - 0.71$) and $K_n / K_p = 2.5$

$$2.5\ (0.57\ V_{out} - 0.71 - 0.6) = [2\ (0.57\ V_{out} - 0.71 - 3.3 + 0.7)\ (V_{out} - 3.3) - (V_{out} - 3.3)^2]$$

Solving the equation

$$0.66\ V_{out}^2 + 0.05\ V_{out} - 6.65 = 0$$

sign

$V_{out} = 3.14$ V (Taking only positive root)

$V_{IL} = 0.57\ (3.14) - 0.71 = \mathbf{1.08\ V}$

$$V_{IH} = \frac{V_{DD} + V_{T0,p} + K_R(2V_{out} + V_{T0,n})}{1 + K_R}$$

$$V_{IH} = \frac{3.3 - 0.7 + 2.5\ (2V_{out} + 0.6)}{1 + 2.5}$$

$V_{IH} = 1.43\ V_{out} + 1.17$

equation :

$(V_{in} - V_{T0,n})\ V_{out} - V_{out}^2] = K_p\ (V_{in} - V_{pp} - V_{T0,\ p})^2$

ituting $V_{in} = 1.43\ V_{out} + 1.17$ in above equation

$(1.43\ V_{out} + 1.17 - 6.6)\ V_{out} - V_{out}^2] = (1.43\ V_{out} - 1.43)^2$

$V_{out}^2 + 6.94\ V_{out} - 2.04 = 0$

$V_{out} = 0.27$ V (Taking only positive root)

$V_{IH} = 1.43\ (0.27) + 1.17$

$V_{IH} = 1.55$ V

se margins :

$NM_L = V_{IL} - V_{OL} = 1.08 - 0 = \mathbf{1.08\ V}$ ...Ans.

$NM_H = V_{OH} - V_{IH} = 3.3 - 1.55 = \mathbf{1.75\ V}$ ...Ans.

mple 4.4.2 *For a CMOS inverter with following specification compute V_{IL}, V_{IH} and find out inverter is symmetric or not ?*

$V_{DD} = 1.8$ V, $V_{tn} = 0.5$ V, $V_{tp} = -0.5$ V, $\mu_n = 580\ cm^2/VS$

$\mu_p = 290\ cm^2/VS$, $(W/L)p = 2(W/L)n$.

GTU : Winter-16, Marks 7

lution : Critical values of CMOS inverter are computed as :

$$V_{IL} = \frac{1}{8}\cdot(3\ V_{DD} + 2\ V_{T0,n})$$

$V_{IL} = 1/8\ (3 \times 1.8 + 2 \times 0.5) = 0.8$ V

$$V_{IH} = \frac{1}{8}\cdot(5\ V_{DD} - 2\ V_{T0,n})$$

$V_{IH} = 1/8\ (5 \times 1.8 - 2 * 0.5) = 1$ V

For a symmetric inverter, the sum of V_{IL} and V_{IH} is always equal to V_{DD}.

$V_{IL} + V_{IH} = V_{DD}$

$0.8 + 1 = 1.8$ V

Therefore, inverter is symmetric.

University Questions

1. *Draw CMOS inverter circuit. Obtain expressions for V_{IL} and V_{th}.* **GTU : Winter-14, Marks**
2. *In CMOS inverter circuit, if the supply voltage is reduced below the summation of $V_{T,n}$ a $|V_{T,P}|$, how the output voltage will follow the change in input voltage (i.e. draw VTC a explain its behavior) ? Assume that enhancement MOSFET devices are used in the CMOS inver circuit.* **GTU : Winter-14, Marks**
3. *Draw the CMOS inverter circuit and Voltage Transfer Characteristic (VTC) for different operati regions of the nMOS and pMOS transistors. Derive critical voltage points V_{OH}, V_{OL}, V_{IL}, V_{IH}* **GTU : Summer-15, Marks**
4. *Draw circuit of CMOS inverter. Drive V_{IH}, V_{IL}, V_{OL} and V_{OH} for CMOS inverter.* **GTU : Winter-21, Marks**
5. *Sketch CMOS inveter circuit and its VTC. Describe various regions of VTC. Derive equation Vth.* **GTU : Summer-21, Marks**

4.5 Short Questions and Answers

Q.1 Draw transfer characteristics of CMOS inverter.

Ans. :

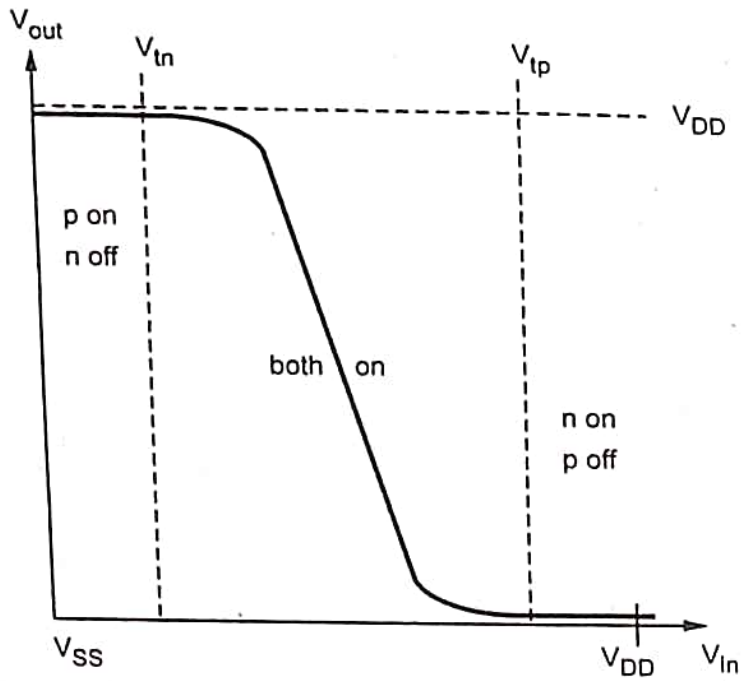


Fig. 4.5.1

Q.1 Why NMOS technology is preferred more than PMOS technology ?

Ans. : NMOS is better than PMOS

- N-channel MOSFETs have some inherent performance advantages over p-channel MOSFET's. The mobility of electrons, which are carriers in the case of an n-channel device, is about two times greater than that of holes, which are the carriers in the p-channel device. Thus an n-channel device is faster than a p-channel device.
- As conductivity is proportional to mobility, NMOS offers less RDS (on) and gate capacitance than PMOS for same die area.
- As the mobility of the carriers in an NMOS is approximately 2 to 3 times higher than that of a PMOS, for the same RDS (on) value, the PMOS must be 2 to 3 times the size of the NMOS.
- In high frequency applications PMOS results into higher switching losses due to their higher gate charge requirement than NMOS for same current rating.

Q.3 Draw the CMOS Inverter circuit.

Ans. :

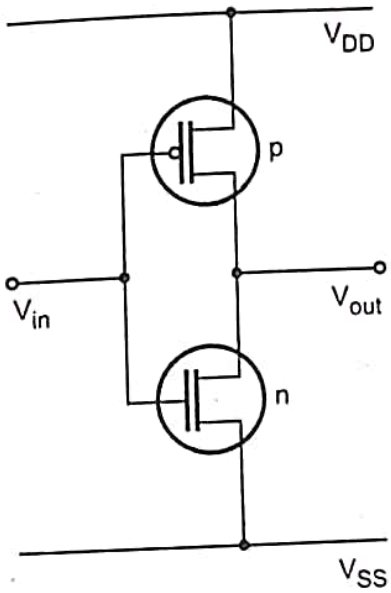


Fig. 4.5.2

Q.4 Explain noise margin.

Ans. : Noise Margin :

- Noise margin is the amount of noise that a CMOS circuit could withstand without compromising the operation of circuit.
- Noise margin does make sure that any signal which is logic '1' with finite noise added to it, is still recognised as logic '1' and not logic '0'. It is basically the difference between signal value and the noise value.

- Noise margin is a parameter closely related to the input-output voltage characteristics.

Q.5 What are the issues involved in driving large capacitive loads in VLSI circuits.

Ans. : Problems of driving large capacitive loads :

- For driving large capacitive loads, inverters should present low pull-down and pull-up resistances. This, in turn means that MOS device must be designed with Low L : W ratios to have low resistance values for $Z_{p.u.}$ and $Z_{p.d.}$ For this channels must be made very wide to reduce resistance value which is consequence makes the inverter occupy a larger area.
- Another limitation on L is that it cannot be reduced below the minimum feature size which makes L : W ratio large. Hence the gate region area L × W becomes significant and a comparatively large capacitance is presented at the input. This in turn increases time required for transitions of voltages at the input.

□□

5 MOS Inverters Switching Characteristics and Interconnect Effects

Syllabus

Introduction, Delay-time definitions, Calculation of Delay times, Inverter design with delay constraints, Estimation of Interconnect Parasitic, Calculation of interconnect delay, Switching Power Dissipation of CMOS Inverters.

Contents

5.1 Introduction

5.2 Delay-Time Definitions

5.3 Calculation of Delay Times . . . . . **Summer-15, Winter-21,** · · · Marks 7

5.4 Inverter Design with Delay Constraints. **Winter-14, 21,**

. . . . . **Summer-21,** · · · · · · · · · · · Marks 7

5.5 Estimation of Interconnect Parasitics

5.6 Calculation of Interconnect Delay. . . . . **Summer-21,** · · · · · · · · · · · Marks 7

5.7 Switching Power Dissipation of CMOS Inverters

. . . . . **Winter-21,** · · · · · · · · · · · · · Marks 3

5.8 Short Questions and Answers

5.1 Introduction

- In any CMOS circuit, the output is dependent on various types of capacitance such as -
  1. Parasitic capacitance.
  2. Gate overlap with diffusion.
  3. Voltage dependent junction capacitance.
  4. Thin-oxide capacitance over gate area.
  5. Metal or polysilicon capacitance.
- The combined capacitance at output node is sum of all capacitances.

5.1.1 Propagation Delay

- **Propagation delay** is the time required for the input to be propagated to th output. In other words, the propagation delay of a logic gate is defined as th time it takes for the effect of change in input to be visible at the output.
- Propagation delay is defined as the difference between the times when th transitioning input reaches 50 % of its final value to the time when the outpu reaches 50 % of the final value showing the effect of input change. Here, 50 % the defined as the logic threshold where output (or, in particular, any signal) assumed to switch its states.

5.1.2 Transistor Sizing

- Certain logic circuits have signals with large capacitive loads due to large fan o e.g. clock and reset signals and row line decoders. In these cases, it advantageous to size logic gates to improve the delay between the stages.
- The level at which this can be exercised depends on design style used. Standar cell designs might have a wider range of size options for gates.
- Custom designs allow continuous variability of each transistor size in each gate.
- In 8 input AND gate, the total area of each implementation is proportional to th total size of transistors used.
- In row decoder, it is better to reduce the area and fan in to the smallest possib value because multiple row drivers are needed to drive different rows of memory
- In designing circuits at gate level, one must balance the time to optimize suc gates versus overall effect on the system. Minimum size means different things different technologies.
- In custom design, a transistor that is minimum size that geometric design rules allow. In gate array system it is transistor pair 5 to 10 times the size of smalle pair that can be fabricated.

5.2 Delay-Time Definitions

- The input and output voltage waveforms of a typical inverter is shown in Fig. 5.2.1.

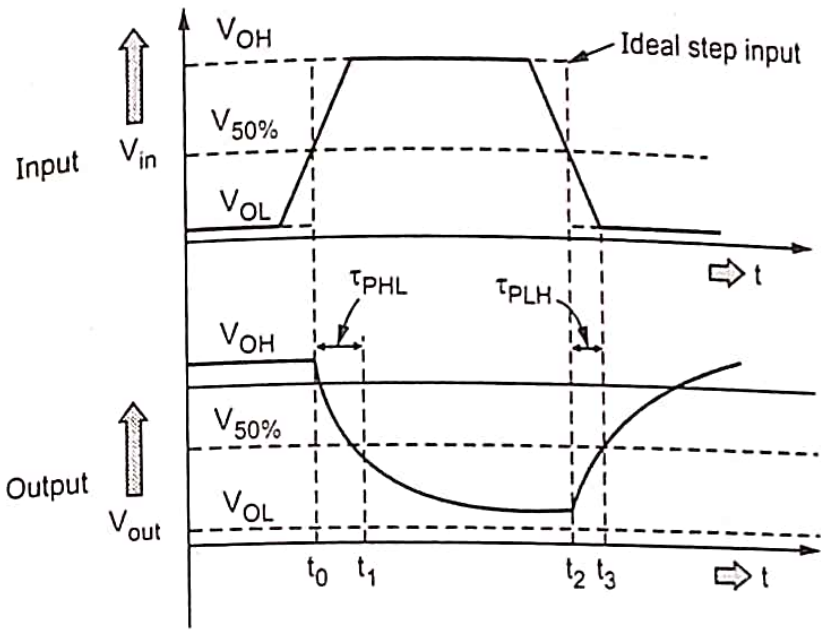


Fig. 5.2.1 Input output waveform of inverter

τ_{PHL} :

- τ_{PHL} is defined as the time delay between $V_{50\ \%}$ transition of the rising input voltage and the $V_{50\ \%}$ transition of the falling output voltage.

$$\tau_{PHL} = t_1 - t_0$$

τ_{PLH} :

- τ_{PLH} is defined as the time delay between $V_{50\ \%}$ transition of the falling input voltage and the $V_{50\ \%}$ transition of the rising output voltage.

$$\tau_{PLH} = t_3 - t_2$$

Average propagation delay (τ_p)

- The average propagation delay is defined as average time required for the input signal to propagate through the inverter.

$$\tau_p = \frac{\tau_{PHL} + \tau_{PLH}}{2}$$

5.3 Calculation of Delay Times

GTU : Summer-15, Winter-21

- The propagation delay times are dependent on average capacitance during char down and charge up respectively.
- If a constant average current is I_{avg}, then delay times are :

$$\tau_{PHL} = \frac{C_{Load} \cdot \Delta V_{HL}}{I_{avg} \cdot HL}$$

$$\tau_{PHL} = \frac{C_{Load} \cdot (V_{OH} - V_{50\%})}{I_{avg} \cdot HL}$$

$$\tau_{PLH} = \frac{C_{Load} \cdot \Delta V_{LH}}{I_{avg} \cdot LH}$$

$$\tau_{PLH} = \frac{C_{Load} \cdot (V_{50\%} - V_{OL})}{I_{avg} \cdot LH}$$

After substituting and solving different parameters, the propagation delay time f high-to-low output transition (τ_{PHL}) is expressed as

$$\tau_{PHL} = \frac{C_{Load}}{K_n(V_{OH} - V_{T,n})} \left[\frac{2V_{T,n}}{V_{OH} - V_{T,n}} + ln\left\{ \frac{4(V_{OH} - V_{T,n})}{V_{OH} + V_{OL}} - 1 \right\} \right]$$

For CMOS inverter, $V_{OH} = V_{DD}$ and $V_{OL} = 0$

$$\therefore \quad \tau_{PHL} = \frac{C_{Load}}{K_n(V_{DD} - V_{T,n})} \left[\frac{2V_{T,n}}{V_{DD} - V_{T,n}} + ln\left\{ \frac{4(V_{DD} - V_{T,n})}{V_{DD}} - 1 \right\} \right]$$

University Questions

1. *Derive the expression for τ_{PHL} for CMOS inverter for ideal step as an input to CMOS inverter.*

   GTU : Summer-15, Marks 7

2. *Define the propagation delay τ_{PHL} and τ_{PLH} derive τ_{PHL} for CMOS inverter.*

   GTU : Winter-21, Marks 7

5.4 Inverter Design with Delay Constraints

GTU : Winter-14, 21, Summer-21

- In any digital circuit, timing or delay specification determines the overa performance of complex systems.
- The delay constraints are considered with other constraints such as noise margin logic threshold, silicon area and power dissipation.
- For optimum overall performance most of the conflicting requirements a balanced by design engineer.

- General circuit for inverter design is shown in Fig. 5.4.1.
- For satisfying certain timing requirements, the channel dimensions (W_n, W_p) of nMOS and pMOS transistors are to detemine.
- The load capacitance C_{Load} mainly consists of extrinsic components such as interconnect, wiring capacitance and fanout capacitance.

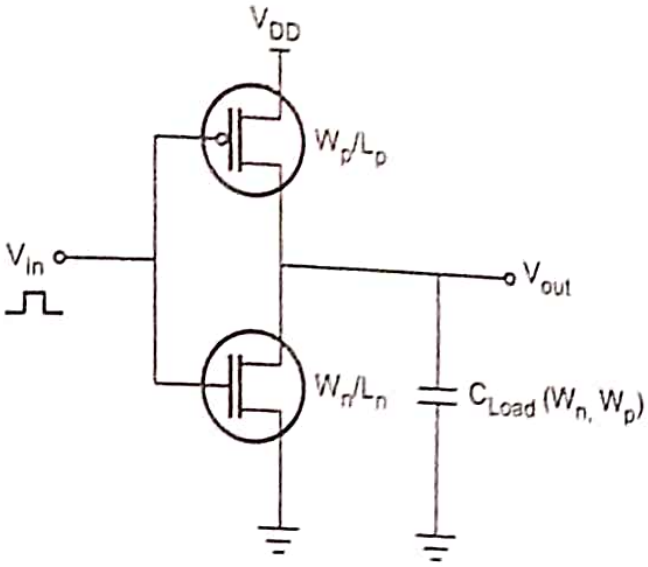


Fig. 5.4.1 Inverter circuit for design

- Accurate estimation of C_{Load} reduces the inverter design problem significantly.

(W/L)-ratio for nMOS

$$\left(\frac{W_n}{L_n}\right) = \frac{C_{Load}}{\tau^*_{PHL}\mu_n C_{ox}(V_{DD} - V_{T,n})}\left[\frac{2V_{T,n}}{V_{DD} - V_{T,n}} + ln\left\{\frac{4V_{DD} - V_{T,n}}{V_{DD}} - 1\right\}\right]$$

(W/L)-ratio for pMOS

$$\left(\frac{W_p}{L_p}\right) = \frac{C_{Load}}{\tau^*_{PLH}\mu_p C_{ox}(V_{DD} - |V_{T,p}|)}\left[\frac{2|V_{T,p}|}{V_{DD} - |V_{T,p}|} + ln\left\{\frac{4V_{DD} - |V_{T,p}|}{V_{DD}} - 1\right\}\right]$$

where, τ^*_{PHL} and τ^*_{PLH} are target delay of respective transistors.

5.4.1 CMOS Ring Oscillator Circuit

- A three stage ring oscillator circuit consisting of three identical CMOS inverters is shown in Fig. 5.4.2.
- Three stage inverters forms a voltage feedback loop.

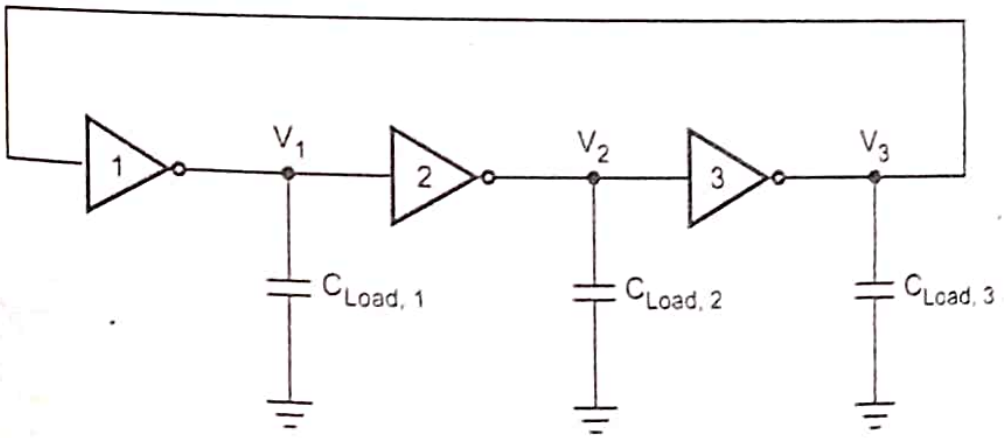


Fig. 5.4.2 CMOS ring oscillator

- The oscillator circuit does not have a stable operating point but only DC operatin point at which the input and output voltages of all inverters are equal to log threshold (V_{th}).
- A closed-loop cascade circuit of any odd number will oscillate if any of th inverter input or output voltage deviate from unstable operating point (V_{th} Hence, the circuit is called **ring oscillator.**
- Output voltage waveforms of three inverters during oscillation is shown i Fig. 5.4.3.

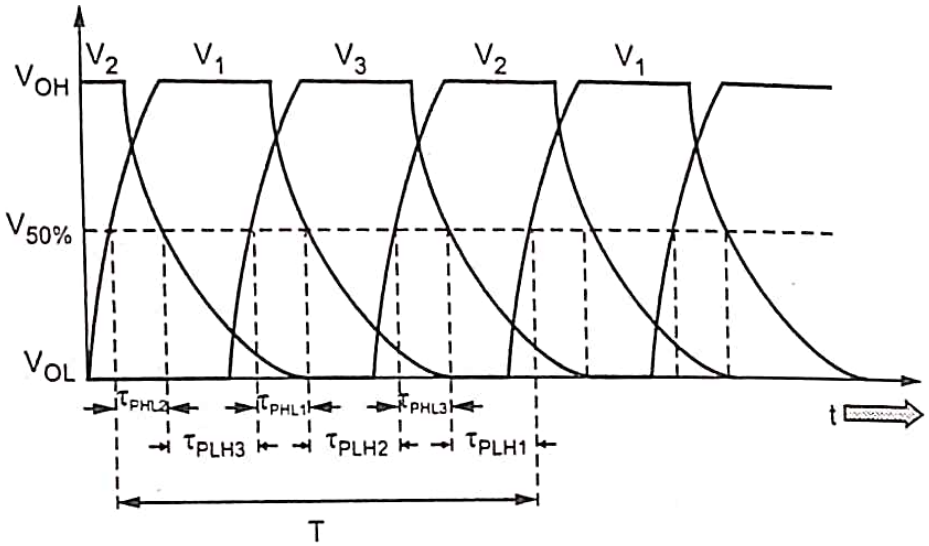


Fig. 5.4.3 Voltage waveforms of ring oscillator

- The oscillation period T is sum of six propagation delay times.

$$T = \tau_{PHL1} + \tau_{PLH1} + \tau_{PHL2} + \tau_{PLH2} + \tau_{PHL3} + \tau_{PLH3}$$

$$T = 2\,\tau_p + 2\,\tau_p + 2\tau_p$$

$$T = 6\,\tau_p$$

- For an odd number (n) of cascaded inverters, the oscillating frequency

$$f = \frac{1}{T} = \frac{1}{2n\tau_p}$$

- Propagation delay is given by

$$\tau_p = \frac{1}{2n(f)}$$

University Questions

1. *Derive expression for frequency of oscillation for three stage ring oscillator circuit. Draw necessary circuit and waveforms.* **GTU : Winter-14, Marks 7**
2. *Write a note on CMS ring oscillator circuit.* **GTU : Summer-21, Marks 4**
3. *What is CMOS ring oscillator ? Explain in brief.* **GTU : Winter-21, Marks 3**
4. *Describe CMOS ring oscillator.* **GTU : Summer-21, Marks 3**

5.5 Estimation of Interconnect Parasitics

5.5.1 CMOS Parasitic Equivalent Circuit

- Because of parasitic elements a condition called Latchup exists where high current flows through V_{DD} and Ground. The Latchup condition may result in malfunctioning sometimes destruction of device.
- Fig. 5.5.1 shows equivalent circuit of parasitic elements of a CMOS inverter. Inverter consists of n-channel and p-channel MOSFET. Two parasitic bipolar transistors vertical pnp (Q_1) and lateral (Q_2) is shown.

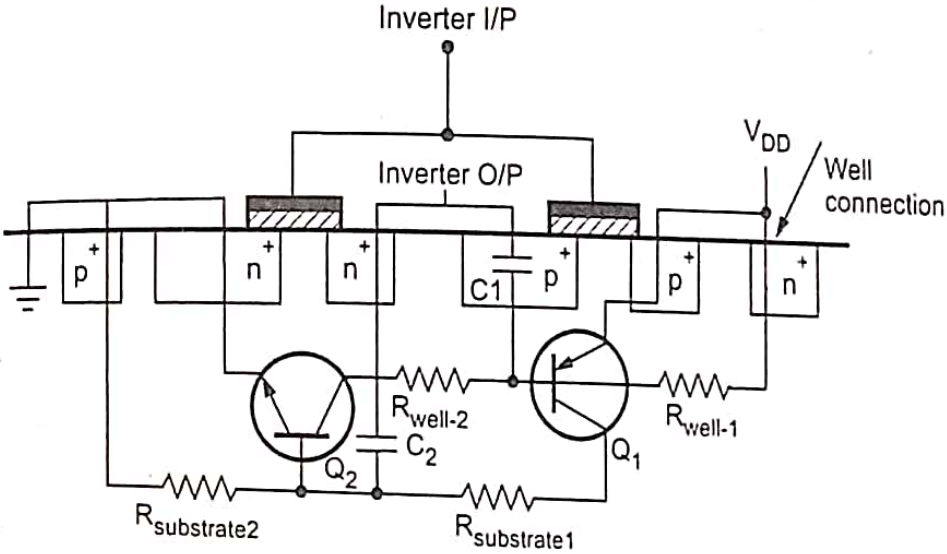


Fig. 5.5.1 CMOS with parasitic elements

- R_{well-1} and R_{well-2} are resistances of n-well and $R_{substrate\,1}$ and $R_{substrate\,2}$ are resistances of substrate. C_1 and C_2 represents the drain implant depletion capacitance.

5.5.2 Interconnect Parasitics

- Interconnects are used to link transistors. Interconnect plays a major role in performance of a system. A pair of adjacent interconnect is shown in Fig. 5.5.2.

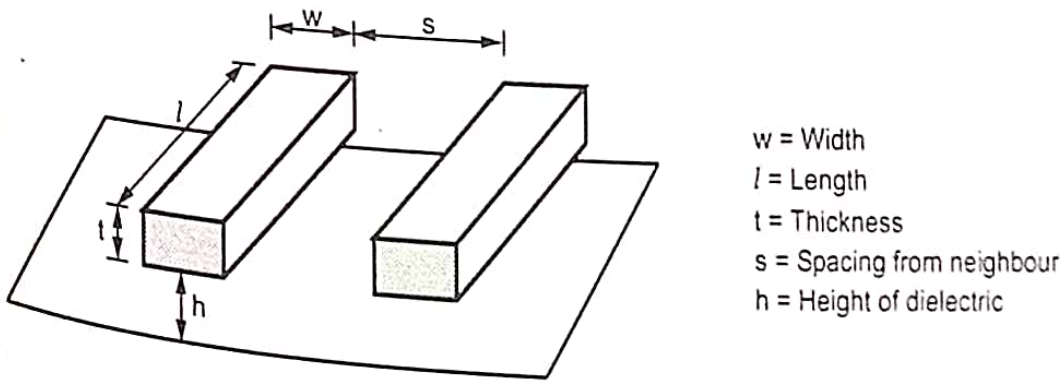


Fig. 5.5.2 Interconnect geometry

$$\text{pitch} = w + s$$

$$\text{aspect ratio} = \frac{t}{w}$$

- The dielectric is of low-k material, usually SiO_2. The dielectric is placed betwee interconnect and conducting layer below.

5.5.2.1 Resistance

- The resistance of a uniform slab of conducting material is given by -

$$R = \frac{\rho}{t} \cdot \frac{l}{w}$$

$$R = R_{\square} \cdot \frac{l}{w}$$

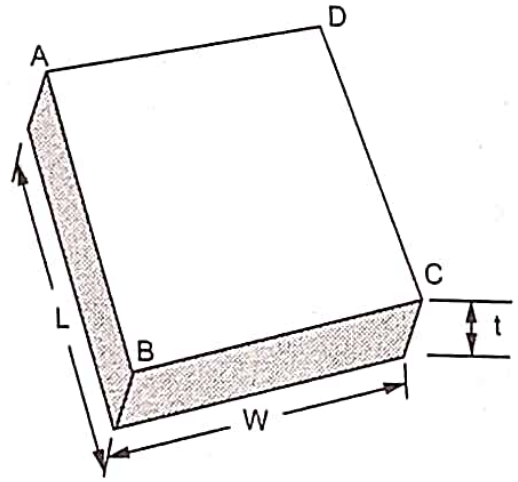
where,

ρ is resistivity

$R_{\square}$ is sheet resistance. (Ω/square)

Sheet Resistance (R_s)

- The sheet resistance R_s is a measure of resistance of a thin film that have uniform thickness. The sheet resistance is used to evaluate the outcome semiconductor doping, metal deposition and resistive paste printing.
- A uniform slab of conducting material is shown in Fig. 5.5.3 with dimensions.



Width = W
Thickness = t
Length = L
Resistivity = ρ
Area = A

Fig. 5.5.3 Sheet resistance model

- Resistance of sheet along its length AB is given by

$$R_{AB} = \frac{\rho L}{A}\ \Omega$$

$$R_{AB} = \frac{\rho L}{tW} \qquad \because W = L \text{ for uniform slab}$$

$$R_{AB} = \frac{\rho}{t} \qquad \text{Sheet resistance } R_s = R_{AB}$$

$$R_s = \frac{\rho}{t}\ \Omega/\text{cm}$$

Therefore, R_s is independent of area and length.

- The V-I characteristics of MOS transistors is non-linear therefore channel resistance is to be estimated.

Channel resistance $R_C = k \cdot \frac{L}{W}$

where, $$k = \frac{1}{\mu C_{ok} (V_{gs} - V_t)}$$

Value of k for p-channel and n-channel is ranging between 1000 to 30,000 Ω/Sq.

- Typical sheet resistance for various layers is summarized in Table 5.5.1.

Material layer	Sheet resistance (R_s) Ω/SQ		
	1.2 μm	2 μm	5 μm
Metal	0.05	0.07	0.03
Polysilicon	30	20	15
Silicide	6	3	2
Diffusion (n^+, p^+)	100	25	10
Silicide Diffusion	10	4	2
n-well	5 k	2 k	1 k

Table 5.5.1

5.5.2 Capacitance

- A wire over sustrate is considered as a conductor over a ground plane. The two components of wire capacitance are :
  1. Parallel plate capacitance 2. Fringing capacitance.
- The parallel plate capacitance exists between bottom of wire to ground and fringing capacitance arises from fringing fields along the edge of conductor. The effects of fringing fields capacitance is shown in Fig. 5.5.4.

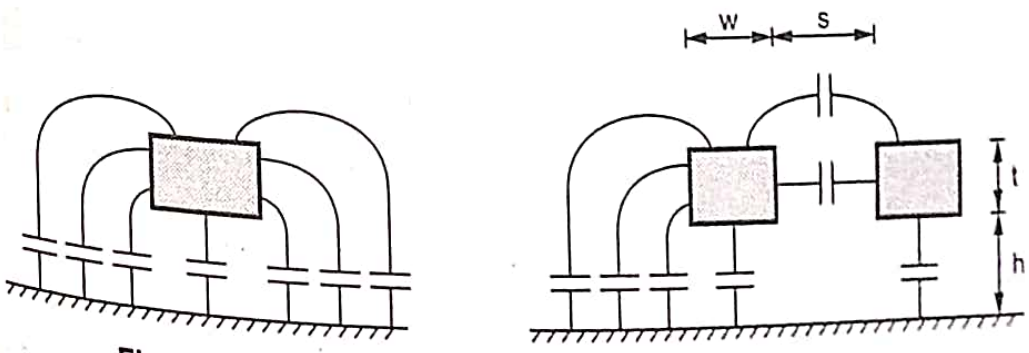


Fig. 5.5.4 Parallel plates and fringing field capacitance

Capacitance due to Fringing Fields

- The effect of fringing fields is to increase the effective area of the plates formi parallel plate capacitance. The capacitances due to the fringing fields can b major component of the overall wiring capacitances. When metallization is in form of fine lines, the value of fringing capacitance may be comparable to tha the area capacitance. Hence for accurate evaluation of circuit performar fringing field capacitances are required to be taken into consideration. Frin field capacitance (C_{ff}) is given by,

$$C_{ff} = \varepsilon_0 \varepsilon_{SiO_2}\, l \left[\frac{\pi}{ln\left\{1+\frac{2d}{t}\left(\sqrt{1+\frac{t}{d}}+1\right)\right\}} - \frac{t}{4d} \right]$$

where t is thickness of wire (metallization), l is wire length and d is distance betw wire and substrate. Total wire capacitance comes out to

$$C_w = C_{ff} + C_{area}$$

5.5.2.3 Delay

- Interconnect increases circuit delay because -
  1. The wire capacitance adds loading to each gate.
  2. The wire resistance adds distributed RC delay (flight time).
- The distributed R and C of interconnect can be approximated by numbe lumped elements such as L-model, π-model and T-model.

5.5.2.4 Crosstalk

- When one interconnect (wire) switches, it affect its adjacent interconnect through capacitive coupling, this effect is called **crosstalk.** Fig. 5.5.5 shows capacitive coupling between interconnect A and B.
- The Miller Coupling Factor (MCF) is multiplied to adjacent capacitance gives effective capacitance. Usually MCF = 1.5 is used as statistical compromise for estimating propagation delays.

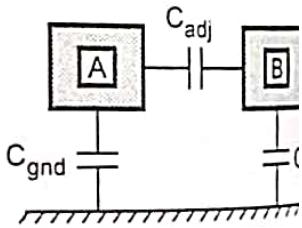


Fig. 5.5.5 Capacitive coupling

5.6 Calculation of Interconnect Delay

GTU : Summer-2

5.6.1 RC Delay Model

- The delay of the logic gate is computed as the product of RC, where R b effective driver resistance and C is load capacitance.
- Minimum device length reduces delay area and power consumption. The dela logic gate depends on widths of transistors in gate and capacitance of the load.

Equivalent RC model

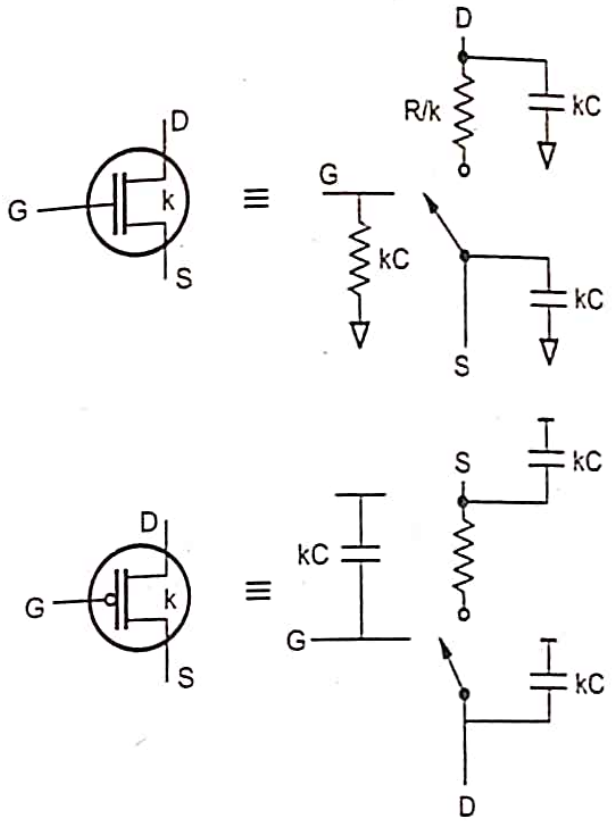


Fig. 5.6.1

5.6.1.1 Effective Resistance and Capacitance

- When multiple transistors are in series, the equivalent resistance is the sum of each individual resistance. When multiple transistors are in parallel, the equivalent resistance is lower.
- Capacitance is contributed by gate capacitance and source/drain diffusion capacitance. The gate capacitance includes fields terminating on the channel, source and drain.

5.6.1.2 Diffusion Capacitance Layout Effects

- Diffusion capacitance can be reduced by sharing diffusion nodes wherever possible. The uncontacted diffusion nodes between series transistors are smaller than contacted nodes. The uncontacted nodes have less capacitance.
- The gate capacitance can be determined from the widths in the schematic. The diffusion capacitance depends on the layout.

5.6.2 Elmore Delay Model

- Elmore delay is a simple approximation, often used where speed of calculation is important but the delay through the wire itself cannot be ignored.
- Elmore delay uses the R and C values of the wire segments in a simple calculation. The delay of each wire segment is the R of that segment times the C. Then all delays are summed from the root.

- For tree structured network (in most chips), the Elmore delay can be calculated in time with two tree traversals. If the network is not tree structured the Elmore delay can still be computed, but involves matrix calculations.
- Considering ON transistors as resistors, a chain of transistors can be represented as an RC ladder as shown.

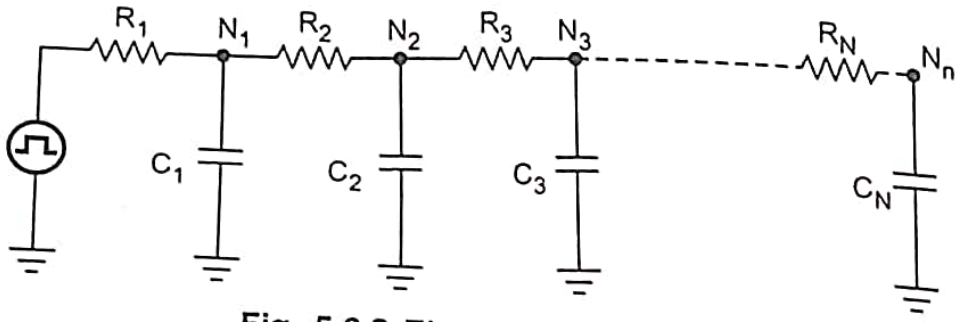


Fig. 5.6.2 Elmore delay model

- The Elmore delay model computes the delay of an RC ladder as the sum of RC delay at each node.

$$t_{pd} = \sum_{i} R_{n-i} C_i$$

$$t_{pd} = \sum_{i=1}^{N} C_i \sum_{j=1}^{i} R_j$$

- Boosting the width of the transistors decreases the resistance but increases the capacitance so that parasitic delay is ideally independent of gate size.

5.6.2.1 Applications of Elmore Delay

- Elmore delay metric is a widely used model to compute signal delays for both analog and digital circuit interconnects.
- The Elmore delay formula is extremely useful in analyzing wires and to approximate the propagation delay of complex transistor networks.
- In the switch model, transistors are replaced by their equivalent, linearized on-resistance. The evaluation of the propagation delay is then reduced to the analysis of the resulting RC network.
- More precise minimum and maximum bounds on the voltage waveforms in an RC tree have further been established. These bounds have formed the base for most computer-aided timing analyzers at the switch and functional level.
- An interesting result is that the exponential voltage waveform with the Elmore delay as time constant is always situated between these min and max bounds, which demonstrates the validity of the Elmore approximation.
- The lower bound of the delay formulation is considered as the basis for the development of the new model.

Design

3 Linear Delay Model

The propagation delay of a gate is given by

$$d = f + p$$

where,

is parasitic delay when no load is attached.

is effort delay (stage effort).

The effort delay or stage effort is depend on complexity and fanout of the gate.

$$fd = gh$$

where,

is complexity or logical effort; h is fanout.

An inverter has logical effort of 1. A gate driving h identical gates of similar types has fanout or electrical effort of h. When load is not driving identical gates the electrical effort (fanout) is given as -

$$h = \frac{C_{out}}{C_{in}}$$

where,

C_{out} is capacitance of external load (being driven)

C_{in} is input capacitance of gate.

- For an idealized inverter and 2 input NAND gate a plot for normalized delay versus fanout is shown in Fig. 5.6.3.

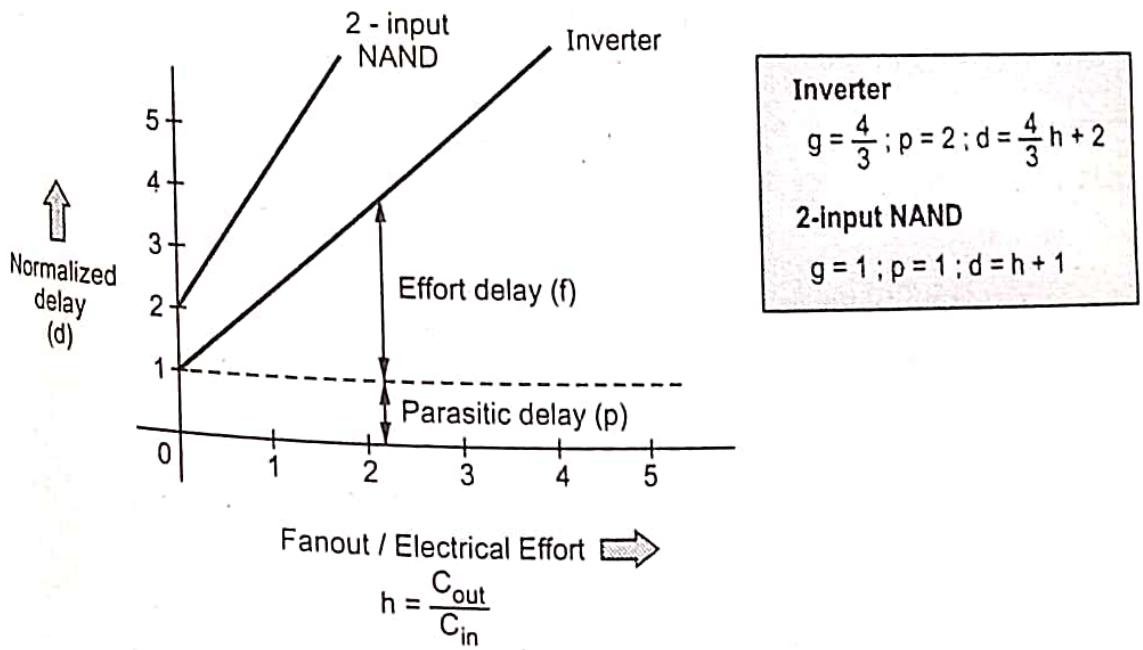


Fig. 5.6.3

- The delay when the gate drives no load is called parasitic delay. The y-interc on graph shows parasitic delay and slope of the lines indicates logical effort. T logical effort term and synopsys term. Relationship is shown in Table 5.6.1.

Logical effort term	Synopsys term
d	delay
p	intrinsic
C_{out}	Capacitance
$\frac{g}{C_{in}}$	Resistance

Table 5.6.1

University Question

1. *Explain the elmore delay calculation method used for complex RC network.*

GTU : Summer-21, Marks 7

5.7 Switching Power Dissipation of CMOS Inverters

GTU : Winter-2

- Static CMOS gates are very power efficient.
- Since they dissipate nearly zero power while idle.

Power Dissipation in CMOS Inverter

- Power dissipation in a CMOS inverter is given as -

$$P = V_{DD} \cdot I_{DC}$$

where,

V_{DD} is power supply voltage.

I_{DC} is current drawn from power supply in steady state.

- In CMOS inverter power dissipation is caused due to various factors, depend on which it is categorized into two types.
  1. **Static power dissipation :** Static power dissipation is due to the leakage cur drawn from the power supply.
  2. **Dynamic power dissipation :** Dynamic power dissipation is due to switch transient current and charging and discharging of load capacitance.

5.7.1 Static Power Dissipation

- The causes of static power dissipation are -
1. Subthreshold conduction through OFF transistors.
2. Tunneling current through gate oxide.
3. Leakage through reverse biased diodes.
4. Contention current in ratioed circuits.
- Static power dissipation is also known as steady state power dissipation. The static power dissipation is given by -

$$P_{static} = I_{static} \cdot V_{DD}$$

where,

I_{static} is the current flowing from supply to ground in absence of switching.

- Static power dissipation is due to leakage current flowing through the reverse biased junctions of transistors. The leakage current per unit drain area typically ranging between 10 to 100 pA/μm^2 at room temperature.
- Junction leakage current is also due to thermally generated carriers. These carriers increases with increase in temperature. The power dissipation increases with increase in package density or number of devices. The total static power dissipation is given as -

$$P_{Total\ static} = \sum_{1}^{n} I_{static} \cdot V_{DD}$$

5.7.2 Dynamic Power Dissipation

- The causes of dynamic power dissipation are -
1. Charging and discharging of load capacitances.
2. "Short circuit" current while both pMOS and nMOS networks are partially ON.
- During switching of transistors current pulses of small duration are produced. Also the load capacitance C_L gets charged through the pMOS transistor, its voltage rises from 0 to V_{DD} and draws energy from power supply.
- If capacitive load increases its charge and discharging dominates the current drawn from power supply.
- For a square wave input of frequency f_p and voltage V_{in}, dynamic power P_d during switching is given by -

$$P_{dynamic} = C\ V_{DD}^2\ f_p$$

5.7.3 Short Circuit Power Dissipation

- It indicates the power dissipation due to short circuit current in an unloaded inverter. Short circuit power dissipation is given as,

$$P_{sc} = \frac{K}{12}(V_{DD} - V_t)\frac{3\,t_{rf}}{t_p}$$

where,

t_p is period of input waveform.

t_{rf} is rise time of input waveform.

K is a constant.

5.7.4 Total Power Dissipation

- Total power dissipation of CMOS inverter is expressed as sum of all power dissipation.

$$P_{total} = P_{static} + P_{dynamic} + P_{sc}$$

5.7.5 Need to Estimate Power Dissipation

Power dissipation affects

1. Performance
2. Reliability
3. Packaging
4. Cost
5. Portability

5.7.6 Factors Influencing CMOS Power Consumption

- There are three major sources of power dissipation in digital CMOS circuits which are summarized in the following equation :

$$P_{avg} = P_{switching} + P_{short-circuit} + P_{leakage}$$

1. Capacitive voltage transitions

- The switching or dynamic component of power consumption arises when the capacitive load (C_L) of a CMOS circuit is charged through PMOS transistors to make a voltage transition from 0 to the high voltage level, which is usually the supply V_{dd}.

2. Logic function

- The type of logic function NAND, NOR or a more complex gate, will all yield different static transition probabilities.

3. **Logic style**

- One basic choice of logic style is between static and dynamic CMOS. For a dynamic implementation of the 2 input NOR gate which is implemented as an NMOS-tree with PMOS precharge, power is consumed during the precharge operation for the times when the output capacitor was discharged the previous cycle.

5.7.7 Minimizing Power Reduction in CMOS Circuits

- Power dissipation is unavoidable especially as technology scales down. Techniques must be devised to reduce power dissipation. Techniques must be devised to accurately estimate the power dissipation.
- Estimation and modeling of the sources of power dissipation can be used for simulation purposes.
- Several techniques can be used for reducing power reduction.

1. Prime choice : Reduce voltage
   - Recent years have seen an acceleration in supply voltage reduction
   - Design at very low voltages still open question (0.6 ... 0.9 V by 2010!)
2. Reduce switching activity.
3. Reduce physical capacitance
   - Device Sizing
4. Reducing the number of transitions on Buses using sign magnitude.
5. Reducing activity for arithmetic computation.
6. Ordering of input signals.
7. Minimizing glitching activity.
8. Degree of resource sharing.
9. Voltage reduction using algorithmic transformations.
10. Minimizing the number of operations : Vector Quantization :
    a] Full search vector quantization.
    b] Tree-structured vector quantization.
    c] Differential codebook tree-structure vector quantization.

iversity Question

1. *Prove that switching dissipation of CMOS inverter is proportional to switching frequency.*

GTU : Winter-21, Marks 3

5.8 Short Questions and Answers

Q.1 Explain the importance of wiring capacitance of a MOS transistor.

Ans. : Wiring capacitance :

- The wiring capacitance depends upon the length and width of the connecting wires and is a function of the fan-out from the driving gate and the number of fan-out gates.
- Wiring capacitance is growing in importance with the scaling of technology.
- The effect of fringing fields is to increase the effective area of the plates forming a parallel plate capacitance.
- The capacitances due to the fringing fields can be a major component of the overall wiring capacitances.
- Wiring capacitance can be classified as :
  i) Capacitance due to fringing fields.
  ii) Interlayer capacitance.
  iii) Peripheral capacitance.

Q.2 List out the sources of static and dynamic power consumption.

Ans. : Power dissipation -

- Power dissipation can be of two types i.e. static and dynamic.
- Static power dissipation takes place
  a) When device holds a particular state.
  b) Due to the leakage current drawn from the power supply.
- Dynamic power dissipation takes place
  a) When device changes its state.
  b) Charging and discharging of load capacitance.

Q.3 Define fan-in and fan-out.

Ans. : Fan-in :

- The number of inputs that a gate can have is called as fan-in. The additive 'ON' resistance of a series transistors limits the fan-in of CMOS gates.

Fan-out :

- The maximum number of inputs of the same logic family that a gate can drive maintaining its output levels within the specified limits is called fan-out of that gate.

Q.4 What are the various ways to reduce the delay time of a CMOS inverter ?

Ans. : Various ways for reducing the delay time are given below :

a) The width of the MOS transistors can be increased to reduce the delay. This is known as gate sizing, which will be discussed later in more details.

b) The load capacitance can be reduced to reduce delay. This is achieved by using transistors of smaller and smaller dimensions as provided by future generation technologies.

c) Delay can also be reduced by increasing the supply voltage V_{dd} and/or reducing the threshold voltage V_t of the MOS transistors.

❑❑❑

Notes

6 Combinational MOS Logic Circuits

Syllabus

Introduction, MOS logic circuits with Depletion nMOS Loads, CMOS logic circuits, Complex logic circuits, CMOS Transmission Gates (TGs).

Contents

6.1 Introduction

6.2 MOS Logic Circuits with Depletion nMOS Loads ..... **Summer-17, Winter-14, 17, 21,** ..... Marks 7

6.3 CMOS Logic Circuit ..... **Winter-21,** ..... Marks 7

6.4 Complex Logic Circuit ..... **Summer-16, 17, 21, Winter-21** ..... Marks 7

6.5 Transmission Gates (TGs) ..... **Summer 15, 16, 17, 21, Winter-21** ..... Marks 7

6.6 Short Questions and Answers

6.1 Introduction

- The gate of a MOS transistor controls the flow of current between the source and drain. Therefore the gate is a control input. Hence by using CMOS transistor all basic logic gates and complex logic gate can be implemented.
- A general form of combinational logic circuit performing a Boolean function can be represented as multiple-input, single-output system as shown in Fig. 6.1.1.

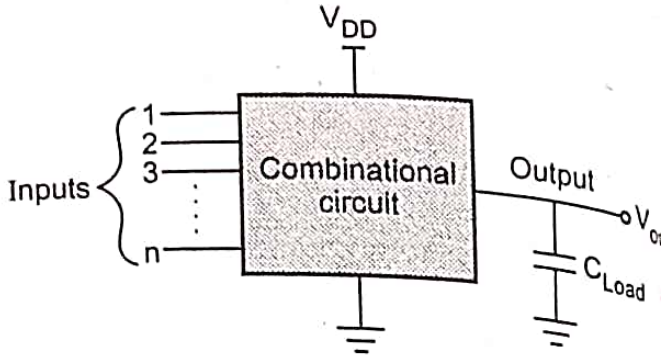


Fig. 6.1.1 General form of combinationa circuit

6.2 MOS Logic Circuits with Depletion nMOS Loads

GTU : Summer-17, Winter-14, 2

6.2.1 Two-Input NOR Gate

- A two input depletion load NOR gate is shown in Fig. 6.2.1.

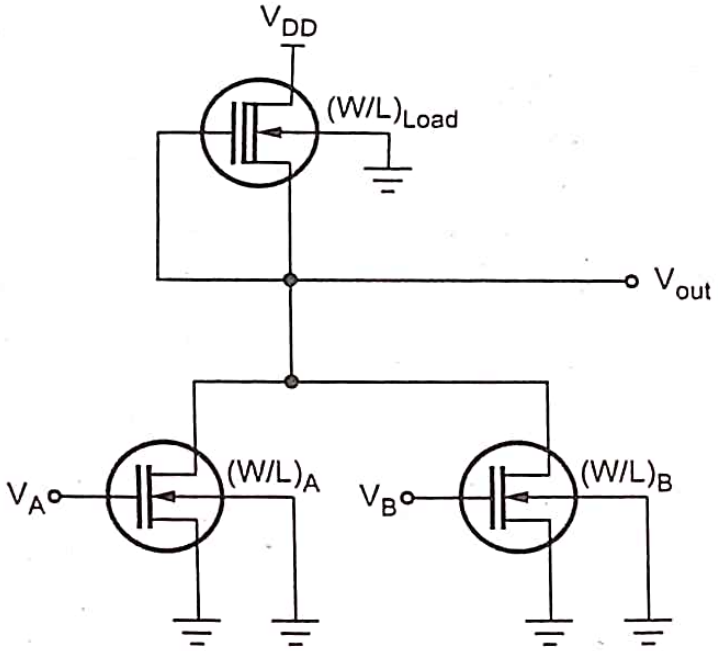


Fig. 6.2.1 Two input depletion load NOR gate

Substrates of all transistors are connected to ground.

- DC analysis of the circuit is simplified by structural similarities between the ci and simple nMOS depletion load invertor.

Calculation of V_{OH}

- By solving linear region current equation gives

$$V_{OH} = V_{DD}$$

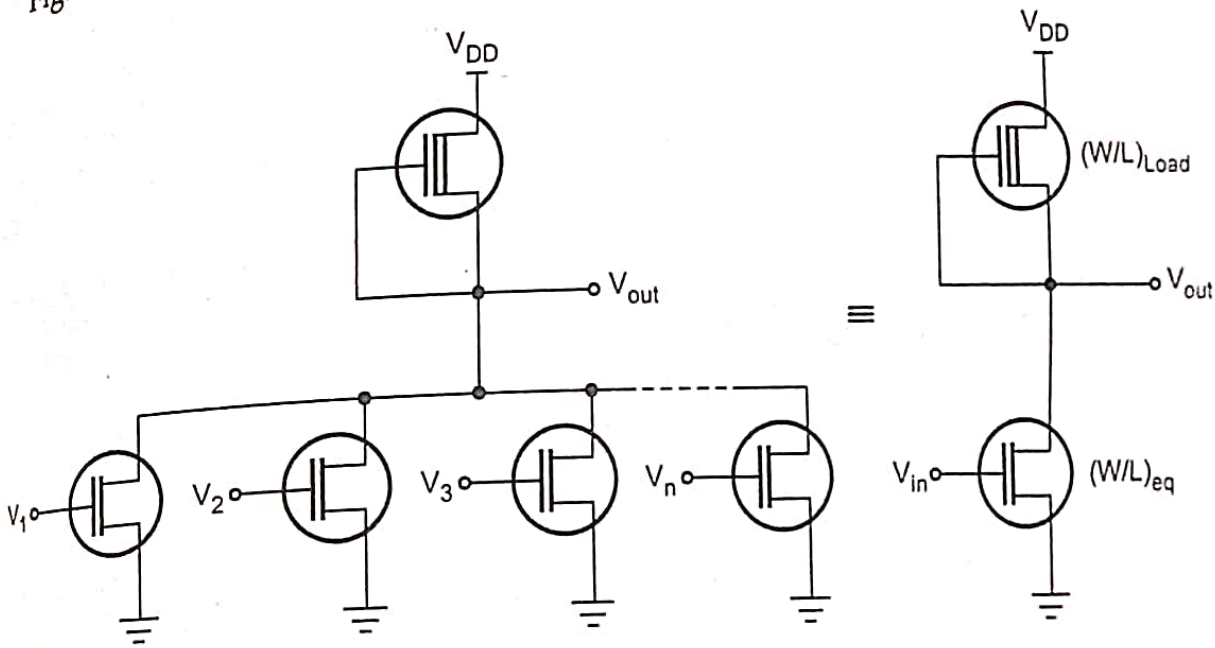
Calculation of V_{OL}

- Considering three different input voltage combination.
- The output-low voltage V_{OL} is given by -

$$V_{OL} = V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{TO})^2 - \left(\frac{K_{load}}{K_{driver}}\right) \cdot |V_{T\,load}(V_{OL})|^2}$$

6.2.2 NOR Structure with Multiple Inputs

- The n-inputs NOR gates consists of an n-parallel driver transistors as shown in Fig. 6.2.2.



(a) General n-input NOR gate (b) Equivalent circuit

Fig. 6.2.2

- The source terminal of all enhancement type nMOS driver transistors are connected to ground. Therefore, the drivers do not experience any substrate bias effect.
- The depletion type n-MOS load transistor is subject to substrate bias effect as source is connected to output.

6.2.3 Two-Input NAND Gate

- Fig. 6.2.3 shows a two input depletion load NAND gate.
- All nMOS transistors except one transistor are subject to substrate bias effect, since their source voltages are larger than zero.

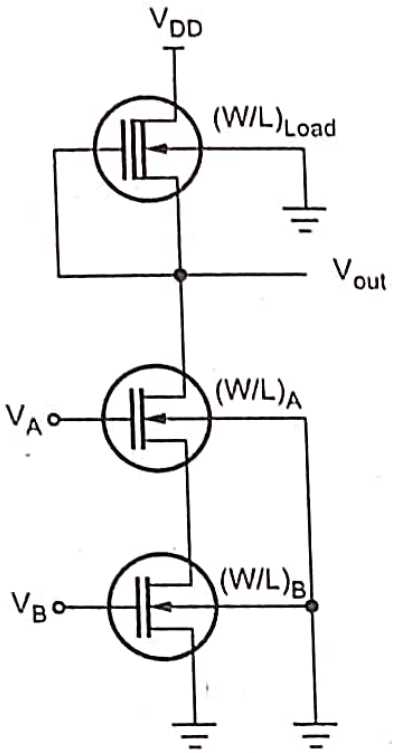
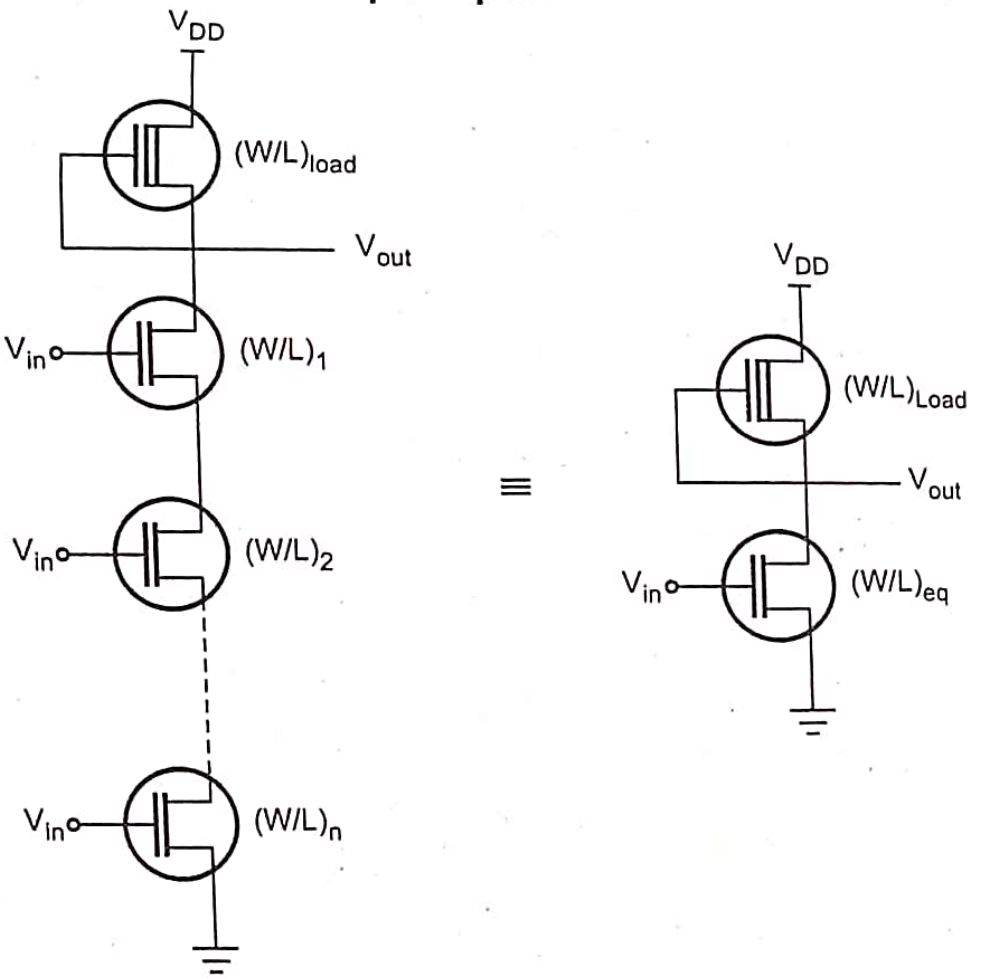


Fig. 6.2.3

6.2.4 NAND Structure with Multiple Inputs



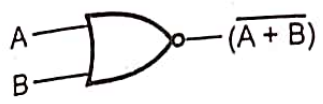
(a) Multiple input NAND structure (b) Equivalent circuit

Fig. 6.2.4 NAND

Example 6.2.1 *Draw the layout of two input CMOS based NOR gate.*

GTU : Summer-17, Mark 1

Solution : 2-input CMOS NOR gate



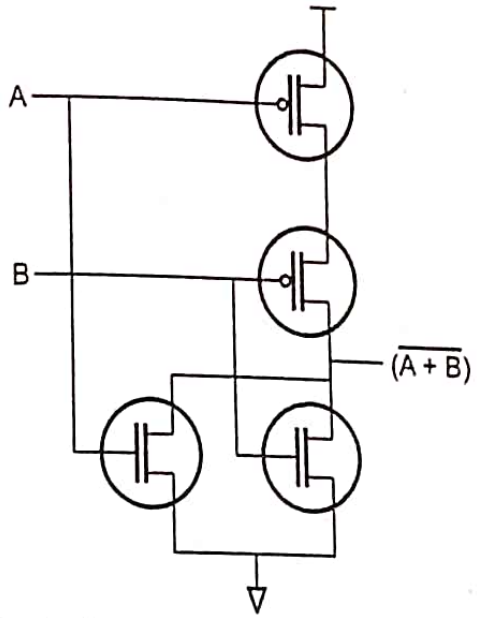


Fig. 6.2.5

University Questions

1. *Draw two-input CMOS NOR and NAND gate circuits. Identify the trasistors suffering from body bias effect in each of these circuits. Assume single-well CMOS technology.*

GTU : Winter-14, Marks 7

2. *Explain generalized NAND structure with multiple inputs using combinational MOS logic circuits.*

GTU : Winter-21, Marks 4

6.3 CMOS Logic Circuit

GTU : Winter-21

- Inverter and NAND gates are examples of complementary CMOS logic gates or static CMOS gates. It has a pull-down network (nMOS) to connect output to '0'(GND) and a pull-up network (pMOS) to connect output to '1'(V_{DD}). The arrangement of network is such that one is ON and other is OFF for any input pattern. Fig. 6.3.1 shows typical arrangement of pull-up and pull-down networks.

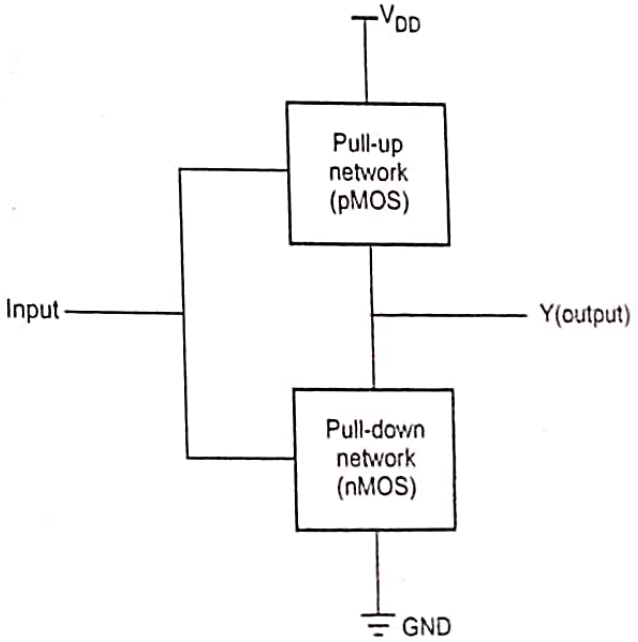


Fig. 6.3.1 Arrangement of pull-up and pull-down network

- The pull-up and pull-down networks may consists of series or paralle combination of pMOS and nMOS transistors. Two or more transistors in series ar ON if all the transistors are ON and the parallel transistors are ON if any of th parallel transistors are ON. By using this principle various combinational logic ca be constructed. Selected connections and behaviour of series and paralle transistors are shown in Fig. 6.3.2.

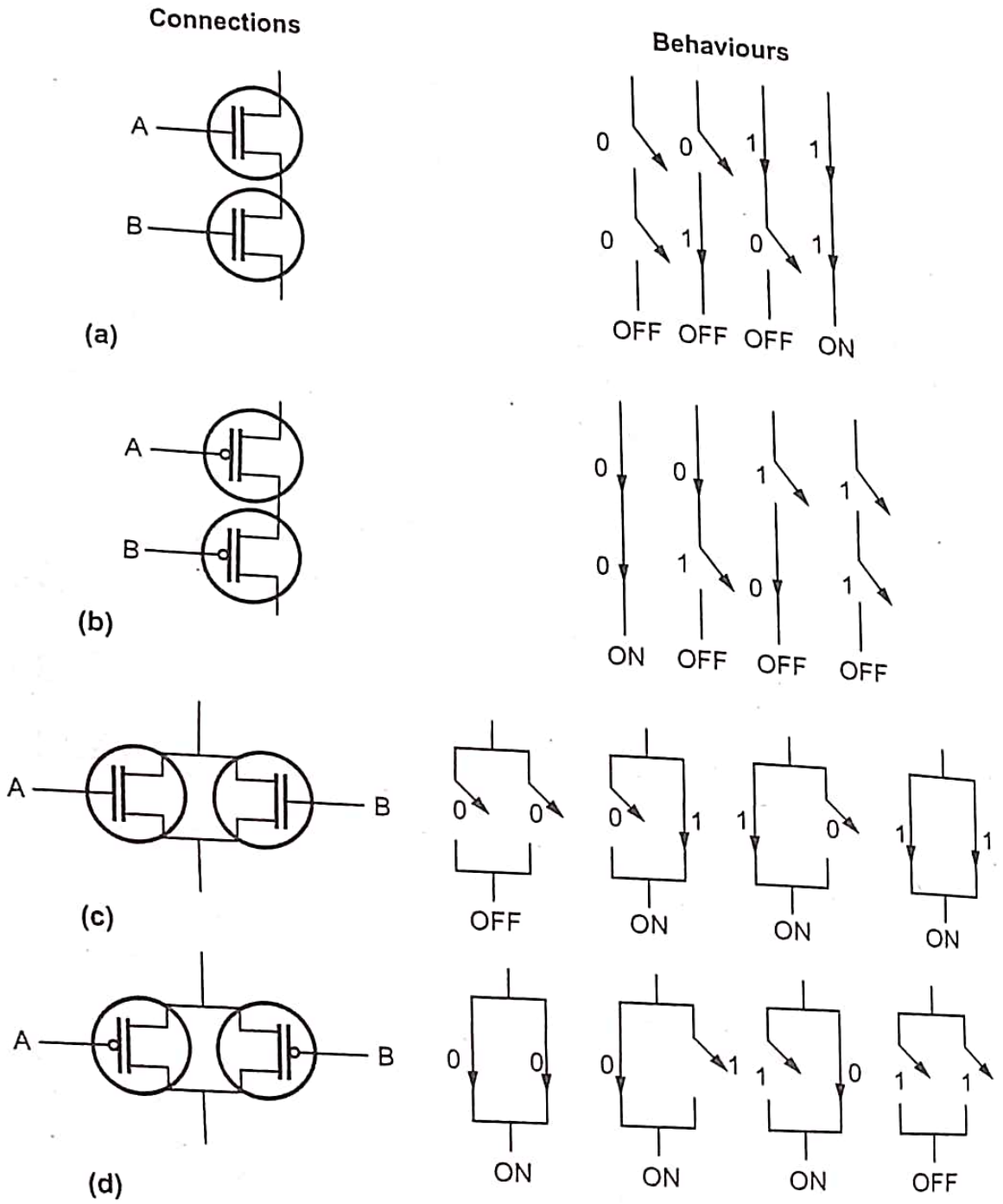


Fig. 6.3.2 Series and parallel connections of transistors

- The output of CMOS logic gate can have four states. Corresponding to var combinations of pull-up and pull-down networks.

	Pull-up OFF	Pull-up ON
Pull-down OFF	Z	1
Pull-down ON	0	Crowbarred (X)

Table 6.3.1 Output states of CMOS logic

- The '1' and '0' levels are the logic states of NAND gate and inverter as either pull-up or pull-down structure is ON. But when both pull-up and pull-down are OFF, the *high impedance* or *floating* Z output state results. The high impedance or floating state is important in multiplexers, memory elements and bus drivers.
- When both either pull-up or pull-down structures are simultaneously ON a state called *Crowbarred* X level, i.e. intermediate level. This is unwanted situation in CMOS digital circuit.

6.3.1 CMOS NOR2 (Two-Input NOR) Gate

- Two input CMOS NOR gate is shown in Fig. 6.3.3.

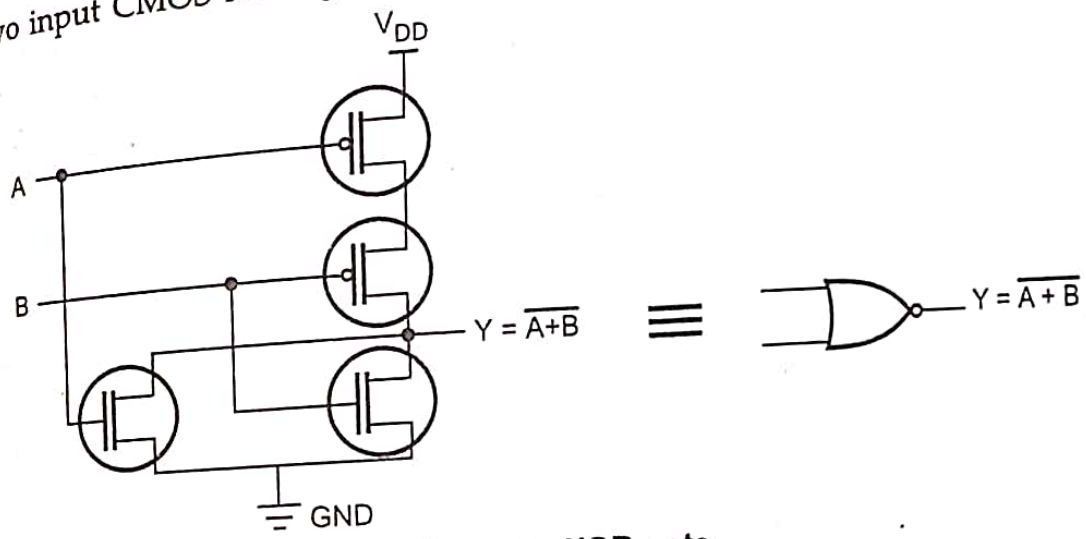


Fig. 6.3.3 NOR gate

- The nMOS transistors are in parallel to pull the output low when any of the input is high. The pMOS transistors are in series to pull the output high when both inputs are low as shown in truth table.

NOR gate truth table		
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

- The crowbarred or floating condition does not exist in NOR gate.

Three input NOR gate

- Fig. 6.3.4 shows schematic for three input NOR gate.

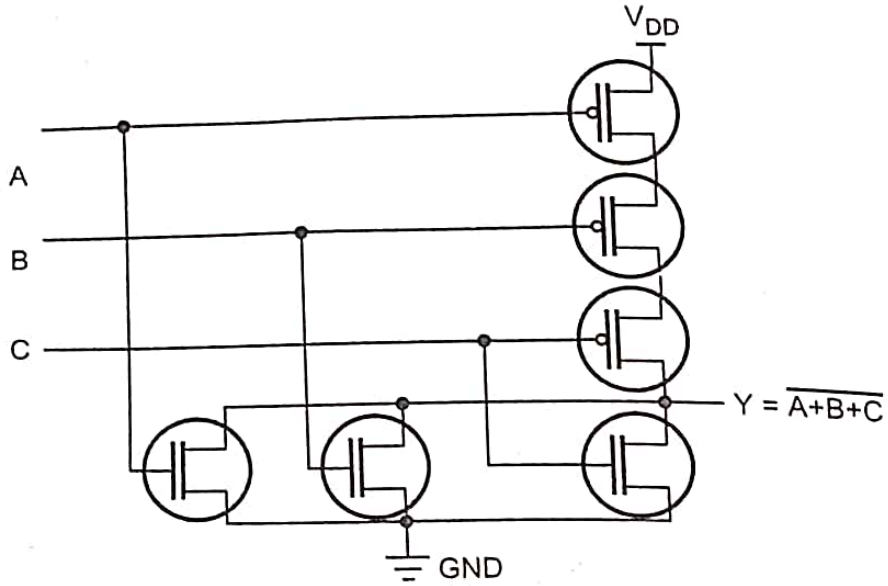
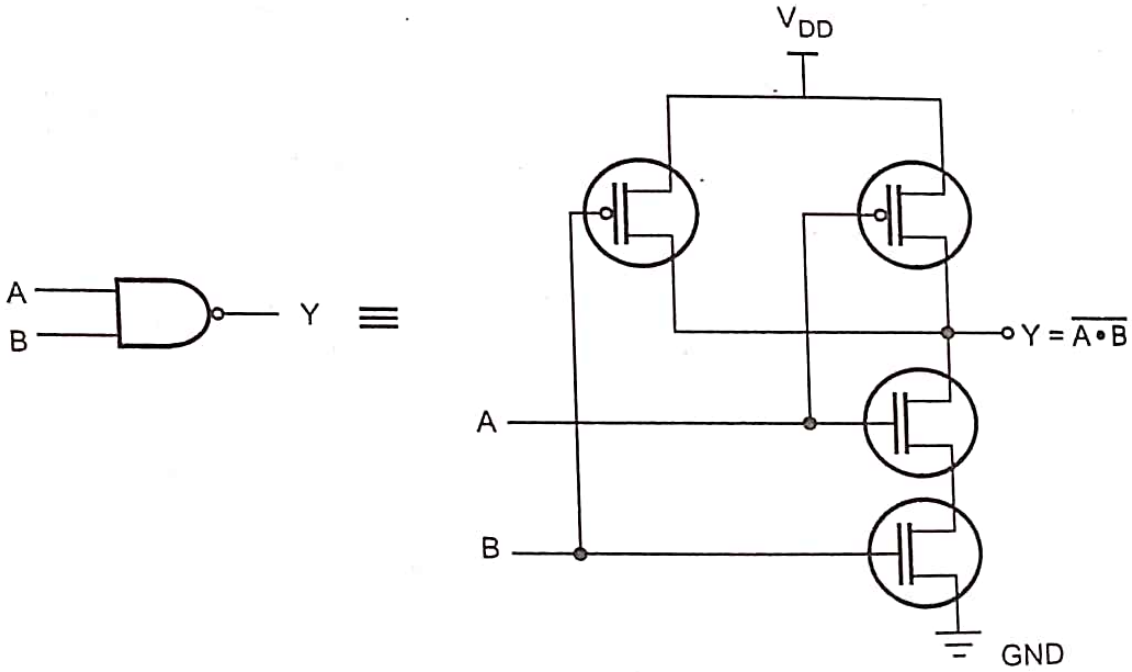


Fig. 6.3.4 Three input NOR gate

- When any of the input is high output is pulled low through parallel nMOS transistors. If all inputs are low, the output is pulled high through the series pMOS transistors.

6.3.2 CMOS NAND2 (Two Input NAND) Gate

- Two input NAND consists of two series nMOS transistors between Y and GND and two parallel pMOS transistors between Y and V_{DD} as shown in Fig. 6.3.5.



(a) Symbol (b) Schematic

Fig. 6.3.5 Two input NAND gate

f either input A or B is '0', one of the nMOS transistor will be OFF, the path between Y and GND. At the same time atleast one of pMOS transistor is ON, making a path from Y to V_{DD}. Therefore, the output will be '1'.

If both the inputs are '1', both nMOS transistors will be ON and both pMOS will be OFF, hence the output will be '0'. Truth table is shown in the Table. 6.3.2.

A	B	Pull-down Network	Pull-up Network	Output (Y)
0	0	OFF	ON	1
0	1	OFF	ON	1
1	0	OFF	ON	1
1	1	ON	OFF	0

Table 6.3.2 Two input NAND gate truth table

e Input CMOS NAND

When any of the inputs are '0', the output is pulled high through the parallel pMOS transistors. When all of the inputs are '1', the output is pulled low through the series of nMOS transistors.

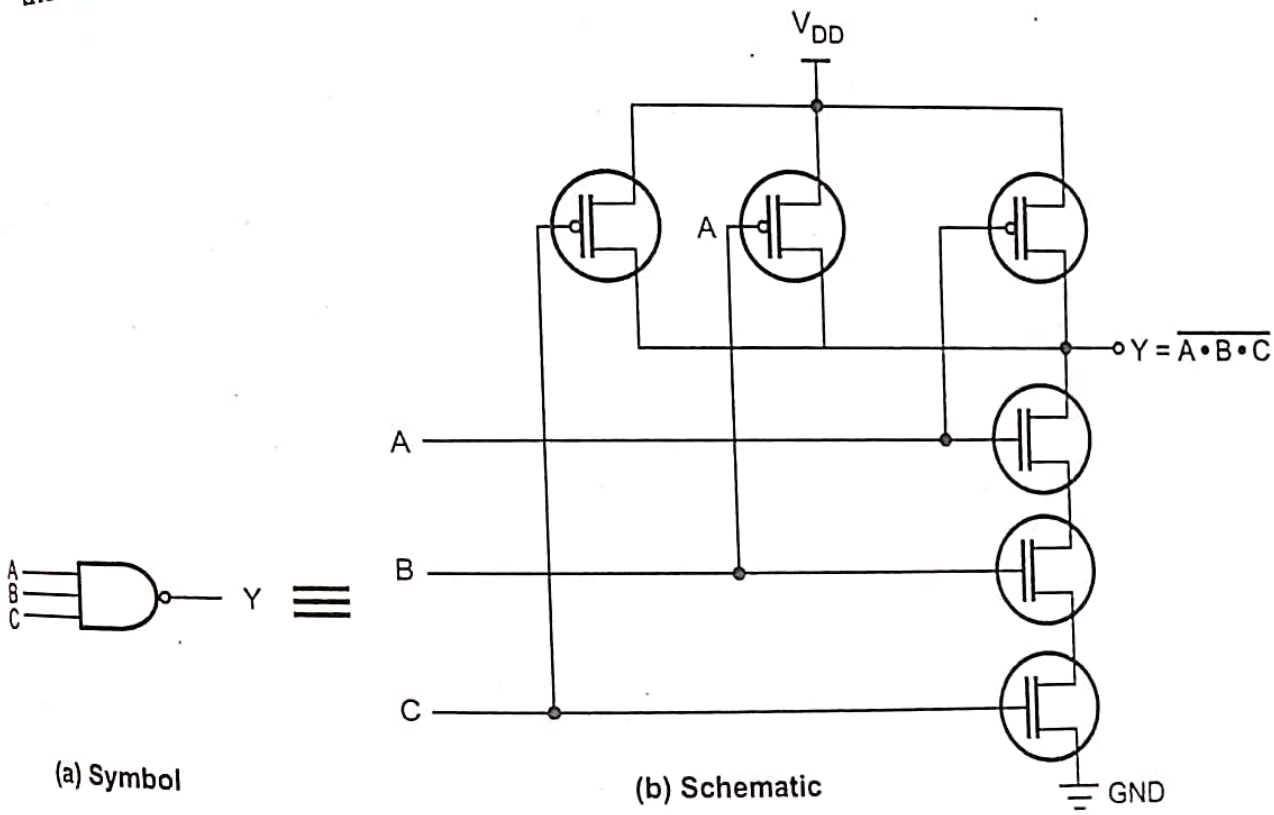


Fig. 6.3.6 Three input NAND gate

6.3.3 Stick Diagram and Layout of CMOS NAND Gate

Example 6.3.1 *Design a stick diagram and layout for two input CMOS NAND gate indicating all the regions and layers.*

Solution : CMOS NAND circuit diagram

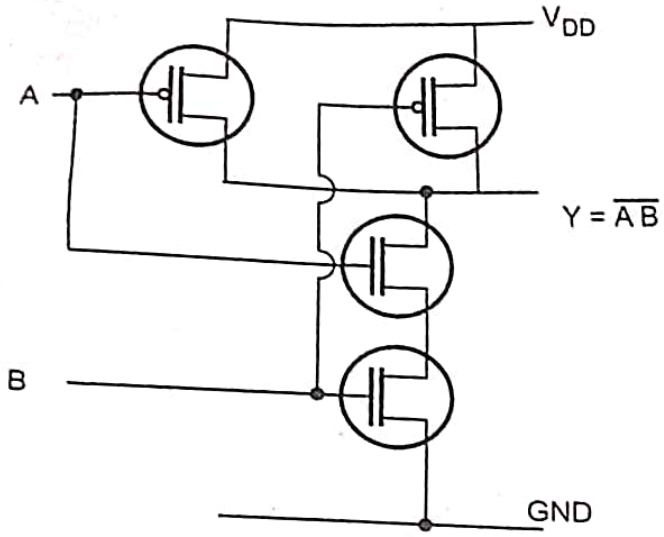


Fig. 6.3.7

Stick diagram

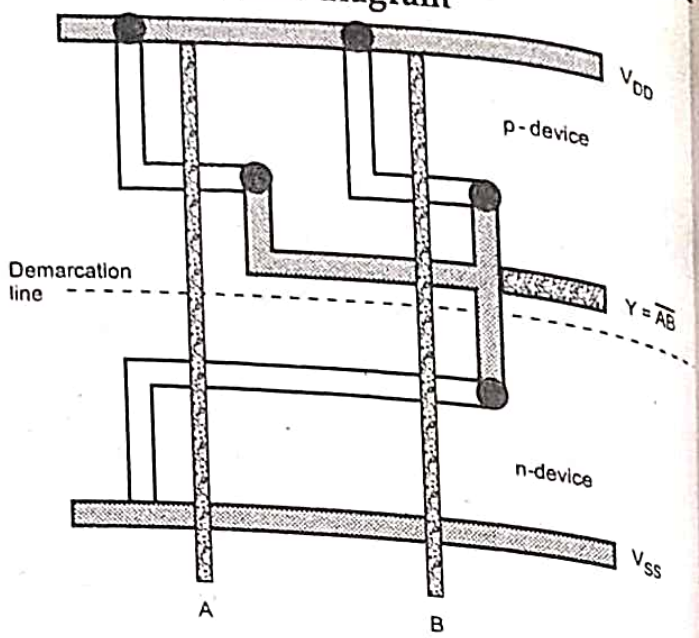


Fig. 6.3.8

Layout diagram

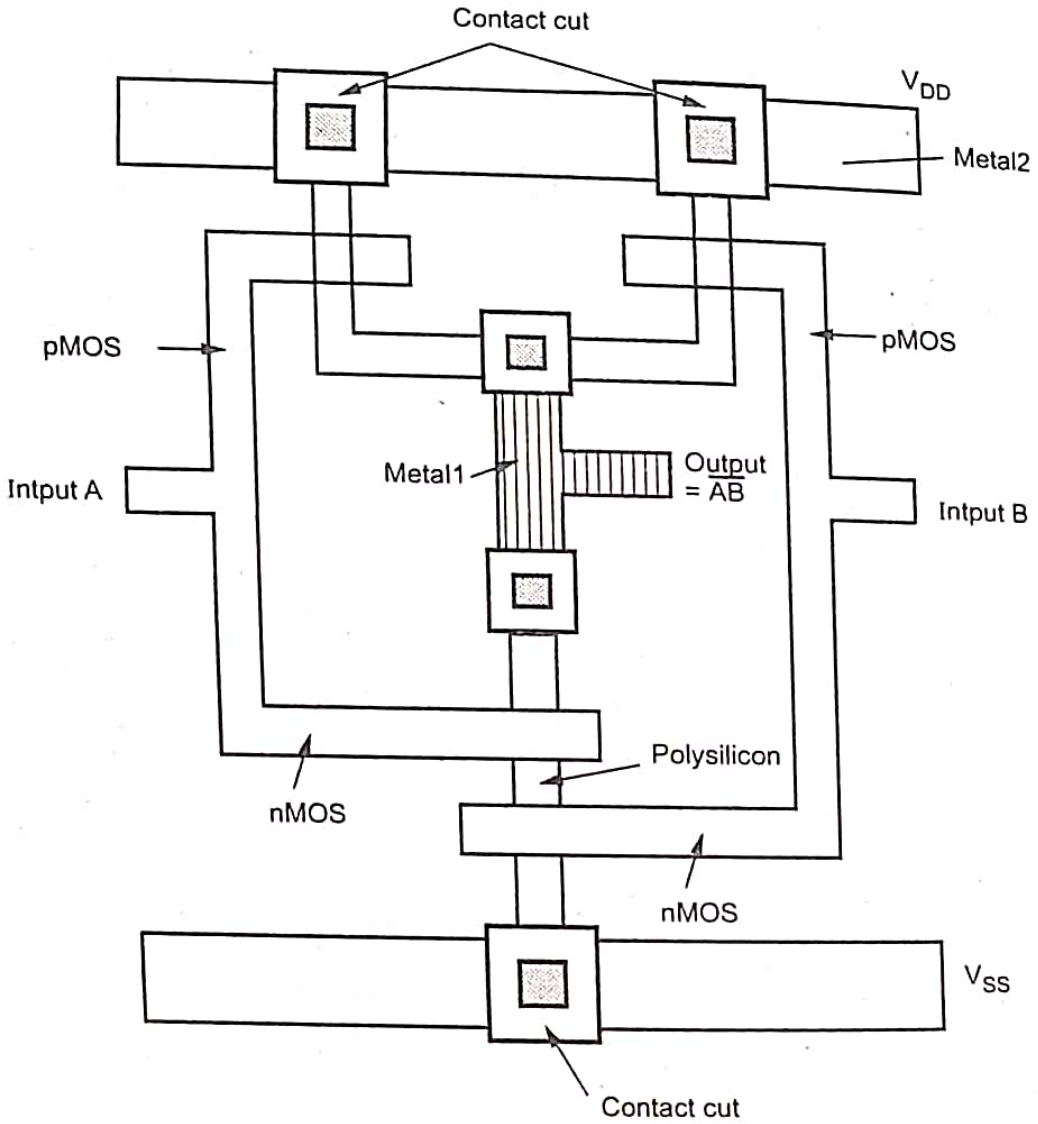


Fig. 6.3.9 Layout diagram for CMOS NAND

University Question

1. *Draw circuit of CMOS two inputs NOR gate. Derive V_{TH} of the same.*

GTU : Winter-21, Marks 7

6.4 Complex Logic Circuit

GTU : Summer-16, 17, 21, Winter-21

- The compound gate comprises of combination of series and parallel switch structures. A function $Y=\overline{(A\cdot B)+(C\cdot D)}$ is a good example of compound gate. The function is called AND-OR-INVERT-22 (AOI 22) because it performs the NOR of a pair of two input ANDs. Fig. 6.4.1 shows implementation of function $Y=\overline{(A\cdot B)+(C\cdot D)}$.

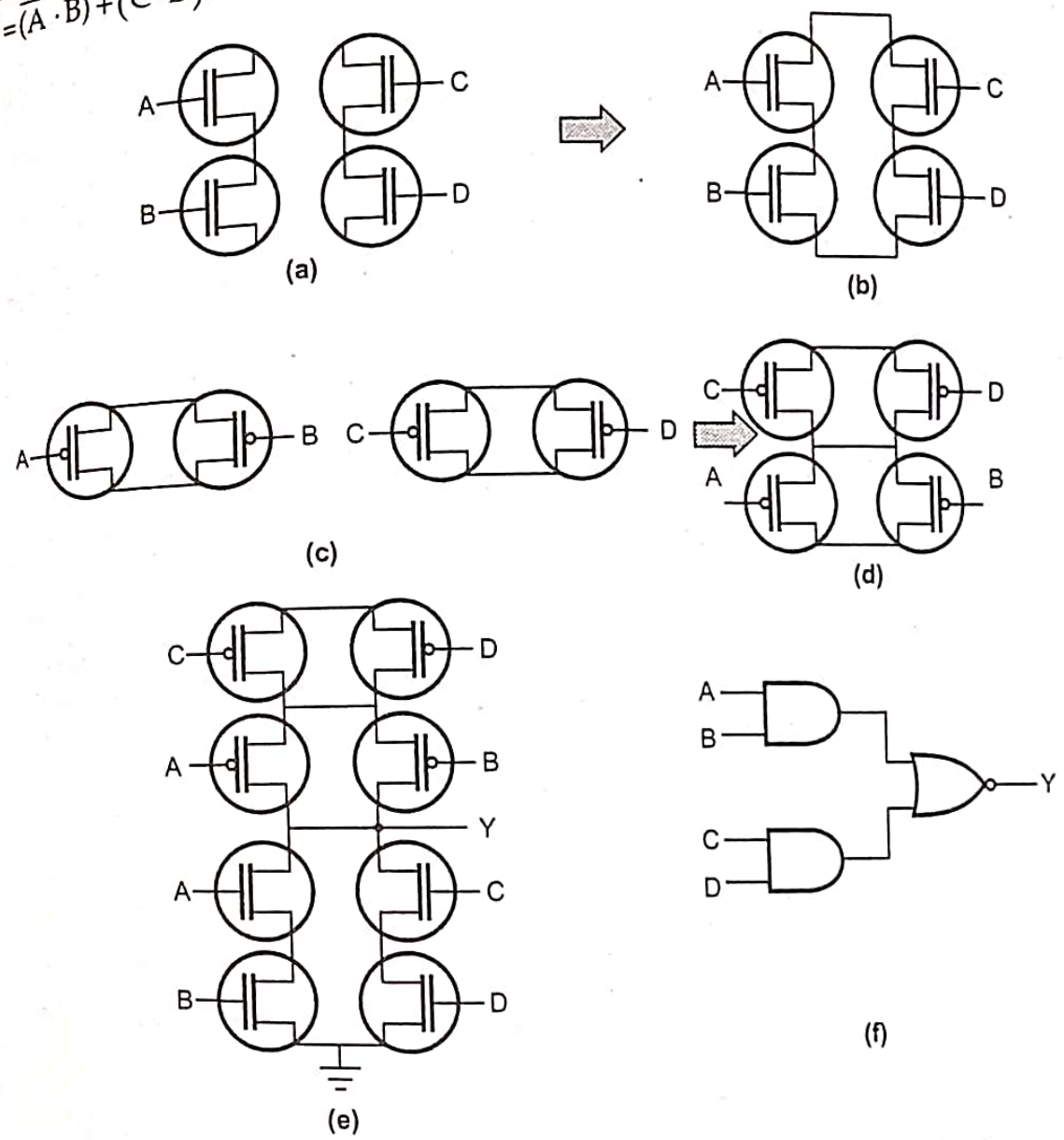


Fig. 6.4.1 CMOS compound gate of function $Y=\overline{(A\cdot B)+(C\cdot D)}$

Solved Examples

Example 6.4.1 *Realize the Boolean function by using nMOS complex logic gate.* $Z=\overline{A(D+E)+BC}$.

GTU : Summer-21, Marks 3

Solution : Realization for $Z = \overline{A(D+E)+BC}$

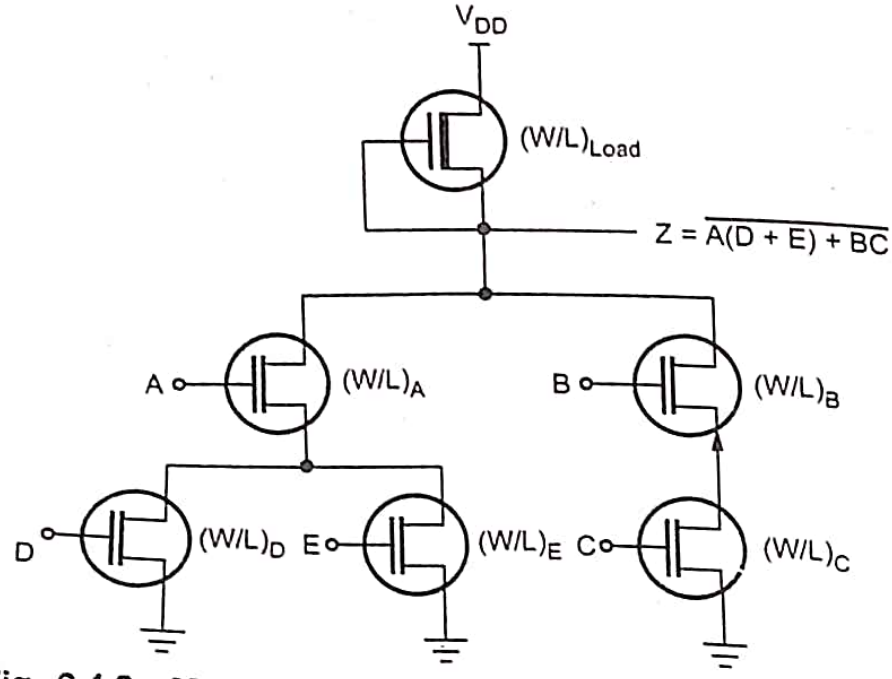


Fig. 6.4.2 nMOS complex logic gate for $Z = \overline{A(D+E)+BC}$

Example 6.4.2 *Realize the Boolean function by using CMOS logic gate.*

$Z = \overline{A(D+E)+BC}$

Solution :

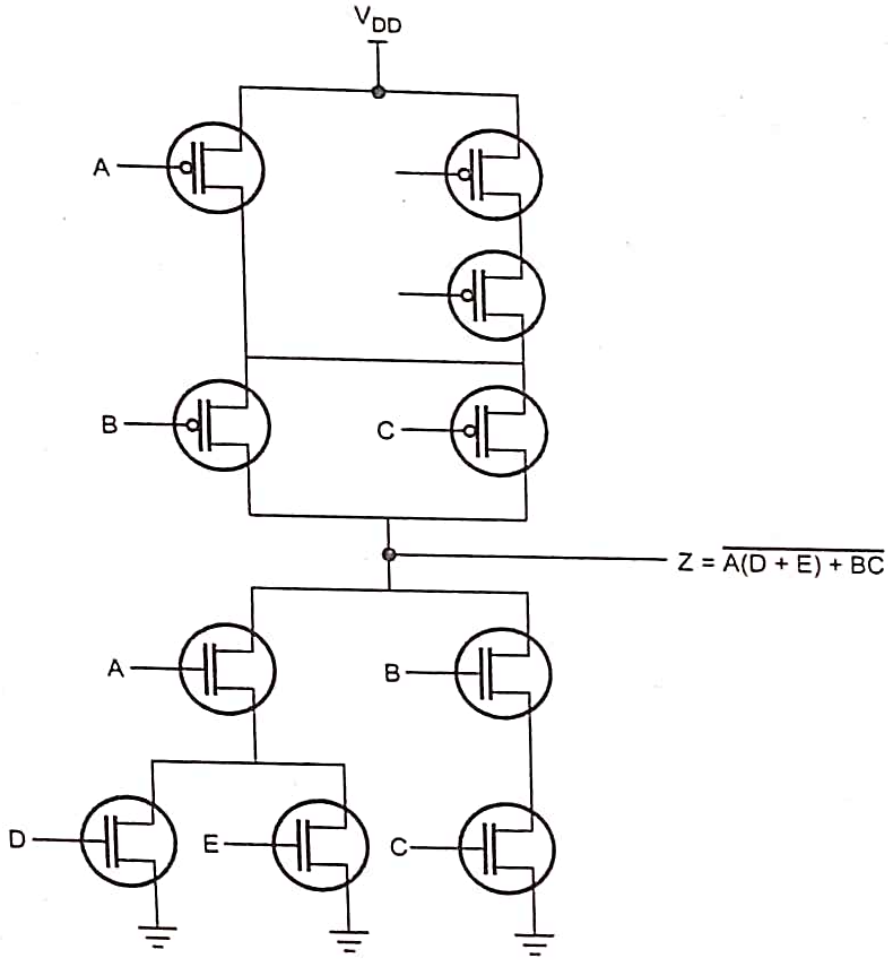


Fig. 6.4.3

Example 6.4.3 *Implement the following Boolean function using CMOS $F = [(C+D+E)(B+A)]'$ find a equivalent CMOS inverter circuit simultaneous switching of all inputs. Assume $(W/L)p = 15$ for all pmos transistor and $(W/L)n = 10$ for all nmos transistor.*

GTU : Summer-16, Marks 7, Winter-21, Marks 4

Solution :

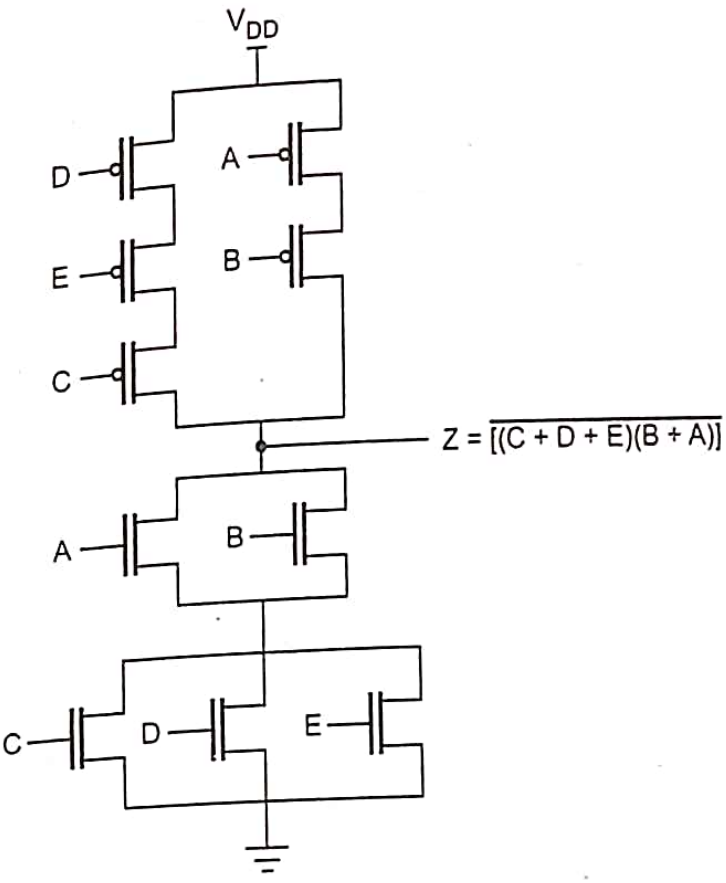


Fig. 6.4.4

$$\left(\frac{W}{L}\right)_{n,eq} = \frac{1}{\dfrac{1}{\left(\frac{W}{L}\right)_C\left(\frac{W}{L}\right)_D\left(\frac{W}{L}\right)_E} + \dfrac{1}{\left(\frac{W}{L}\right)_A\left(\frac{W}{L}\right)_B}} = \frac{1}{\frac{1}{30}+\frac{1}{20}} = 12$$

$$\left(\frac{W}{L}\right)_{P,eq} = \frac{1}{\dfrac{1}{\left(\frac{W}{L}\right)_C} + \dfrac{1}{\left(\frac{W}{L}\right)_D} + \dfrac{1}{\left(\frac{W}{L}\right)_E}} + \frac{1}{\dfrac{1}{\left(\frac{W}{L}\right)_A\left(\frac{W}{L}\right)_B}}$$

$$= \frac{1}{\frac{1}{15}+\frac{1}{15}+\frac{1}{15}} + \frac{1}{\frac{1}{15}+\frac{1}{15}} = 12.5$$

Example 6.4.4 *Draw the stick diagram for CMOS inv*

Solution : Stick diagram of CMOS inverter.

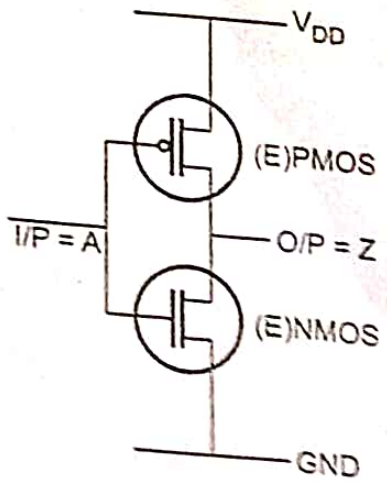


Fig. 6.4.5

6.5 Transmission Gates (TGs)

GTU : Summer 15, 16, 17, 21, Winter-21

- The strength of signal is measured by how closely it approximates ideal voltage source. A strong signal can sink or source more current. The strongest source of '1's and '0's are power supplies or rails (V_{DD} and GND).
- An nMOS transistor passes a strong '0' since it is almost perfect switch when passing a '0'. But it is imperfect at passing a '1'. A pMOS transistor has opposite behaviour i.e. passing a strong '1's and degraded '0's. Fig. 6.5.1 shows transistor symbols and their behaviour.

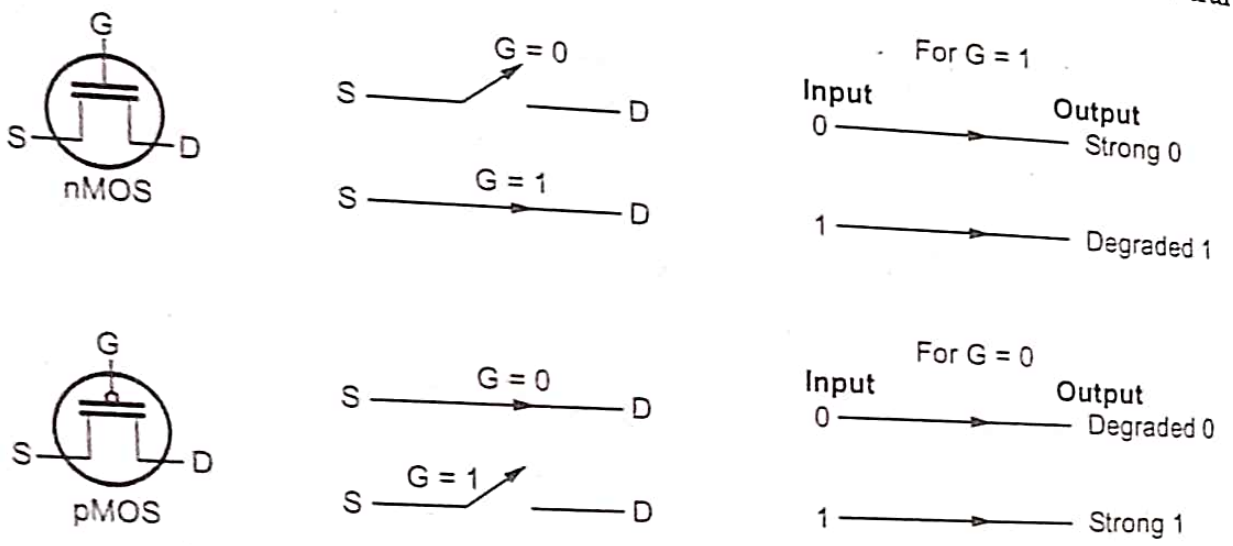


Fig. 6.5.1 Pass transistor strong and degraded outputs

- A pMOS or nMOS transistor alone is (an imperfect switch) called as **pass transistor.** A parallel combination of nMOS and pMOS transistor is called as **transmission gate** or **pass gate.** The transmission gate turns on when a '1' is applied to G in which '0's and '1's are both passed in an acceptable form.

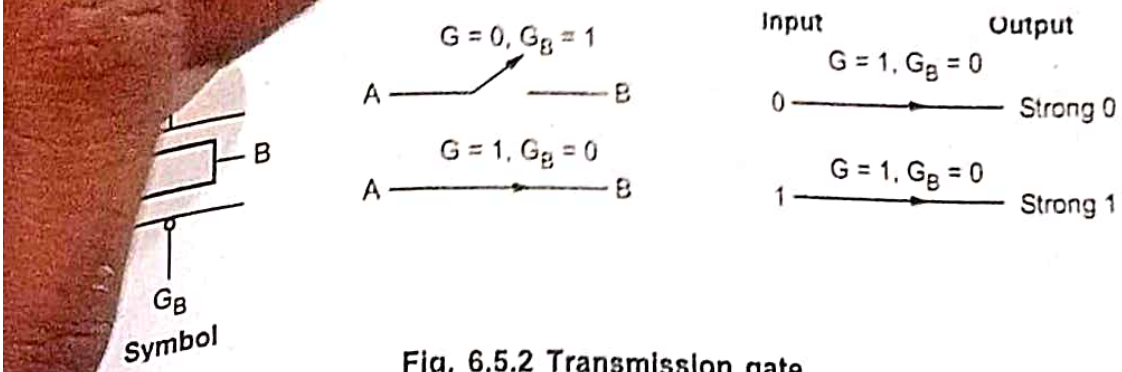


Fig. 6.5.2 Transmission gate

Transmission gate using parallel combination of pMOS and nMOS transistors is shown in Fig. 6.5.3.

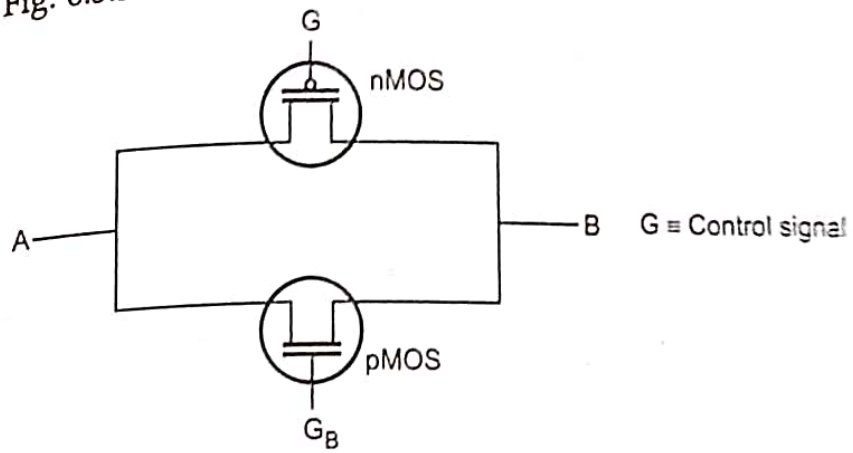


Fig. 6.5.3

- The control signal (G) is applied to nMOS transistor and its complement (G_B) is applied to pMOS transistor. The transmission gate is used as bidirectional switch controlled by control signal (G). When control signal (G) is high (G_B = low), both pMOS and nMOS transistors are ON and provides a low resistance path between terminals A and B vice versa. CMOS transmission gates are compact circuits which requires minimum transistors.
- Another popularly accepted representation of CMOS Transmission Gate (TG) are shown in Fig. 6.5.4.

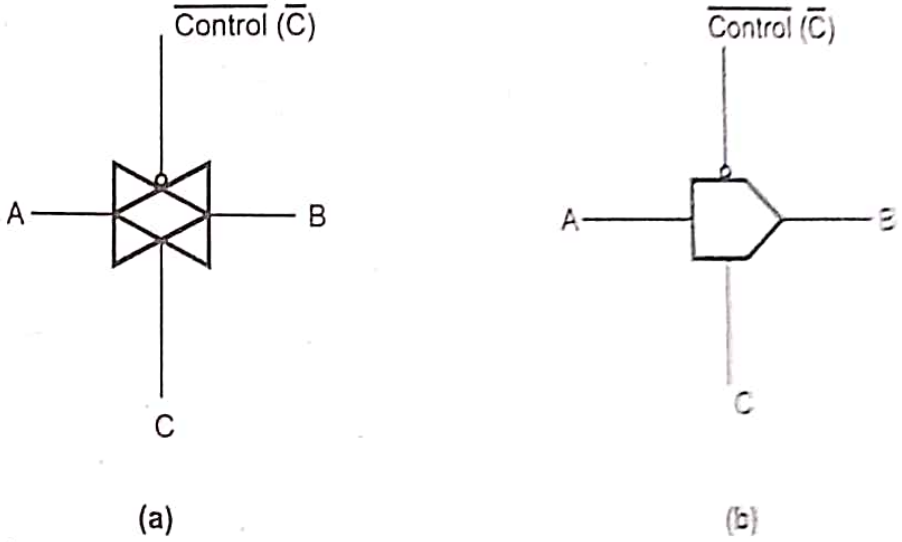


Fig. 6.5.4 Different representations of CMOS Transmission Gate (TG)

6.5.1 Applications of TG

- TG are used for following implementations.
  1. OR operation 2. XOR operation 3. 2-input multiplexer.

1. OR operation

- OR gate using TG is shown in Fig. 6.5.5.

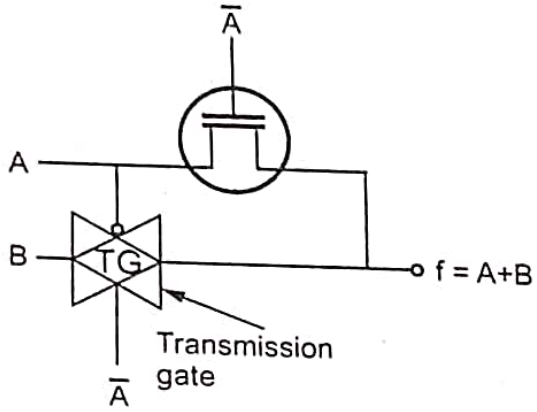


Fig. 6.5.5 OR operation

2. XOR operation

- XOR operation using transmission gate is shown in Fig. 6.5.6.

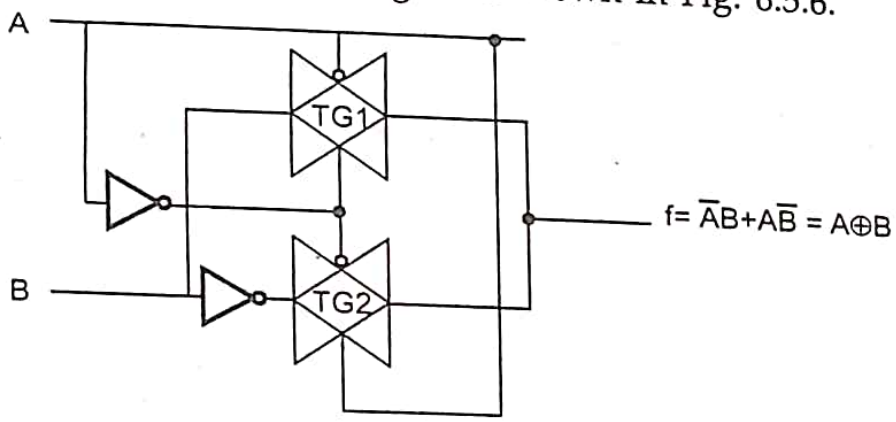


Fig. 6.5.6 XOR operation

3. Two input multiplexer

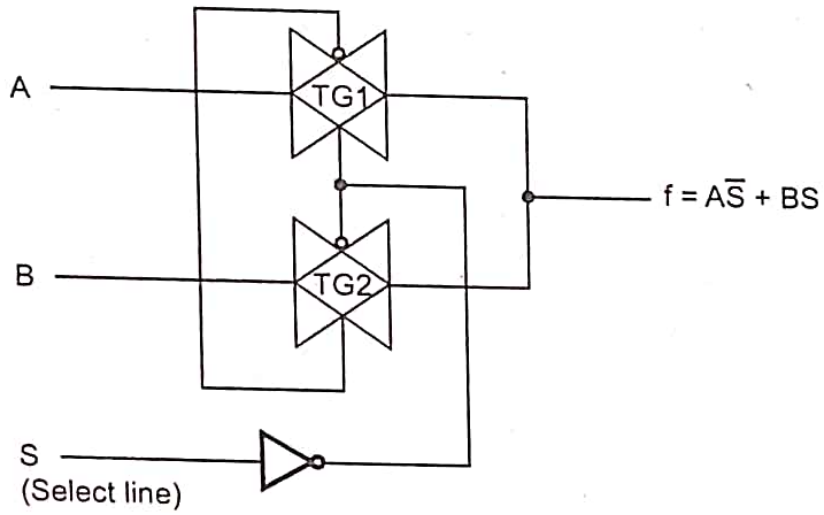


Fig. 6.5.7 Two input MUX

6.5.2 Transmission Gate Adder

Transmission gate XOR

- An adder can be implemented by using transmission XOR gate. Fig. 6.5.8 shows XOR gate.

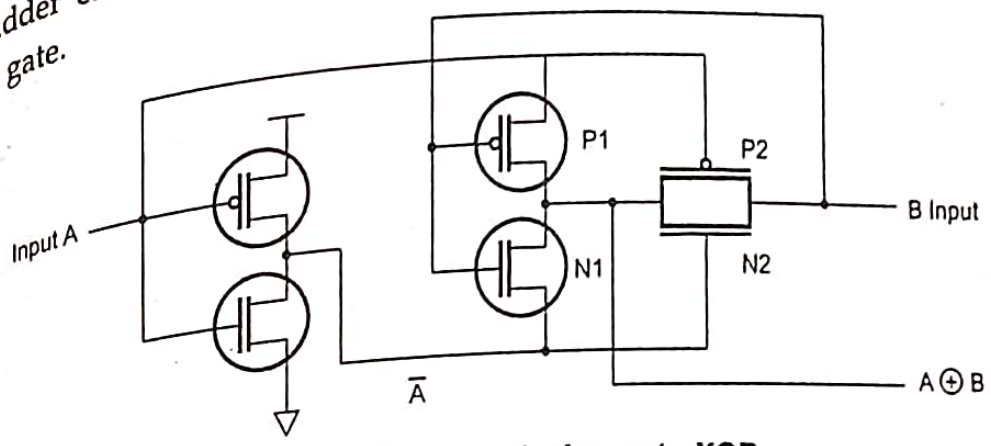


Fig. 6.5.8 Transmission gate XOR

- When A is high ($\overline{A}$ is low), transistors P_1 and N_1 acts as an inverter, the output is $\overline{B}$. Transmission gate formed by transistor pair P_2 and N_2 is open.
- When A is low ($\overline{A}$ is high), the transmission gate P_2, N_2 is closed passing B to the output.

Transmission gate adder

- An adder can be constructed by using four transmission gates, four inverters and two XOR gates as shown in Fig. 6.5.9.

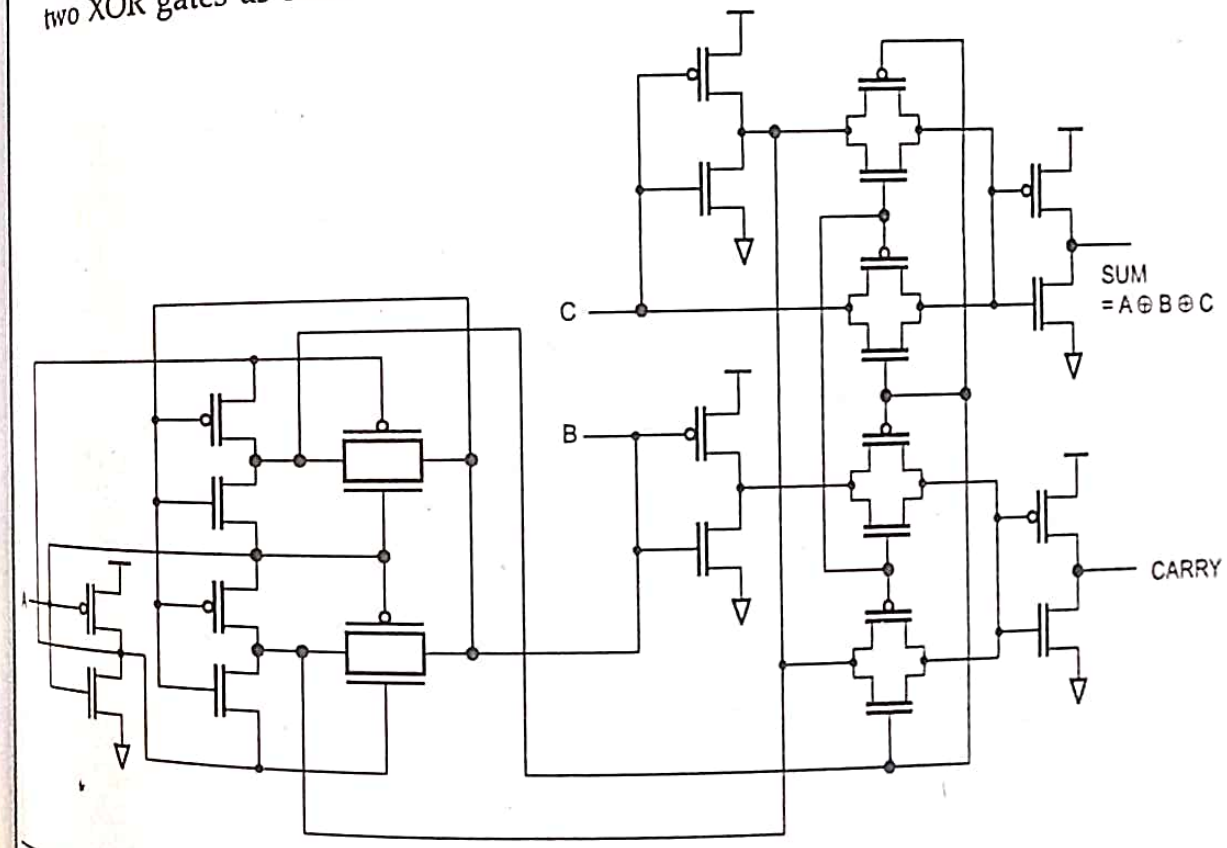


Fig. 6.5.9 Transmission gate adder

- The SUM is formed by multiplexer controlled by $A \oplus B$ (and complement). Carry = C, when $A \oplus B$ is true and carry = A (or B) when $A \oplus B$ is false. This implementation has equal SUM and carry delay times and are non inverted. It uses 24 transistors.
- The number of transistors can be reduced when speed is not the ultimate goal. The two transistors may be replaced by inverter on the output of XOR gate also with some optimization output buffers may be eliminated. Fig. 6.5.10 shows TG adder with optimization.

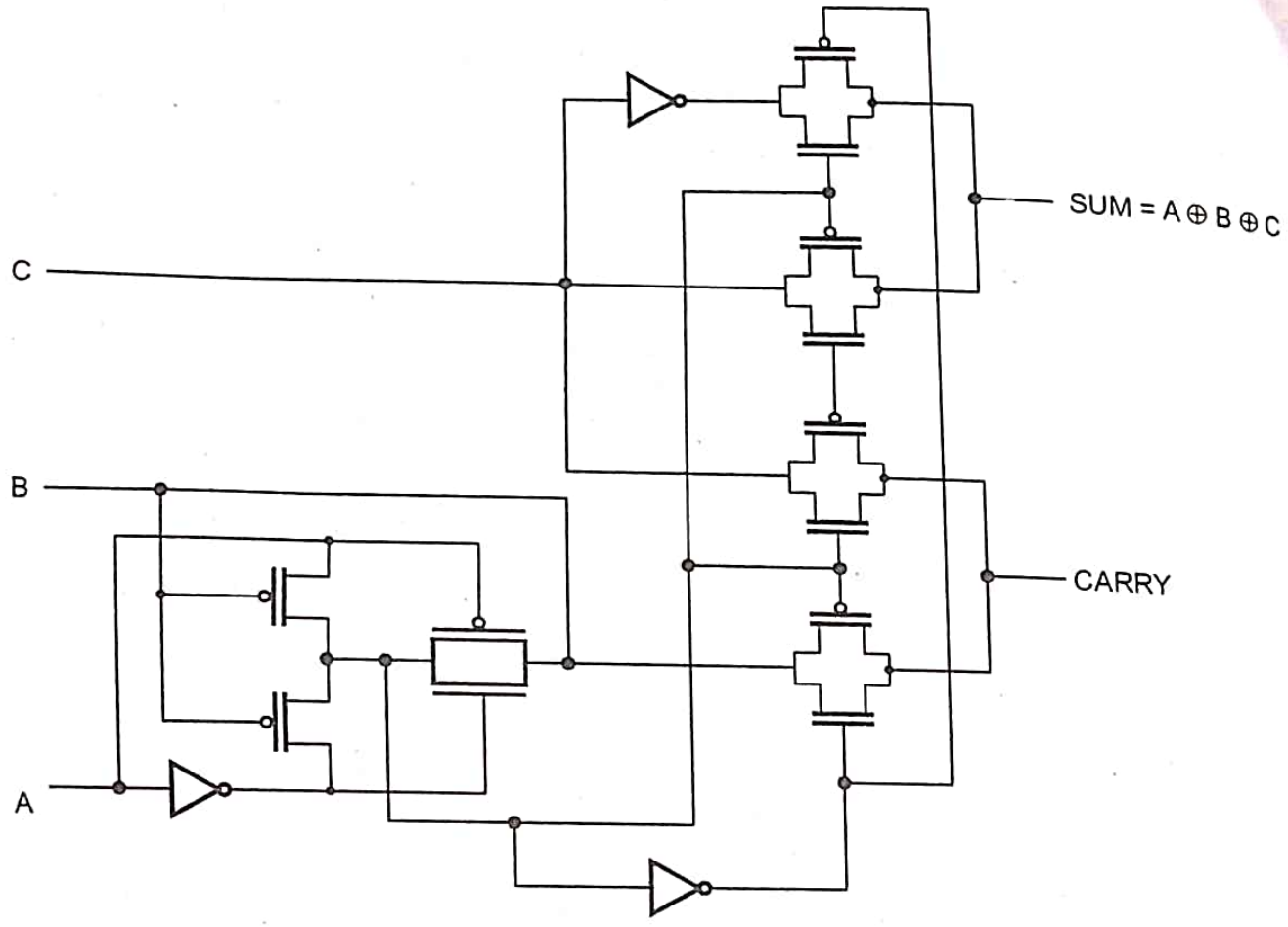


Fig. 6.5.10 TG adder with optimization

- Since the propagation times of SUM and CARRY are similar, these adders are used in multiplier arrays.

6.5.3 Design of Multiplexers

- Multiplexers are important components in CMOS memory elements and data manipulation structures. A multiplexer selects one of the inputs to the output line depending on select lines.

- Let in a tw
$\bar{S}$. The inp
when select
The logic exp
Truth table fo

and D_1 and select lines are S and
nd $\bar{S}$ = 0 and input D_1 is selected

in Table 6.5.1.

D_1	D_0	S	$\bar{S}$	Output Y
X	0	0	1	0
X	1	0	1	1
0	X	1	0	0
1	X	1	0	1

Table 6.5.1 Two input MUX truth table

- Two input multiplexer can be formed by connecting two transmission gates together as shown in Fig. 6.5.11. Any one of transmission gate is enabled depending on select line inputs.

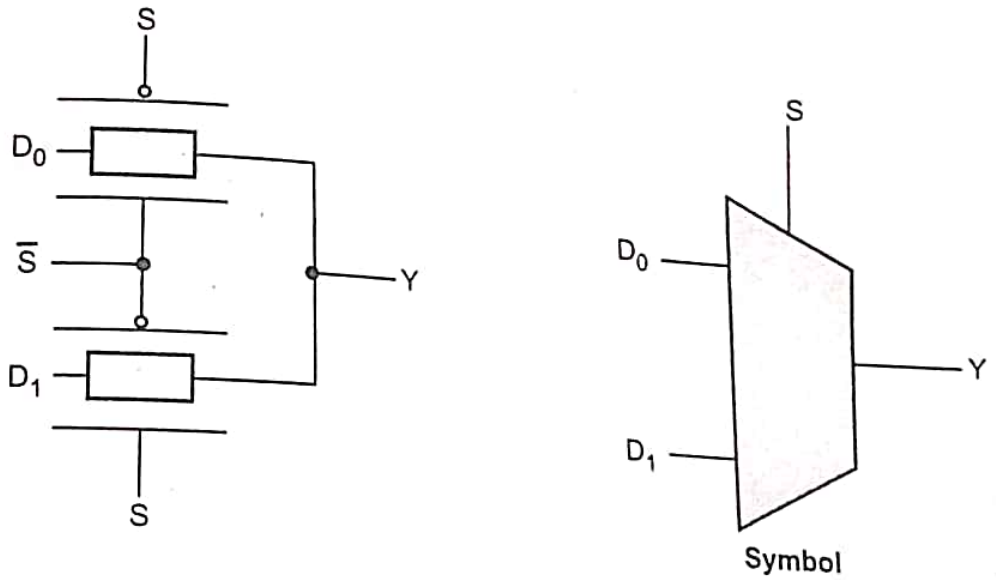


Fig. 6.5.11 Transmission gate multiplexer and symbol

- The transmission gates make multiplexer non-restoring. A restoring, inverting multiplexer can be built. One method is by using compound gate as shown in Fig. 6.5.12.

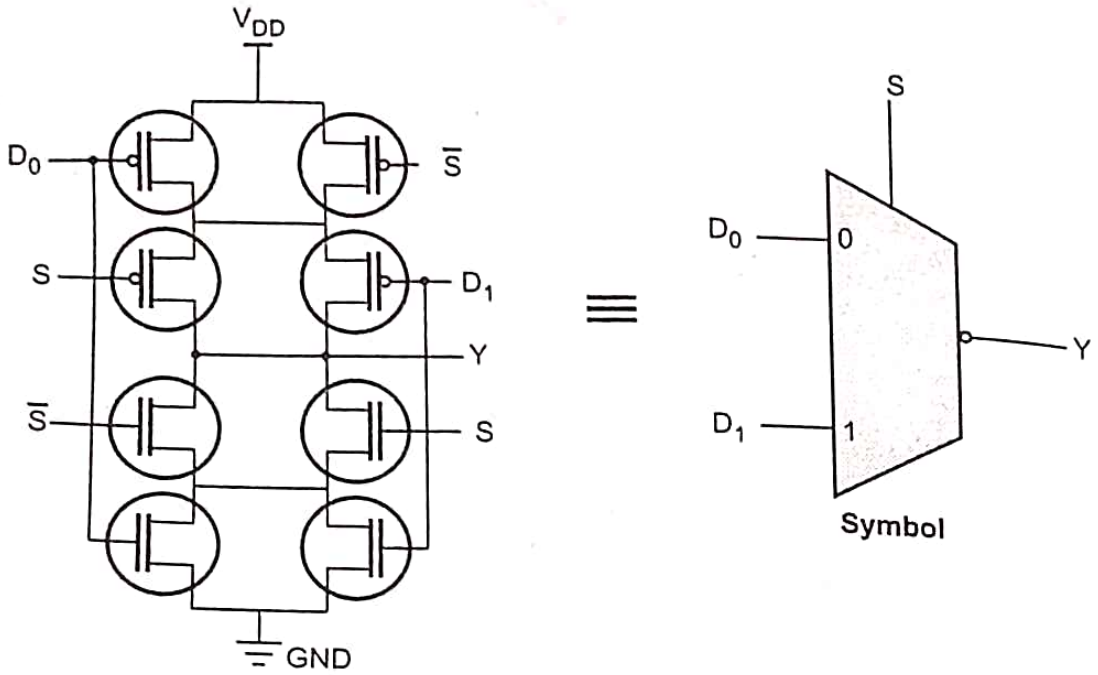


Fig. 6.5.12 Inverting MUX using compound gate

- The other method is to gang together two tristate inverters as shown in Fig. 6.5.13.

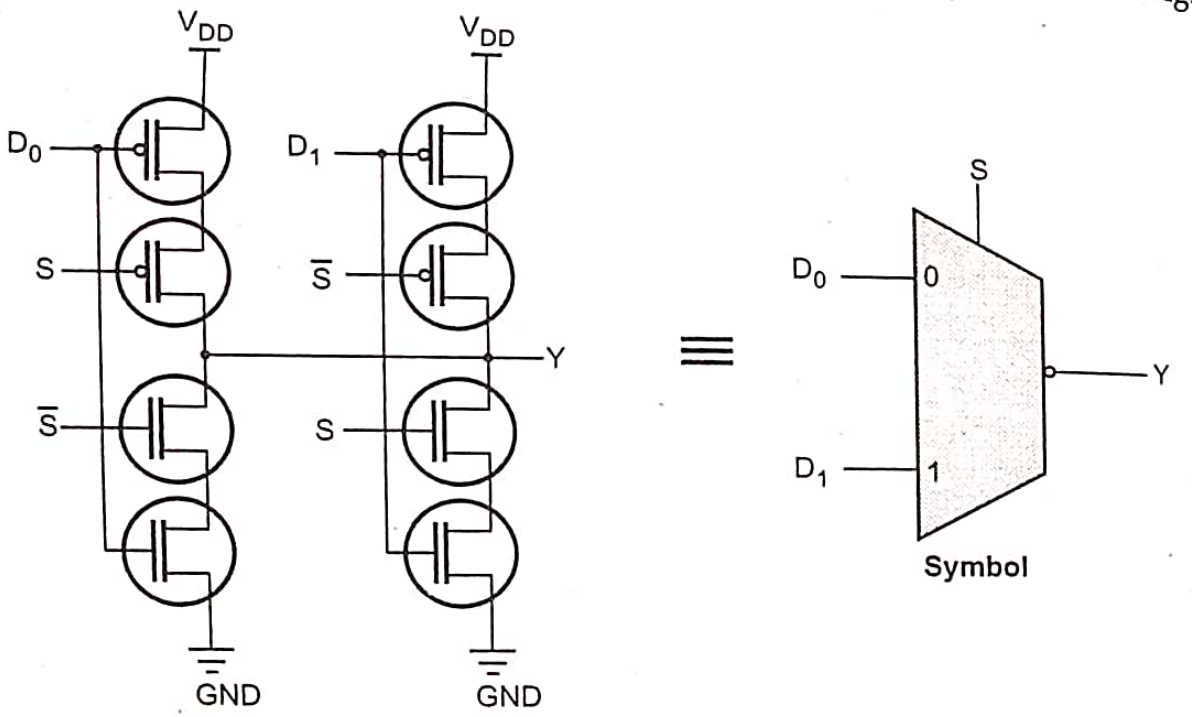


Fig. 6.5.13 Inverting MUX using tristate inverters

- The tristate inverter method is more compact and faster as it requires less internal wire. In the equivalent symbol of MUX the complementary select is generated within cell, hence it is omitted from the symbol.
- Larger multiplexers can be built by ganging together several tristates. A 4:1 MUX using tristates is shown in Fig. 6.5.14 (see Fig. 6.5.14 on next page), where decoded select signals for each tristate is needed.

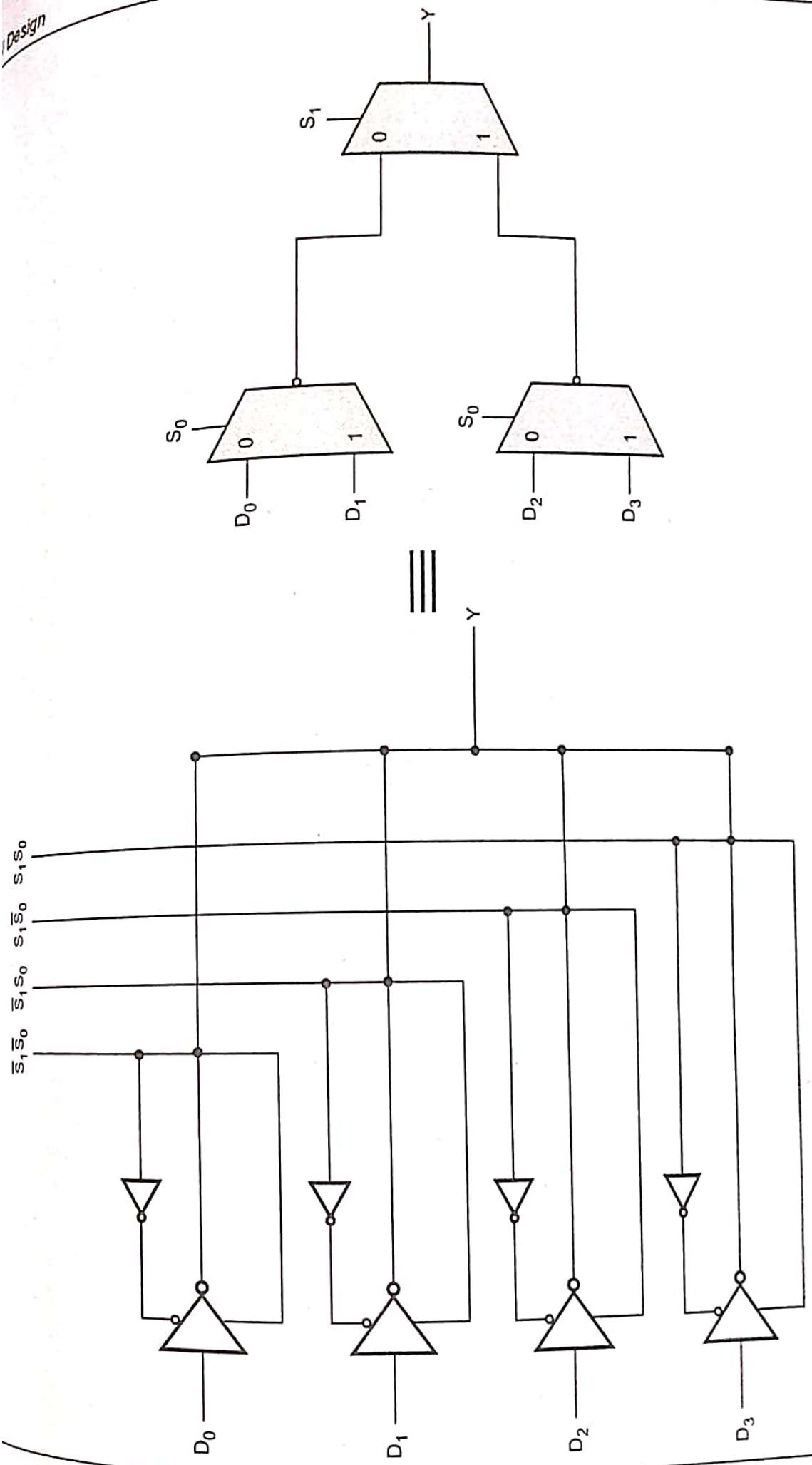


Fig. 6.5.14 4:1 MUX

4:1 MUX

- MUX using transmission gates, compare this schematic with conv

$$Y = AS_0S_1 + B\overline{S}_0S_1 + CS_0\overline{S}_1 + D\overline{S_0S_1}$$

$$Y = \left(AS_0 + B\overline{S}_0\right)S_1 + \left(CS_0 + D\overline{S}_0\right)\overline{S}_1$$

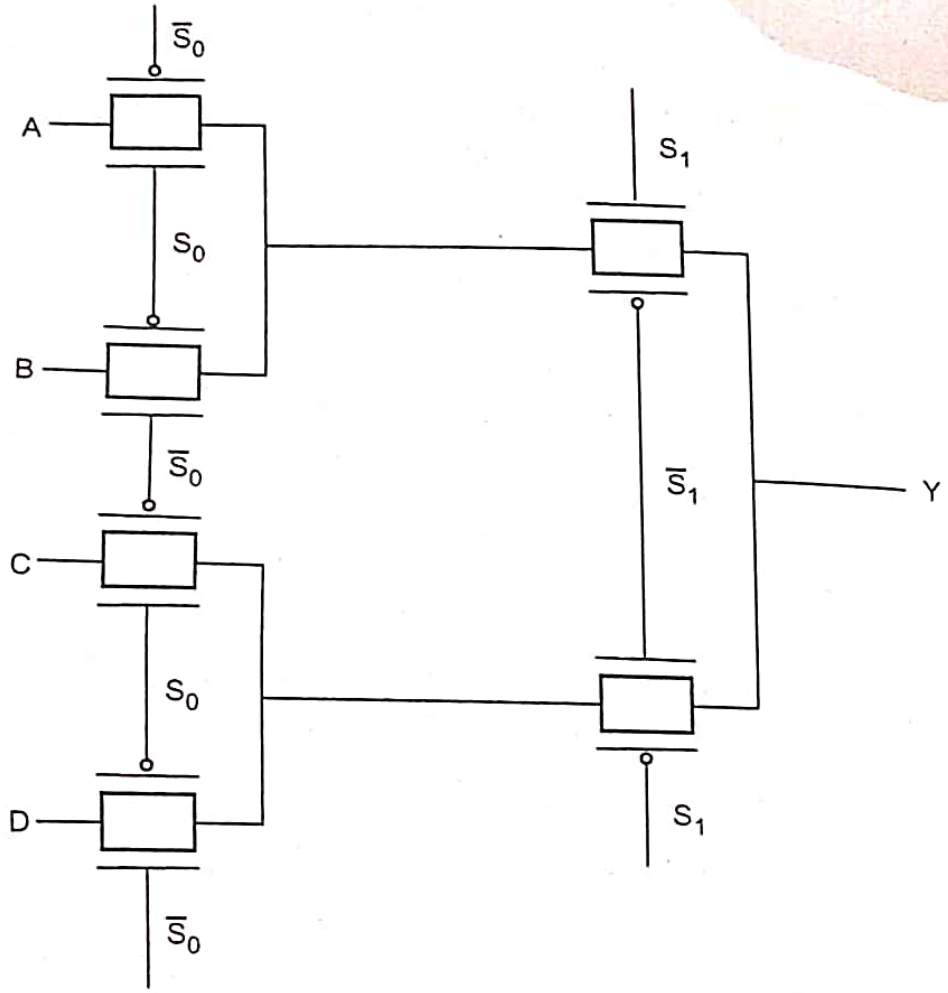


Fig. 6.5.15

8:1 MUX :

8:1 MUX using transmission gates is shown in the Fig. 6.5.16.

Refer Fig. 6.5.16 on next page.

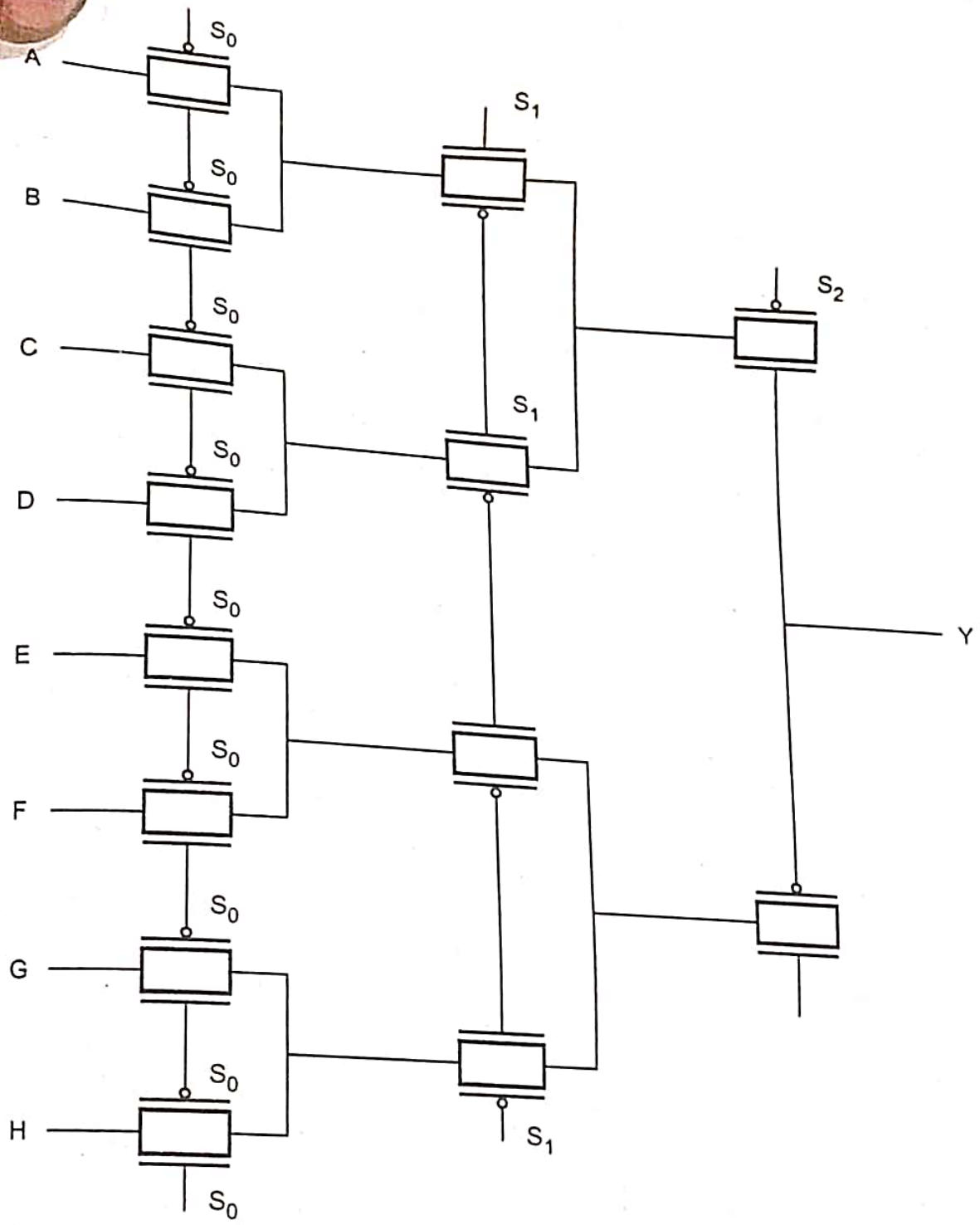
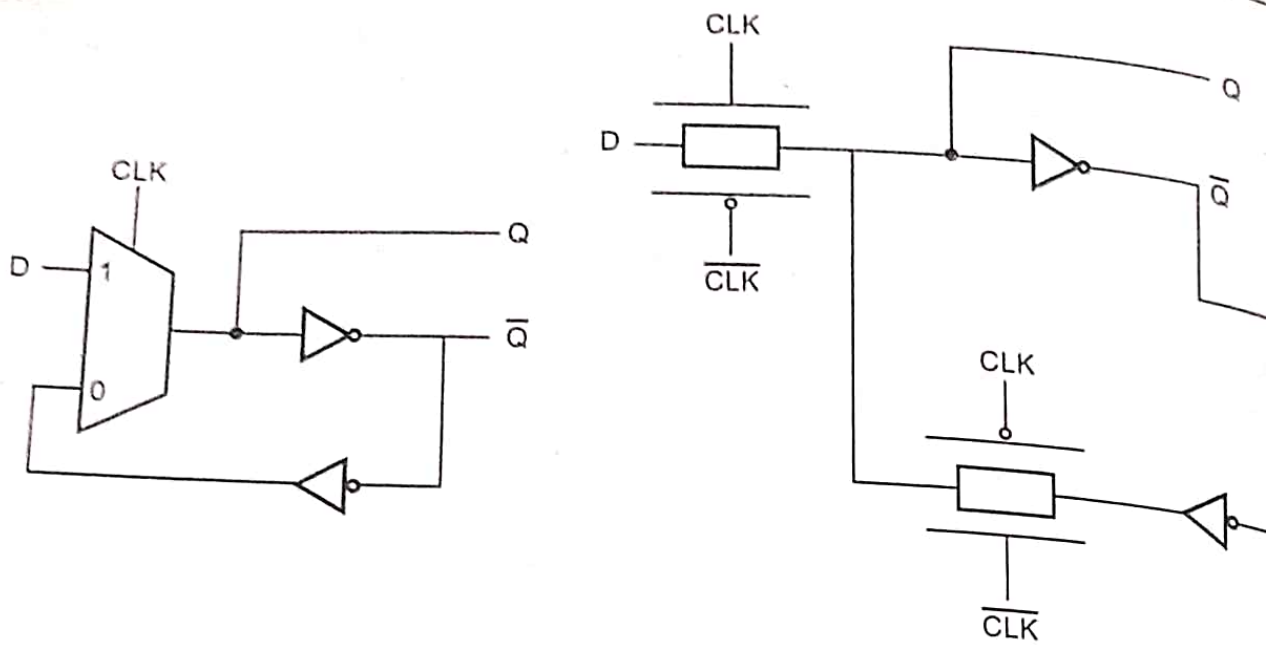


Fig. 6.5.16

6.5.4 Design of Latches and Flip-Flops

- A D-latch using one two inputs multiplexer and two inverter is shown in Fig. 6.5.17. It has a data input (D), a clock input (CLK) and true and complemented outputs (Q and $\overline{Q}$).



(a) D-latch using MUX

(b) D-latch using transmission gate

Fig. 6.5.17 D-latch using MUX and transmission gate

- When CLK = 1, the latch is transparent, its equivalent circuit is sho Fig. 6.5.18. The output Q = D and $\overline{Q} = \overline{D}$.

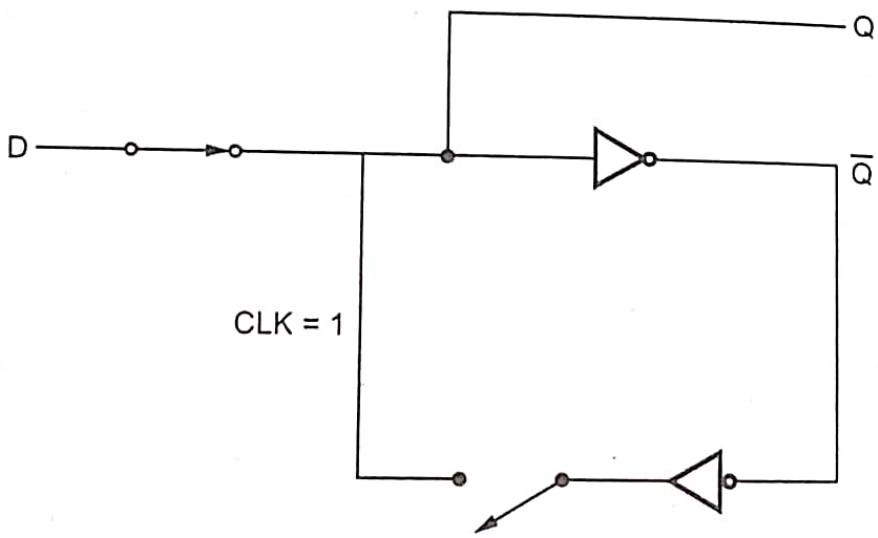


Fig. 6.5.18 Transparent D-latch

- When CLK = 0, the latch becomes opaque. A feedback path around the in pair is established to hold the current state of Q. When latch is opaque input is ignored. Fig. 6.5.19 shows the opaque state of D-latch.

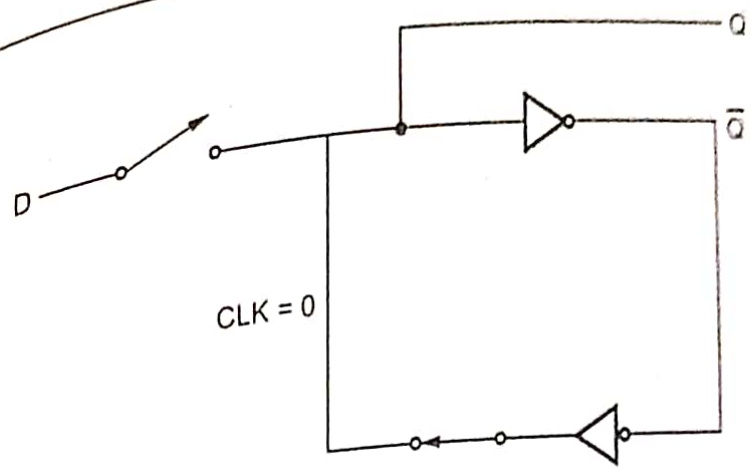


Fig. 6.5.19 Opaque D-latch

- A D-latch is level-sensitive latch because the state of the output is dependant on the level of the clock signal as shown in Fig. 6.5.20.

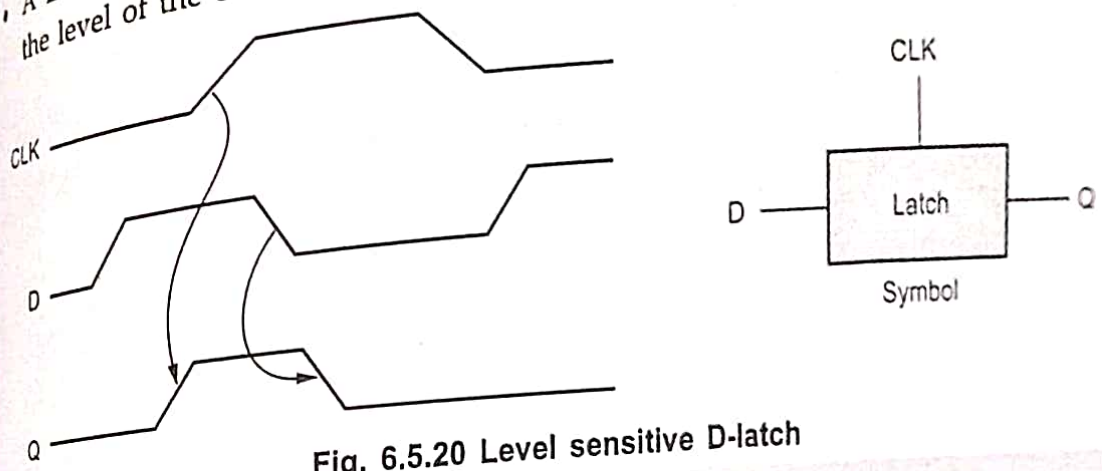


Fig. 6.5.20 Level sensitive D-latch

Example 6.5.1 *Realize functions using TG : $F = AB + A'C' + AB'C$.* **GTU : Summer-16, Mark 1**

Solution : $F = AB + A'C' + AB'C$

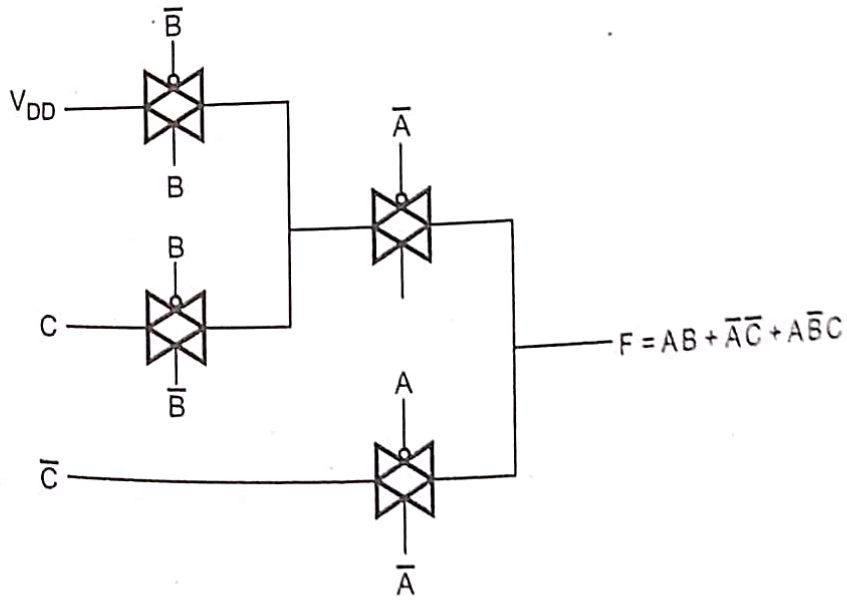


Fig. 6.5.21

Example 6.5.2 *Realize the following Boolean function using CMOS Transmission Gates.*

$F = AB + A'C' + AB'C.$

GTU : Summer-17, Mark

Solution :

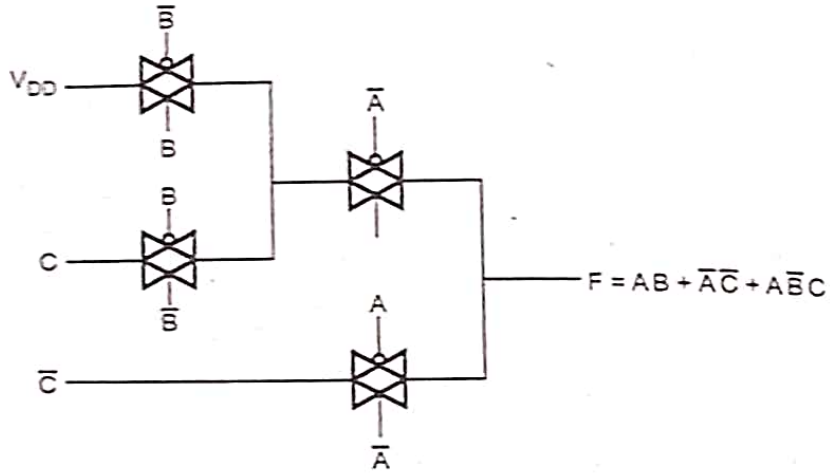


Fig. 6.5.22

Example 6.5.3 *Implement the following Boolean function using CMOS.*

$F = [(C + D + E).(B + A)]'$

GTU : Summer-17, Mark

Solution :

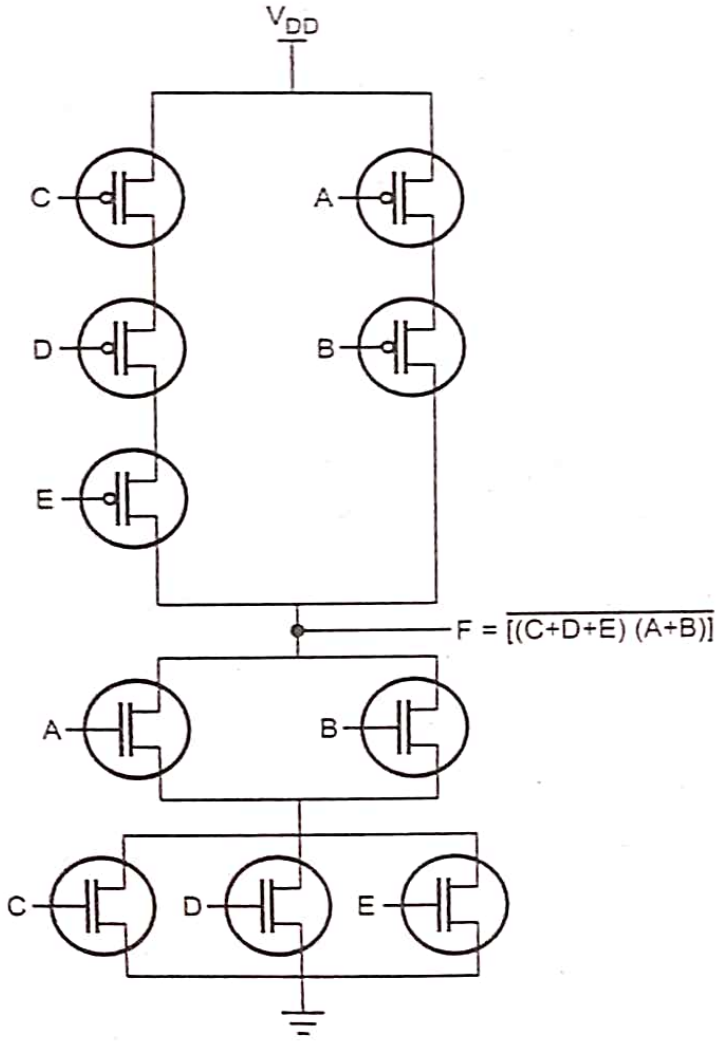
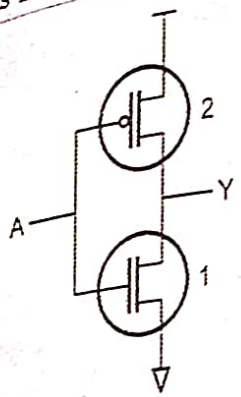
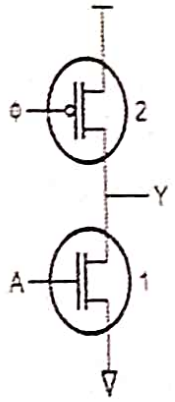


Fig. 6.5.23

6.5 Comparison of Static CMOS and Dynamic CMOS Logic

S.N.	Static CMOS Logic	Dynamic CMOS Logic
1.	Static CMOS circuits use complementary nMOS pulldown and pMOS pullup networks to implement logic gates or logic functions in integrated circuits.	Dynamic gates use a clocked pMOS pullup. The implemented logic function or the logic gate is achieved through 2 modes of operation: Precharge and Evaluate.
2.	 Static	 Dynamic
3.	Power consumption is more.	Less power consumption campared to static CMOS logic
4.	Lower switching speed hence speed is slower.	Higher switching speed hance faster

University Questions

1. *Derive R_{eq} of all regions in CMOS transmission gates and draw a graph of equivalent R plotted as a function of output voltage.* **GTU : Summer 15, Marks 7**
2. *Implement 2-to-1 MUX and XOR functions using CMOS Transmission Gates (TGs).* **GTU : Summer-21, Marks 3**
3. *Why does NMOS pass strong logic - 0 and weak logic - 1 ?* **GTU : Summer-21, Marks 7**
4. *Implement following boolean logic equation using Transmission Gate (TG).* $Y = AB + A'C' + AB'C$. **GTU : Winter-21, Marks 4**
5. *Write a short note on CMOS transmission gate.* **GTU : Winter-21, Marks 7**

6.6 Short Questions and Answers

Q.1 What are pull-ups and write about the resistor pull-up and its usage.

OR What is pull up and pull down device ?

Ans. : Pull up devices :

- Pull-up and Pull-down resistors are used to correctly bias the inputs of digital gates to stop them from floating about randomly when there is no input condition.

- Pull-up resistors are resistors used in logic circuits to ensure a well-define logical level at a pin under all conditions.
- Digital logic circuits have three logic states : high, low and floating (or hig impedance). The high-impedance state occurs when the pin is not pulled to high or low logic level, but is left "floating" instead.
- A good illustration of this is an unconnected input pin of a microcontroller. It i neither in a high or low logic state, and a microcontroller might unpredictabl interpret the input value as either a logical high or logical low.
- Pull-up resistors are used to solve the dilemma for the microcontroller by pullin the value to a logical high state. If there weren't for the pull-up resistor, th MCU's input would be floating when the switch is open and brought down onl when the switch is closed.
- Pull-up resistors are not a special kind of resistors; they are simple fixed-valu resistors connected between the voltage supply (usually + 5 V) and th appropriate pin, which results in defining the input or output voltage in th absence of a driving signal. A typical pull-up resistor value is 4.7 kΩ, but ca vary depending on the application.

Applications for pull-up and pull-down resistors

- Pull-up and pull-down resistors are often used when interfacing a switch or som other input with a microcontroller or other digital gates. Most microcontroller have in-built programmable pull up/down resistors so fewer external component are needed. It is possible to interface a switch with such microcontrollers directly.
- Pull-up resistors are in general used more often than pull-down resistors although some microcontroller families have both pull-up and pull-down available.
- They are often used in analog to digital converters to provide a controlled current flow into a resistive sensor.
- Another application is the I2C protocol bus, where pull-up resistors are used to enable a single pin to act as an input or an output. When not connected to a bus, the pin floats in a high-impedance state. Pull-down resistors are also used on outputs to provide a known output impedance.

Q.2 What are the uses of stick diagram ?

Ans. : Uses of stick diagram :

1. Stick diagrams can be drawn much easier and faster than a complex layout.
2. Stick diagrams are especially important tools for layout built from large cells.
3. Stick diagrams are a means of capturing topography and layer information using simple diagrams.

4. Stick diagrams convey layer information through colour codes (or monochrome encoding).
5. Stick diagram acts as an interface between symbolic circuit and the actual layout.

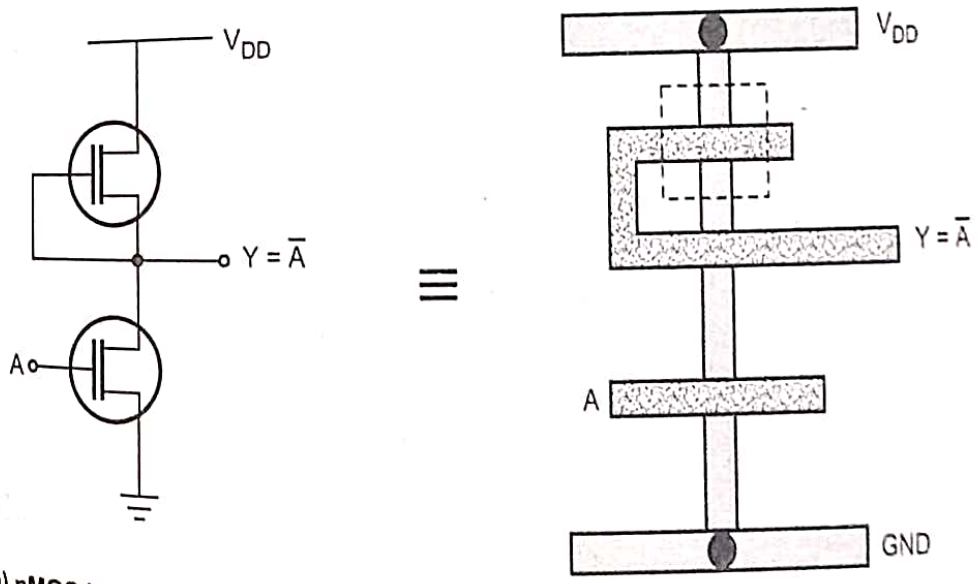
Q.3 Explain difference between stick diagram and layout diagram.

Ans. :

Sr. No.	Stick diagram	Layout diagram
1	A stick diagram is a graphical view of a layout of a chip. Stick diagram is a layout construct from shapes like rectangular or squares meanwhile stick diagrams represents relative positions of transistors with rectangular or mark (X) as component symbols and lines as wires.	The design of physical layout is very tightly linked to overall circuit performance (area, speed, power dissipation) since the physical structure directly determines the transconductances of the transistors, the parasitic capacitances and resistances, and obviously, the silicon area which is used for a certain function.
2	It is easy to draw because it does not need to be drawn to scale.	They may be hand-drawn on a 5 mm squared paper or prepared using CAD VLSI tools that support mask level design.

Q.4 Represent the stick diagram of a NMOS inverter.

Ans. :



(a) nMOS inverter circuit with depletion load

(b) Stick diagram of nMOS inverter

Fig. 6.6.1

Q.5 Give the various color coding used in stick diagram.

Ans. :

P diffusion	Yellow /Brown	Metal 1	Blue
N diffusion	Green	Metal 2	Magenta /Purple
Polysillicon	Red	Metal 3	Cyan/Light Blue
Taps and Contacts	Black		

Q.6 Draw the static CMOS logic circuit for the following expression - Y = (ABCD)′

Ans. :

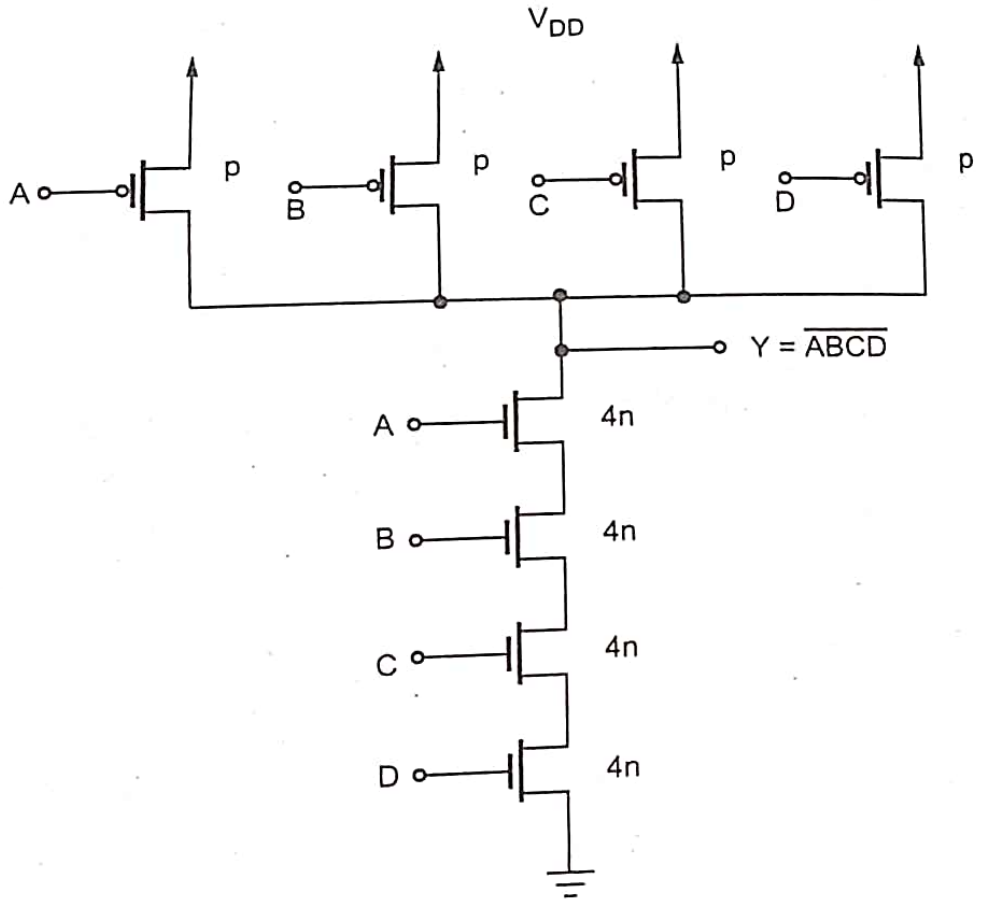


Fig. 6.6.2

Q.7 What is transmission gate ?

Ans. : Transmission gate -

- The circuit constructed with the parallel connection of PMOS and NMOS wi shorted drain and source terminals.
- The gate terminal uses two select signals s and s, when s is high then th transmission gates passes the signal on the input.
- The main advantage of transmission gate is that it eliminates the threshold voltag drop.

Q.8 Explain about switch logic and its usage.

Ans. : Switch logic :

- Switch logic is based on pass transistors or transmission gates. Pass transistor describes several logic families used in the design of integrated circuits.
- The switch logic reduces the count of transistors used to make different logic gates, by eliminating redundant transistors.
- The switch logic approach takes no static current from the supply rails and is faster for small arrays.
- Important features of switch logic approach are that -

  i) It occupies more area,

  ii) Eliminates undesirable threshold voltage and

  iii) Power dissipation is small in switch logic approach since current only flows on switching.

Q.9 How do understand about MOS transistor as a switch ?

Ans. : MOS transistor as a switch :

- MOS transistors can be used as ON/OFF switches. When the gate of an nMOS transistor is '1' the transistor is ON. When the gate of transistor is '0' the transistor is OFF.
- The switch level model with transistor symbol is shown in Fig. 6.6.3.

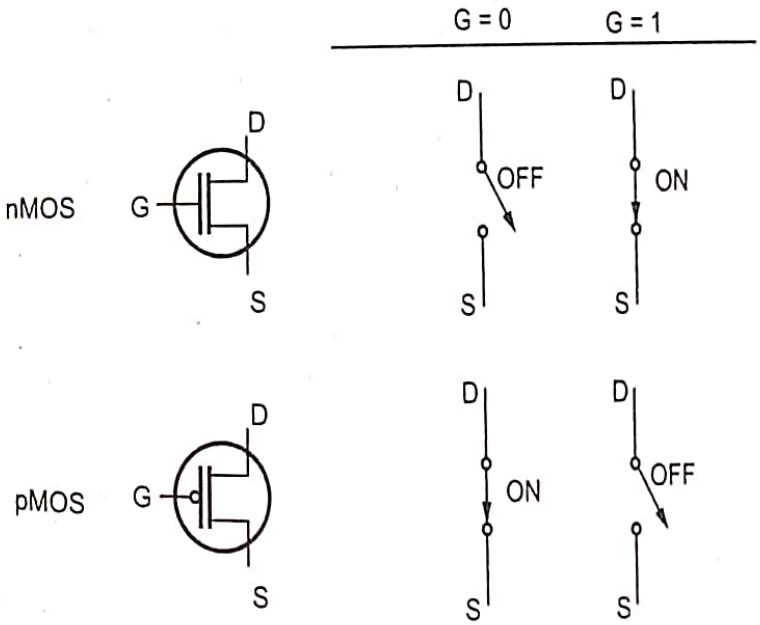


Fig. 6.6.3 : MOSFET as switches

□□□

Notes

7 Sequential MOS Logic Circuits

Syllabus

Introduction, Behavior of Bistable elements, The SR latch circuit, Clocked latch and Flip-flop circuit, CMOS D-latch and Edge-triggered flip-flop.

Contents

7.1 Introduction

7.2 Behavior of Bistable Elements

7.3 SR Latch Circuit ..... Summer-15, Winter-21 ··· Marks 7

7.4 D - Flip-Flop ..... Summer-21, Winter-21, ··· Marks 4

7.5 Multiplexer Based Latches

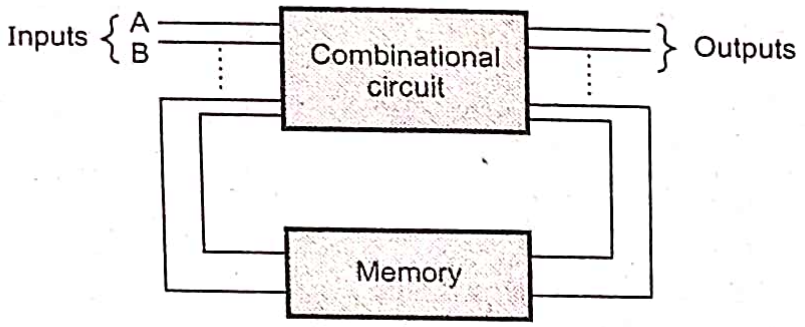
7.6 Master-Slave Edge Triggered Register ..... Winter-14, Summer-21, ··· Marks 7

7.7 Dynamic Latches and Registers..... Winter-21, ····· Marks 4

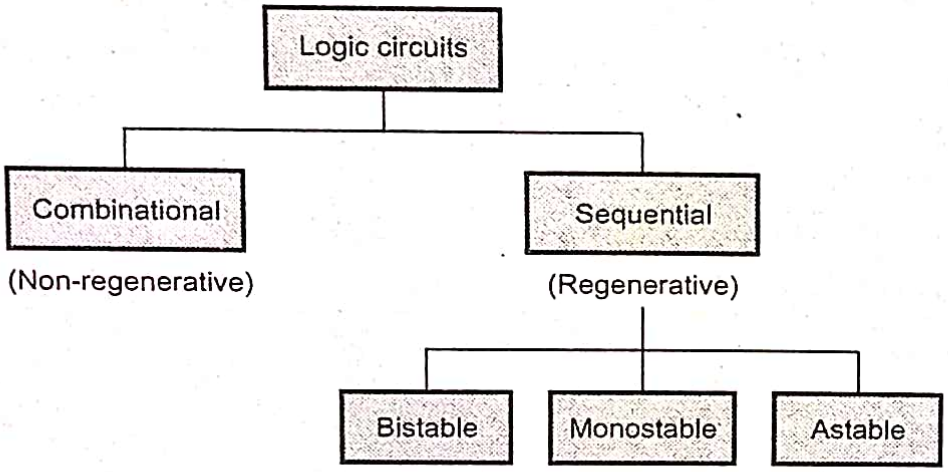
7.8 Short Question and Answer

7.1 Introduction

- The circuit having feedback relationship between output and input is call regenerative circuits. The sequential circuit consists of a combinational circuit a a memory block in feed loop. The critical components of sequential circuits a regenerative circuit which can be classified into three main groups; bistal circuits, monostable circuits and astable circuits Fig. 7.1.1 illustrates block diagra of sequential circuits and classification of logic circuit.



(a) Sequential circuits



(b) Classification of logic circuit

Fig. 7.1.1

7.1.1 Classification of Memory Elements

- Memory elements can be classified based on various features and properties, suc as - foreground and background memory, static and dynamic memory, latches an registers etc.

1. Foreground and background memory

- At a high level, memory is classified into background and foreground memory Memory embedded into logic is called foreground memory.
- The foreground memory is most often organized as individual registers of registe banks.

- Large amounts of centralized memory core are referred to as background memory.
- Background memory achieves higher area densities through efficient use of array structures and by trading off performance and robustness for size.

2. Static and dynamic memory

- Memories can also be categorised as static or dynamic. Static memories preserve the state as long as the power is turned on.
- Static memories are built using positive feedback or regeneration, where the circuit topology consists of specific connections between the output and the input of a combinational circuit.
- Static memories are most useful when the register is not updated for extended periods of time. An example of such is configuration data, loaded at power-up time. This condition also holds for most processors that use conditional clocking or gated clocks, where the clock is turned off for unused modules. In that case, there are no guarantees on how frequently the registers will be clocked and static memories are needed to preserve the state information.
- Memory built with positive feedback fall under the class of elements called multivibrator circuits. The bistable element is its most popular circuit, but other elements such as monostable and astable circuits are also sometimes used.
- Dynamic memories store its state for a short period of time (milliseconds). These are based on the principle of temporary charge storage on parasitic capacitors associated with MOS devices.
- In dynamic logic, the capacitors have to be refreshed periodically to overcome charge leakage. Dynamic memories are simpler, resulting in significantly higher performance and lower power dissipation.
- The dynamic memories are mostly used in datapath circuits that require high performance levels and are periodically clocked.
- It is possible to use dynamic circuitry even when circuits are conditionally clocked, if the state can be discarded when a module goes into idle mode.

3. Latches and registers

- The latches and flip-flops are two commonly used sequencing elements. Both latches and flip-flops are three terminal device i.e. data input (D), clock (Clk) and data output (Q).
- The latch is *transparent* when clock is high i.e. D flows through to Q as if the latch is buffer. The latch is *opaque* when clock is low i.e. latch holds its present Q output even if D changes.

- The flip-flop is an edge triggered device i.e. input data D is copied to Q on rising edge of clock and ignores for remaining time.

7.2 Behavior of Bistable Elements

- Static memories use positive feedback to create a bistable circuit - a circuit having two stable states that represent 0 and 1.

7.2.1 Bistability Principle

- A basic bistable element consisting of two cross-coupled inverters is shown in Fig. 7.2.1.

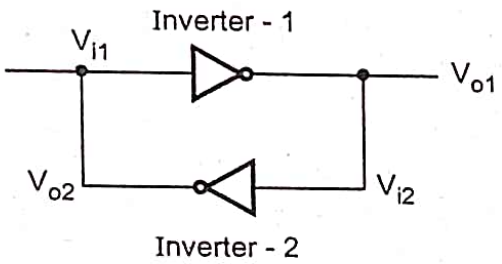


Fig. 7.2.1 Basic bistable element

- From above figure the conclusions can be drawn as : -
  i) $v_{o1} = v_{i2}$
  ii) $v_{o2} = v_{i1}$
- The voltage transfer characteristics of inverter showing three possible operating points is shown in Fig. 7.2.2.
- The two voltage transfer curves intersects at three points. If the circuit is operating at any stable point, it will remain in this state unless forced externally to change its operating point. To change its operating point a sufficiently large external voltage to make voltage gain of inverter loop greater than unity.

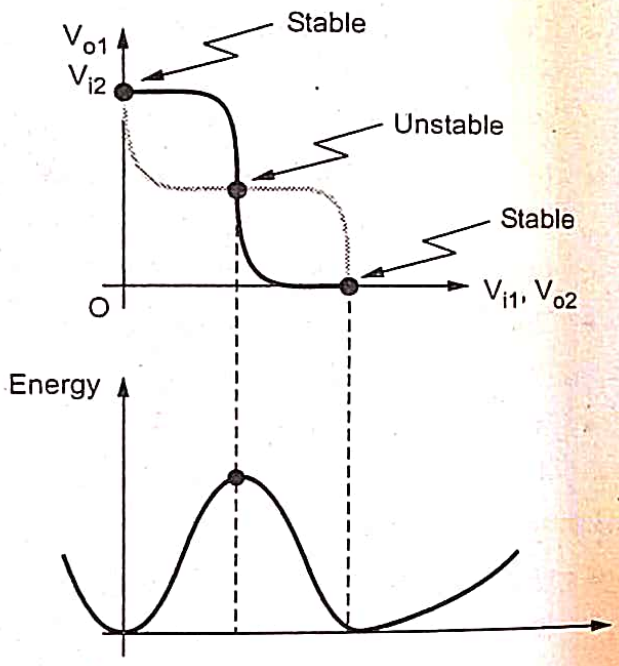


Fig. 7.2.2 Voltage transfer curves and potential energy levels

- If the circuit is biased at unstable point, a small input voltage to any of the inverter will be amplified and causing operating point to move to any of stable operating point. This concludes that the circuit has two stable operating points hence it is called bistable.

- The potential energy is minimum at both stable states as voltage gains of both inverters are equal to zero. The energy is maximum at operating point at which the operating points corresponding to two energy minimum and one stable operating corresponding to maximum potential energy.

7.2.2 CMOS Bistable Element

- CMOS two-inverter bistable element is shown in Fig. 7.2.3.
- Under unstable operating point, all four transistors are in saturation, resulting in maximum loop gain of the circuit. If the initial operating point is set at this point, a small voltage will cause changes in operating modes of transistors. Thus output voltages of two inverters to settle at V_{OH} and V_{OL} respectively. Fig. 7.2.4 shows one possibility for the expected time-domain behavior of the output voltages, if the circuit is initially set at its unstable operating point.

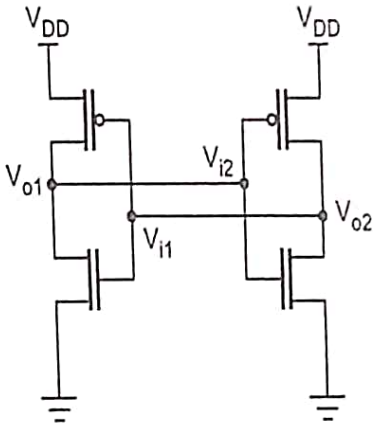


Fig. 7.2.3 CMOS bistable element

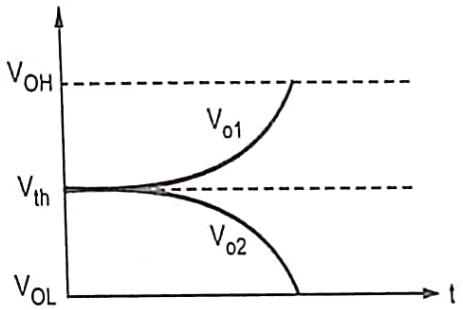


Fig. 7.2.4 Output voltages of two inverters

7.3 SR Latch Circuit

GTU : Summer-15, Winter-21

- The NAND-based SR Latch contains the basic memory cell (back-to-back inverters) built into two NAND gates to allow setting the state of the latch is shown in Fig. 7.3.1.
- **Operation of NAND-based SR Latch :** The circuit responds to active low S and R inputs : If S goes to 0 (while R = 1), Q goes high, pulling Q' low and the latch enters Set state. If R goes to 0 (while S = 1), Q' goes high, pulling Q low and the latch is Reset. Hold state requires both S and R to be high. S = R = 0 if not allowed, it would result in an indeterminate state.

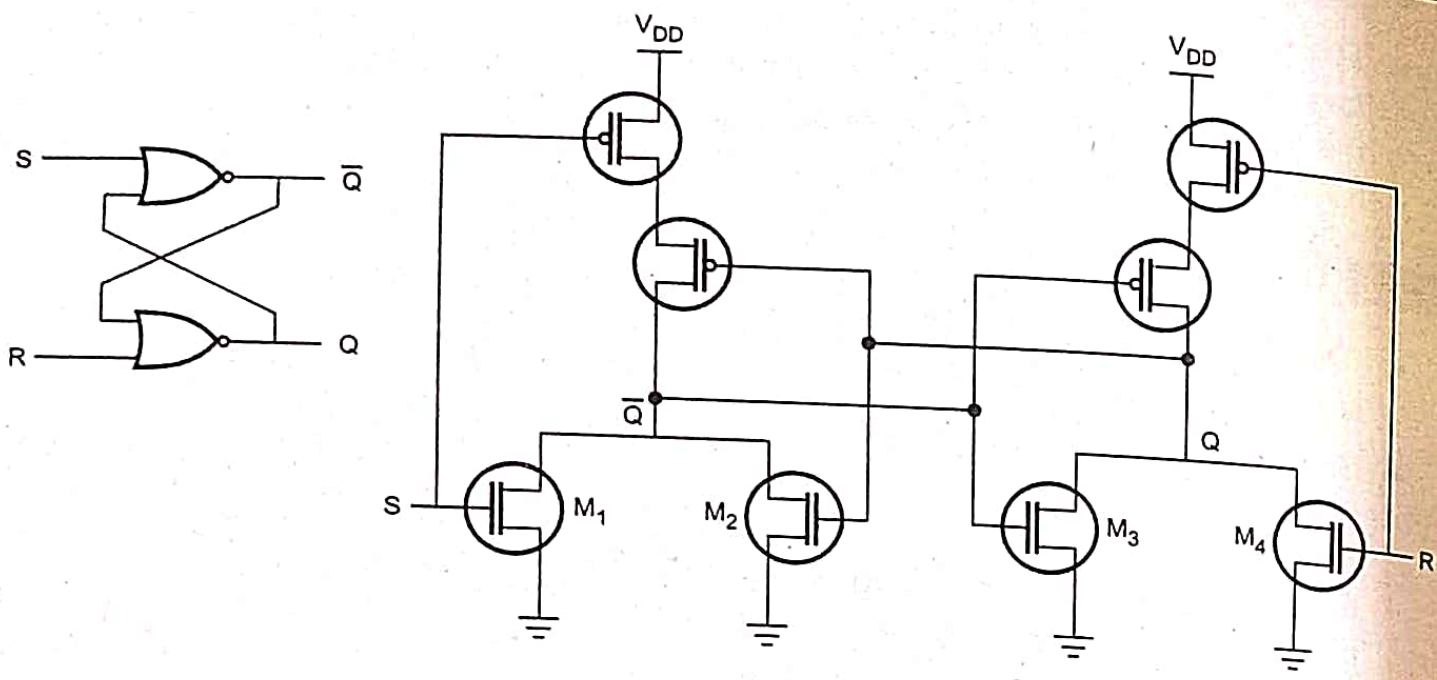


Fig. 7.3.1

Truth table of NAND-based CMOS SR latch circuit is shown in Table.

$\overline{S}$	$\overline{R}$	Q_{n+1}	$\overline{Q_{n+1}}$	Operation
0	0	1	1	not allowed
0	1	1	0	set
1	0	0	1	reset
1	1	Q_n	$\overline{Q_n}$	hold

University Questions

1. *Discuss CMOS SR latch circuits based on NOR2 gates.* **GTU : Summer-15, Marks 7**
2. *Draw NAND gate based CMOS SR latch circuit.* **GTU : Winter-21, Marks 4**

7.4 D - Flip-Flop

GTU : Summer-21, Winter-21

- D Flip-flop, constructed by cascading two D-Latch circuits is shown in Fig. 7.4.1. Master latch is positive level sensitive (receives data when CK = 1). Slave latch is negative level sensitive (receives data Q_m when CK = 0), the circuit is **negative-edge triggered.**
- Master latch receives input D until the CK falls from 1 to 0, at which point it sets that data in the master latch and sends it through to the output Q_s.

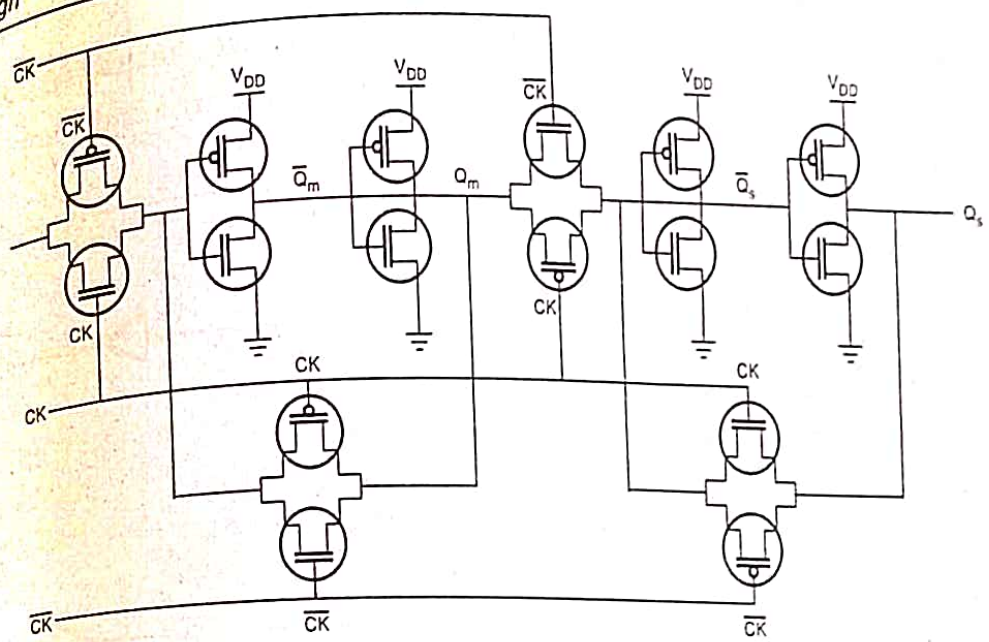


Fig. 7.4.1

University Questions

1. *Sketch and describe D-Latch circuit using CMOS-Transmission Gate(TG).*

GTU : Summer-21, Marks 4

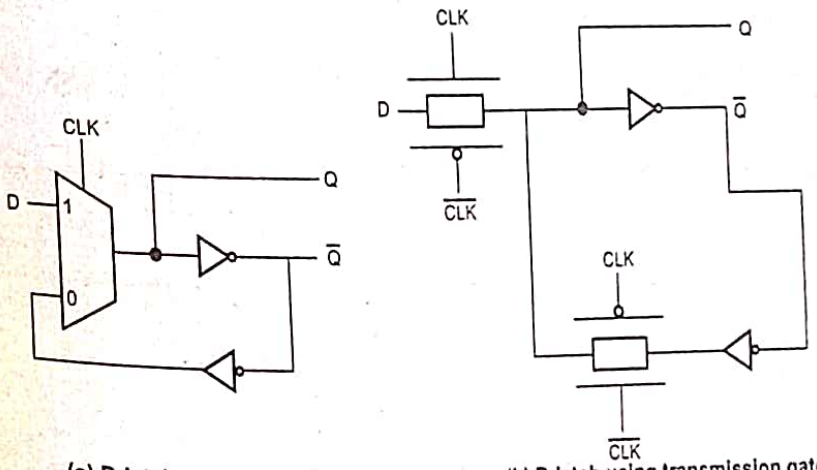
2. *Draw CMOS implementation of D latch with two inverters and two CMOS TG gates.*

GTU : Winter-21, Marks 4

7.5 Multiplexer Based Latches

- There are many approaches for constructing latches. One very common technique involves the use of transmission gate multiplexers.
- Multiplexer based latches can provide similar functionality to the SR latch, but has the important added advantage that the sizing of devices only affects performance and is not critical to the functionality.

Design of latches and flip-flops



(a) D-latch using MUX (b) D-latch using transmission gate

Fig. 7.5.1 D-latch using MUX and transmission gate

- A D-latch using one two inputs multiplexer and two inverter is shown Fig. 7.5.1. It has a data input (D), a clock input (CLK) and true and complemente outputs (Q and $\overline{Q}$).
- When CLK = 1, the latch is transparent, its equivalent circuit is shown in Fig. 7.5.2. The output Q = D and $\overline{Q} = \overline{D}$.
- When CLK = 0, the latch becomes opaque. A feedback path around the inverter pair is established to hold the current state of Q. When latch is opaque the D input is ignored. Fig. 7.5.3 shows the opaque state of D-latch.
- A D - latch is level-sensitive latch because the state of the output is dependant on the level of the clock signal as shown in Fig. 7.5.4.

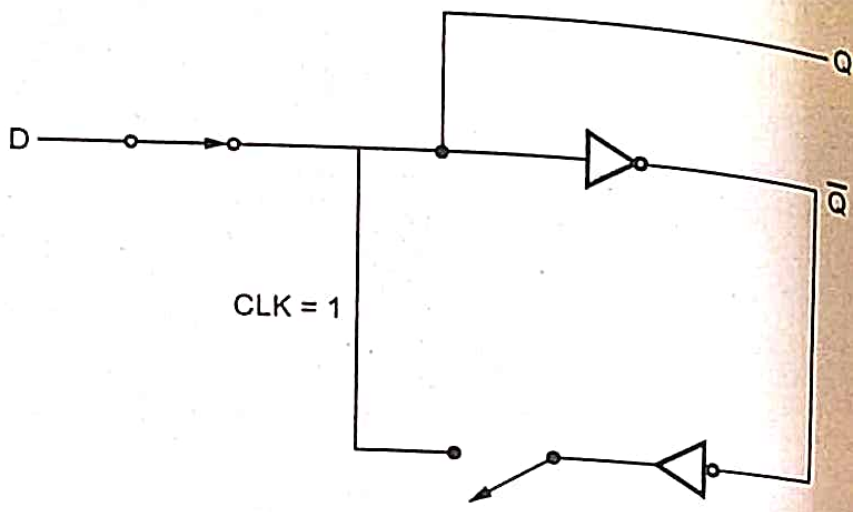


Fig. 7.5.2 Transparent D-latch

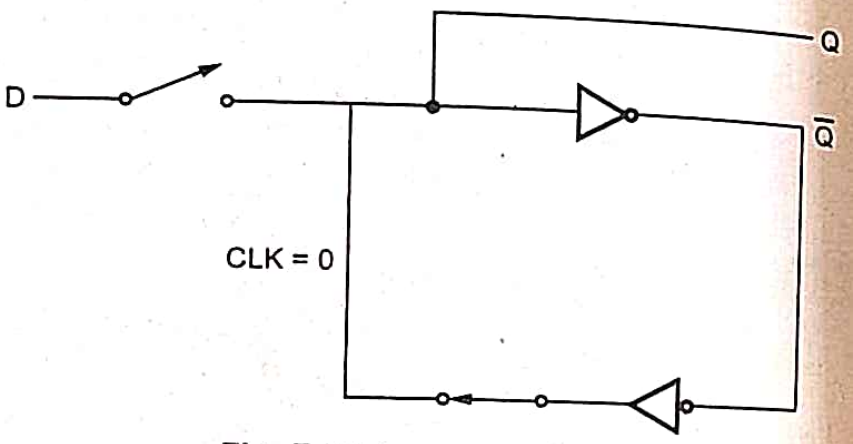


Fig. 7.5.3 Opaque D-latch

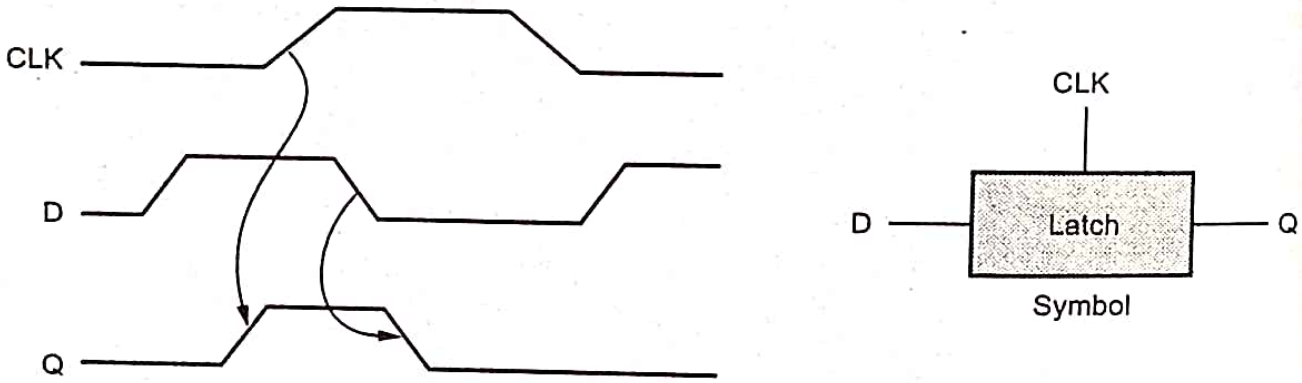


Fig. 7.5.4 Level sensitive D-latch

7.6 Master-Slave Edge Triggered Register

GTU : Winter-14, Summer-21

- The most common approach for constructing an edge-triggered register is to use a master-slave configuration, as shown in Fig. 7.6.1.

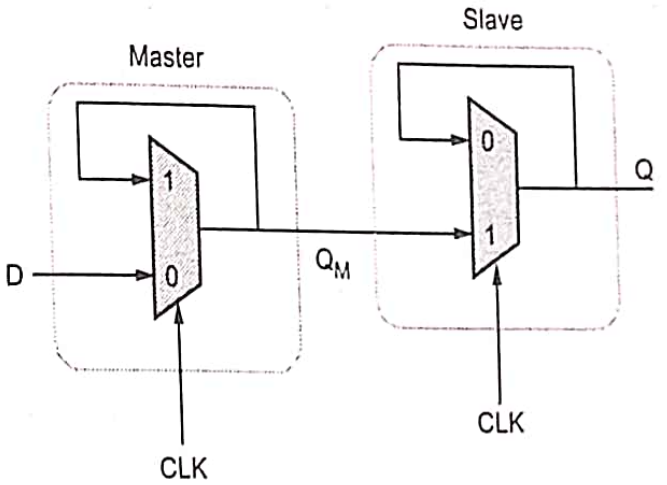


Fig. 7.6.1 Positive edge-triggered register based on a master-slave configuration

- The register consists of cascading a negative latch (master) with a positive latch (slave). A multiplexer-based latch is used in this particular implementation, although any latch could be used. On the negative phase of the clock, the master stage is transparent, and the D input is passed to the master stage output Q_M. The slave stage is in the hold mode i.e. keeping its previous value using feedback.
- On the rising edge of the clock, the master slave stops sampling the input, and the slave stage starts sampling.
- During the high phase of the clock, the slave stage samples the output of the master stage (Q_M), while the master stage remains in a hold mode.
- Since Q_M is constant during the high phase of the clock, the output Q makes only one transition per cycle. The value of Q is the value of D right before the rising edge of the clock, achieving the positive edge-triggered effect.
- Fig. 7.6.2 shows implementation of Master-slave positive edge-triggered register using transmission gates.

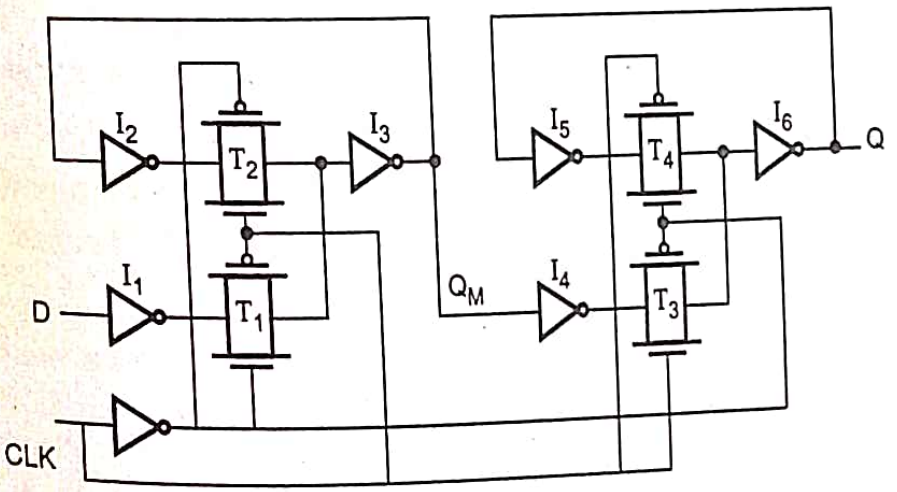


Fig. 7.6.2 Master-slave register using transmission gates

University Questions

1. *Realize negative edge-triggered master-slave D flip-flop using CMOS transmission gates and CMOS inverters. Explain its working with waveforms.* **GTU : Winter-14, Marks 7**
2. *Implement the CMOS edge-trigged master-slave D-flip flop.* **GTU : Summer-21, Marks 3**

7.7 Dynamic Latches and Registers

GTU : Winter-21

- Dynamic or clocked logic gates are used to decrease complexity, increase speed, and lower power dissipation.
- The basic idea behind dynamic logic is to use the capacitive input of the MOSFET to store a charge and thus remember a logic level for use later.

7.7.1 Dynamic TG Based Edge Triggered Registers

- A dynamic positive edge-triggered register based on the master-slave concept is shown in Fig. 7.7.1.

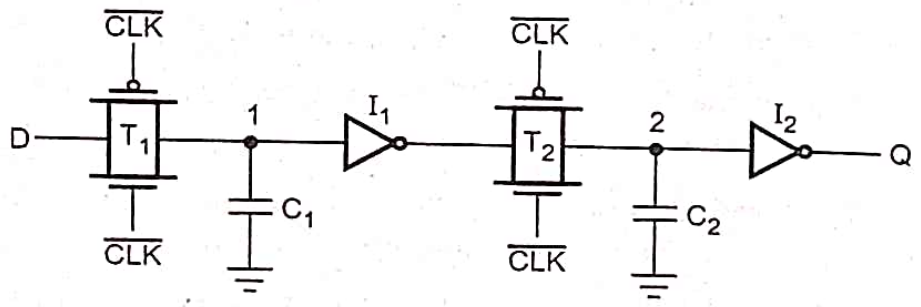


Fig. 7.7.1 Dynamic positive edge-triggered register

- T_1 and T_2 are transmission gates and I_1 and I_2 are invertors, C_1 and C_2 are equivalent capacitances at storage node-1 and node-2.
- One important consideration for such a dynamic register is that the storage nodes (i.e., the state) has to be refreshed at periodic intervals to prevent a loss due to charge leakage, due to diode leakage as well as sub-threshold currents.

Set-up time

- The set-up time of this circuit is simply the delay of the transmission gate and corresponds to the time it takes node 1 to sample the D input.

Hold time

- The hold time of this circuit is approximately zero, since the transmission gate is turned off on the clock edge and further inputs changes are ignored.

Propagation time

- The propagation delay (t_{c-q}) is equal to two inverter delays plus the delay of the transmission gate T_2.

Advantages

- The reduced transistor count is attractive for high-performance and low-power systems.

7.7.2 Clocked CMOS Dynamic Register

- Fig. 7.7.2 shows a positive edge-triggered register based on the master-slave concept which is insensitive to clock overlap. This circuit is called the C^2MOS (Clocked CMOS) register.

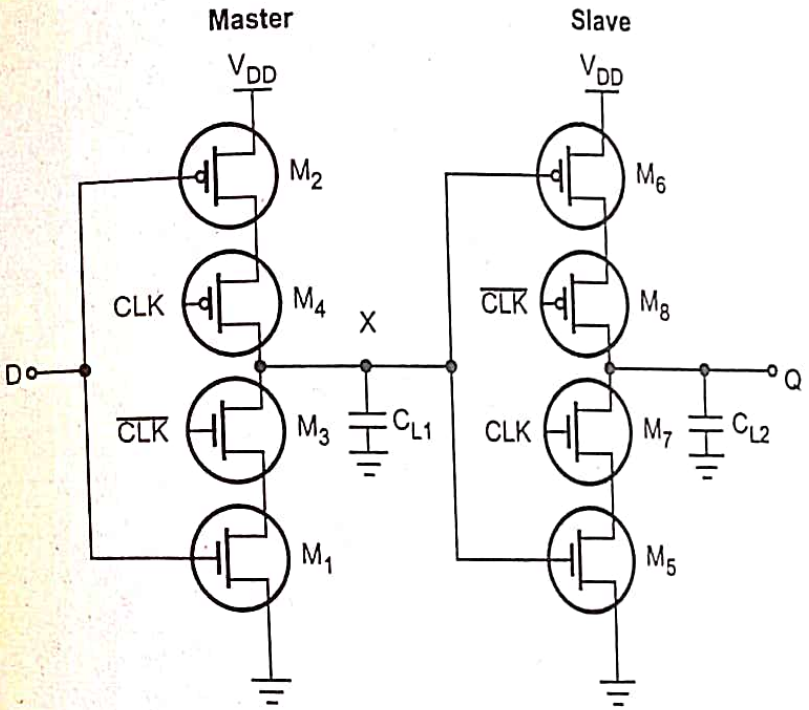


Fig. 7.7.2 C^2MOS master-slave positive edge-triggered register

- The circuit operates as a positive edge-triggered master-slave register similar to the transmission-gate based register. However, there is an important difference :
- A C^2MOS register with CLK - $\overline{CLK}$ clocking is insensitive to overlap, as long as the rise and fall times of the clock edges are sufficiently small.

University Question

1. *Explain clocked JK latch circuit and waveform.* **GTU : Winter-21, Marks 4**

7.8 Short Question and Answer

Q.1 What is latch-up ?

Ans. : Latch-up

- Latchup refers to short circuit formed between power and ground rails in an IC leading to high current and damage to the IC.
- In CMOS transistors, latch up is the phenomenon of low impedance path between power rail and ground rail due to interaction between parasitic pnp and npn transistors.

7.8 Short Question and Answer

Q.1 What is latch-up ?

Ans. : Latch-up

- Latchup refers to short circuit formed between power and ground rails in an IC leading to high current and damage to the IC.
- In CMOS transistors, latch up is the phenomenon of low impedance path between power rail and ground rail due to interaction between parasitic pnp and npn transistors.

❑❑❑

8 Dynamic Logic Circuits

Syllabus

Introduction, Basic Principles of pass transistor circuits, voltage bootstrapping, synchronous dynamic circuit techniques, CMOS dynamic circuit techniques, high-performance dynamic CMOS circuits.

Contents

8.1 Basic Principles of Pass Transistor Circuits . . . **Winter-14, 21, Summer-15, 21,** . . . . . Marks 7

8.2 Voltage Bootstrapping . . . . . **Winter-14, 21,** . . . . . Marks 7

8.3 Synchronous Dynamic Circuit Techniques. . . . **Summer-21,** . . . . . Marks 4

8.4 Dynamic CMOS Circuit Techniques. . . . . **Winter-14, 21, Summer-15,** Marks 7

8.5 High-Performance Dynamic CMOS Circuits. . . **Summer-21,** . . . . . Marks 7

8.6 Short Questions and Answers

8.1 Basic Principles of Pass Transistor Circuits

GTU : Winter-14, 21, Summer-15, 21

- A basic nMOS dynamic logic circuit consists of an nMOS pass transistor driving the gate of another nMOS transistor as shown in Fig. 8.1.1.

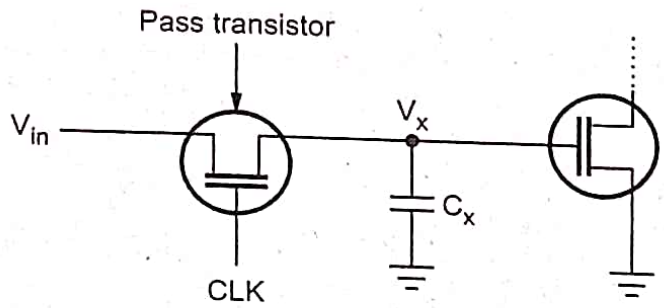


Fig. 8.1.1 Basic nMOS dynamic Logic circuit

- The pass transistor is driven by periodic clock signal (CLK). Depending on input signal V_{in} and clock signal, the pass transistor allows to charge up or charge down the parasitic capacitance C_x. Two possible operations when CLK = 1 are :
  i) Logic "1" transfer i.e. charging up capacitance C_x to logic high level.
  ii) Logic '0' transfers i.e. charging down capacitance C_x to logic low level.
- The pass transistor provides only current path to the capacitive node. When clock is inactive (CLK = 0), the pass transistor is off and charge across C_x determines the output of inverter.

Logic "1" Transfer

- Equivalent circuit for logic "1" transfer is show in Fig. 8.1.2.

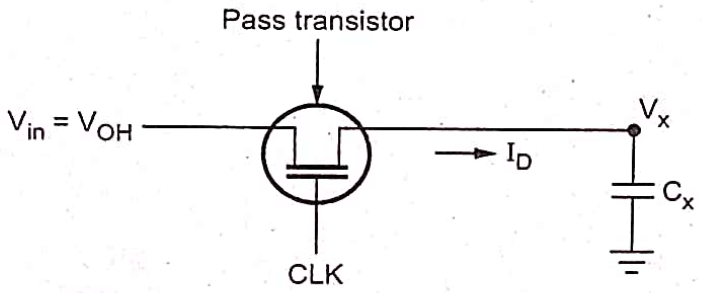


Fig. 8.1.2 Equivalent circuit for logic "1" transfer event

- The variation of node voltage V_x is given by expression.

$$V_{max} = V_{DD} - V_{T0,n} - \gamma(\sqrt{|2\phi_F| + V_{max}} - \sqrt{|2\phi_F|})$$

- The variation of node voltage V_x as a function of time is illustrated in Fig. 8.1.3.

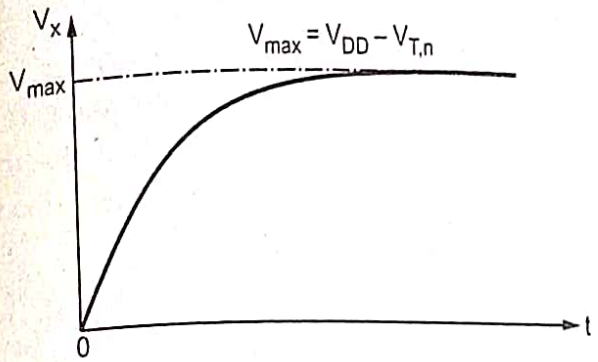


Fig. 8.1.3 V_x for logic "1" transfer

Logic "0" transfer

- Equivalent circuit for logic "0" transfer is shown in Fig. 8.1.4.

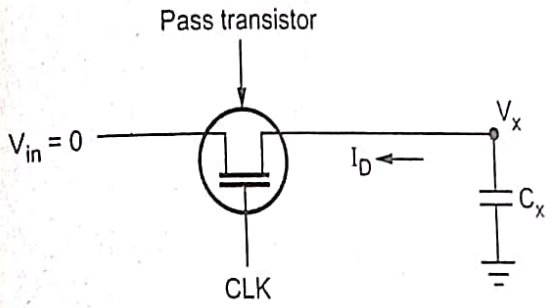


Fig. 8.1.4 Equivalent circuit for logic "0" transfer event

- The variation V_x as a function of time during logic "0" transfer shown in Fig. 8.1.5.

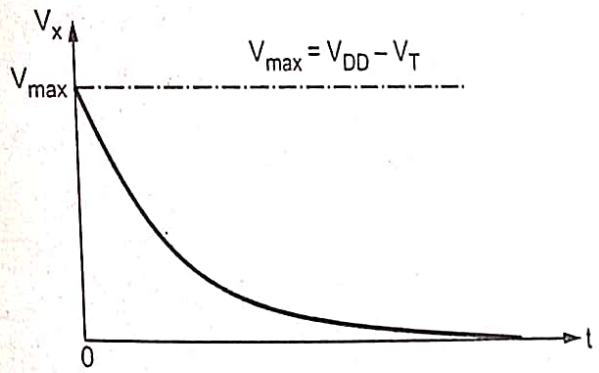


Fig. 8.1.5 V_x for logic "0" transfer

University Questions

1. *Show that the transmission gate (TG) is a good conductor of '1' as well as '0'. Implement two-input multiplexor circuit using is TGs.* **GTU : Winter-14, Marks 7**
2. *Discuss basic principles of pass transistor circuits and logic "1" transfer.* **GTU : Summer-15, Marks 7**
3. *Explain the basic principle of dynamic logic using noMOS pass transistor and discuss the logic '1' and '0' transfer.* **GTU : Summer-21, Marks 7**
4. *Write short note on Complimentary Pass transistor Logic (CPL).* **GTU : Winter-21, Marks 4**

8.2 Voltage Bootstrapping

GTU : Winter-14, 21

- Voltage bootstrapping is a technique used for overcoming threshold voltage drops in digital circuits. Output voltage levels sutfers from threshold voltage drops in circuits such as pass transistor gates or enhancement load inverters and logic gates.
- Consider an enhancement type circuit in which voltage V_x is equal or small than V_{DD} ($V_x \leq V_{DD}$).

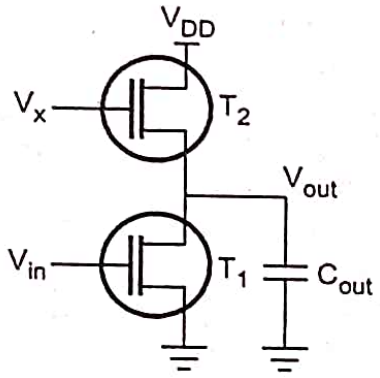


Fig. 8.2.1 Enhancement circuit

- The enhancement type nMOS transistor T2 will operate in saturation. If input voltage V_{in} is low, the maximum output voltage is limited by

$$V_{out}(\max) = V_x - V_{T2}(V_{out})$$

- In order to overcome threshold voltage drop and to obtain a full logic high level V_{DD} at output. The node voltage V_x must be increased.

Dynamic bootstrapping to boost V_x during switching

- A third transistor T3 is added to the circuit. Capacitance C_{boot} and C_S represent the capacitance that dynamically couple the node voltage V_x to the output and ground respectively.
- The circuit produces a high V_x during switching; so that the threshold voltage drop can be overcome at output node.

$$V_x \geq V_{DD} + V_{T2}(V_{out})$$

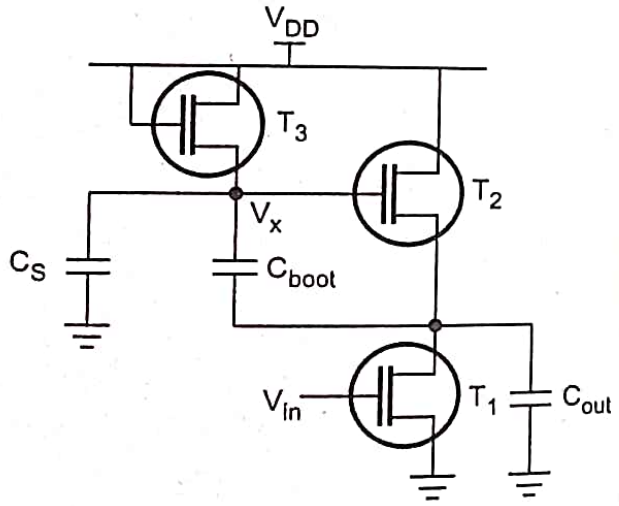


Fig. 8.2.2 Bootstrapping arrangement to boost V_x during switching

University Questions

1. *What is the need of voltage bootstrapping ? Draw MOSFET based voltage bootstrapping circuit. Explain its operation along with necessary mathematical steps.* **GTU : Winter-14, Marks 7**
2. *Explain voltage bootstrapping in detail.* **GTU : Winter-21, Marks 7**
3. *Explain voltage bootstrapping and derive capacitance ration equation.* **GTU : Winter-21, Marks 7**

8.3 Synchronous Dynamic Circuit Techniques

GTU : Summer-21

8.3.1 Comparison of Ratioed

Sr. No.	Ratioed logic	Ratioless logic
1.	In ratioed logic networks, the pull up network and the pull down network are connected in series to each other.	In ratioless logic networks, the pull up network and the pull down network are connected parallel to each other.
2.	If the output voltage is dependent on the aspect ratio of the transistors, then the circuit features ratioed logic.	If there is no dependency between the output voltage and transistors, then the circuit works on the principle of ratioless logic.
3.	NMOS circuits are examples of ratioed logic family.	CMOS circuits can be considered as instances of ratioless logic family.

8.3.2 Ratioed Circuits Design

- Ratioed logic is an attempt to reduce the number of transistors required to implement a given logic function, at the cost of reduced robustness and extra power dissipation.
- In complementary CMOS the pull-up network provides a conditional path between V_{DD} and the output when pull-down network is turned off.
- In ratioed logic, the entire pull-up network is replaced with a single unconditional load device that pulls up the output for a high output.
- Instead of a combination of active pull-down and pull-up networks, such a gate consists of an NMOS pull-down network that realizes the logic function and a simple load device.

8.3.3 Dynamic Pass Transistor Circuits

- A common multistage synchronous circuits (combinational logic stages) interconnected through nMOS pass transistor is show in Fig. 8.3.1.

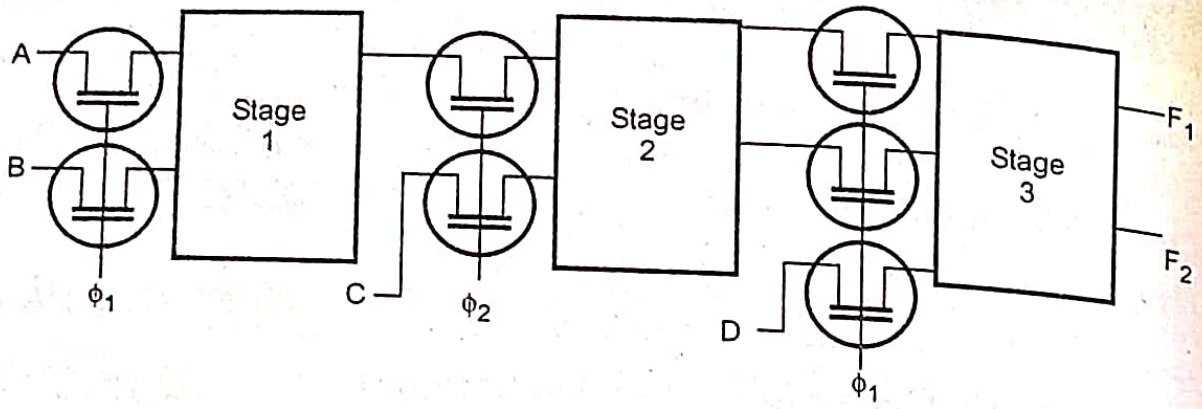
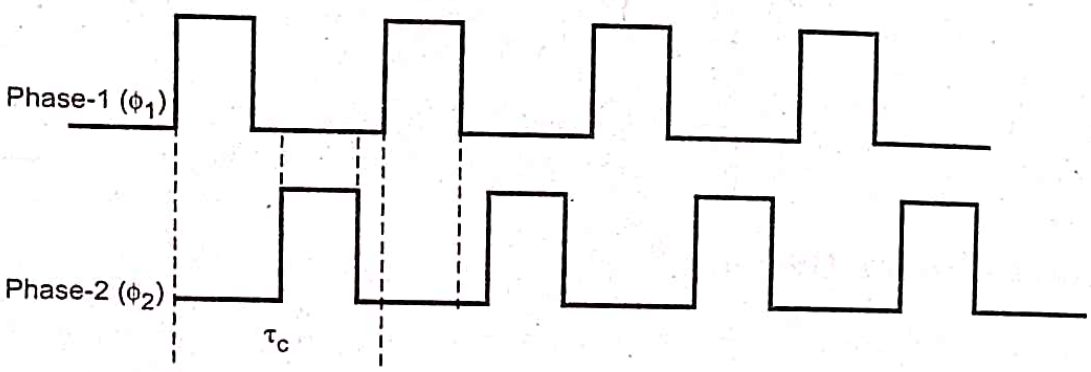


Fig. 8.3.1 Multistage synchronous circuit

Two Phase Clocking

- A two-phase non-overlapping clock signal is generally given to complex clocked sequential circuits due to its distinct advantages. We now consider clocked circuits that work on a two-phase non-overlapping clock signal. Such a signal is readily defined in Fig. 8.3.2 and it gives a great deal of freedom in designing sequential circuits if the clock period and the duration of the signals ph_1 and ph_2 are properly chosen.



Notes : 1. τ_c = clock period
2. $\phi_1(t).\ \phi_2(t) = 0$; all t

Fig. 8.3.2 Two phase clocking

- With proper signals ph_1 and ph_2, data is allowed to stabilize before any further transfer takes place and race conditions are completely avoided. Designing clocked circuitry is considerably earier than the corresponding asynchronous sequential circuitry. However clocked circuits may be slower. Inspite of this, we will concentrate on two-phase clocked sequential circuits alone to simplify design procedures, being the beginners of VLSI design. The clock signal in Fig. 8.3.2 reveal that ph_1 and ph_2 need not be symmetrical as shown. Once the clock period is specified, each clock phase period and its associated underlap period can be varied if required in order to optimize the design. The two-phase clock signals are

characterised by a clock period τ_C and the condition that $ph_1(t)\ ph_2(t)$ for all t. They can be generated by a number of techniques.

Depletion Load dynamic Shift Register

- A three-stage depletion load nMOS dynamic shift register circuit driven with two phase clocking is shown in Fig. 8.3.3.

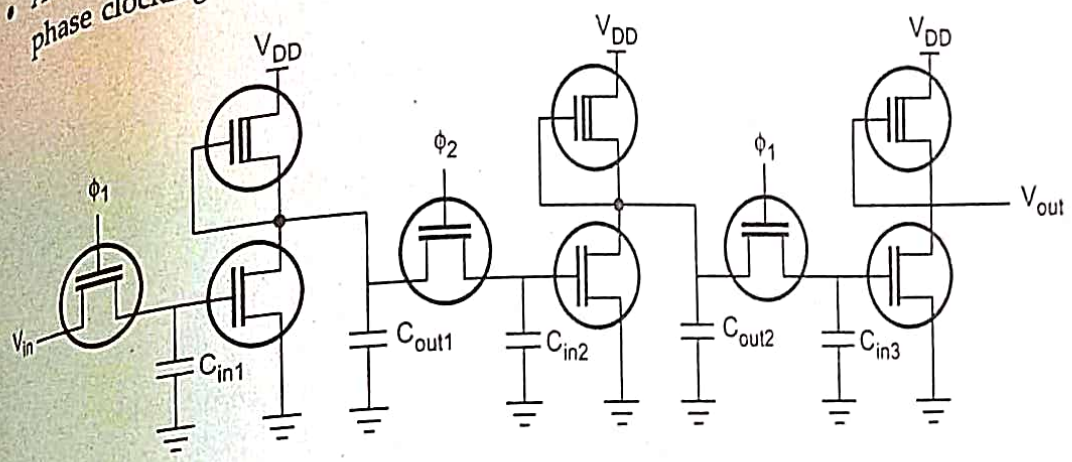


Fig. 8.3.3 Depletion load nMOS dynamic shift register

- The input data (V_{in}) is inverted once and shifted into next stage during each clock phase. The valid low output voltage level V_{OL} of each stage is strictly determined by the driver-to-load ratio, since the output pass transistor (input pass transistor to next) turns on in phase with load transistor. Hence, such circuit arrangement is called **ratioed dynamic logic.**

University Question

1. *Differentiate the rationed and ratioless dynamic logic circuits with examples.*

GTU : Summer-21, Marks 4

8.4 Dynamic CMOS Circuit Techniques

GTU : Winter-14, 21, Summer-15

8.4.1 CMOS Transmission Gate Logic

- In a two-phase synchronous logic circuit, static CMOS gates are used for implementing the logic blocks and CMOS transmission gates are used for transferring the output levels to next stage.
- The transmission gate is controlled by CLK and $\overline{CLK}$.
- Basic building block of a CMOS transmission gate dynamic shift register is shown in Fig. 8.4.1.

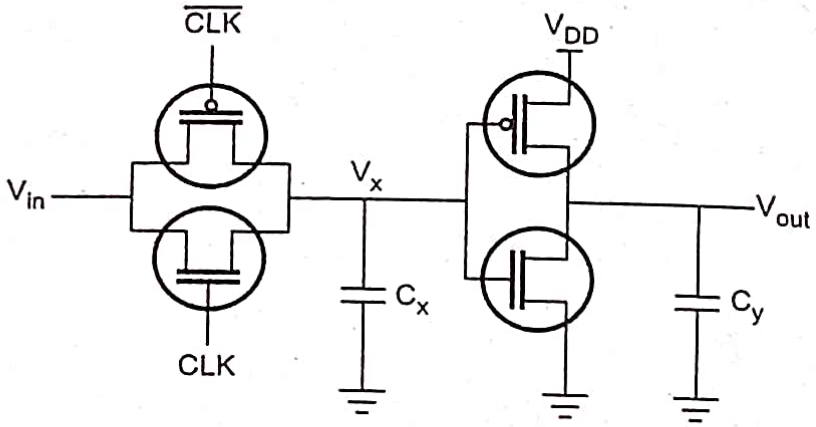


Fig. 8.4.1 CMOS transmission gate dynamic shift register

8.4.2 Advantages of Pseudo NMOS Transistor

Advantages of Pseudo NMOS Transistor :

1. Low area cost-only N+1 transistors are needed for an N input gate
2. Low input gate-load capacitance(C_{gn})
3. High speed
4. low transistor count.

Disadvantage :

1. Non-zero static power dissipation.

8.4.3 Pseudo nMOS Logic

- CMOS logic gates have following drawbacks
  a) Increased area. b) Increased capacitance.
  c) Greater delay d) Complexity.
- Therefore pseudo nMOS logic is an alternate gate circuit that is used as supplement for complementary MOS circuits. Fig. 8.4.2 shows pseudo nMOS logic inverter.

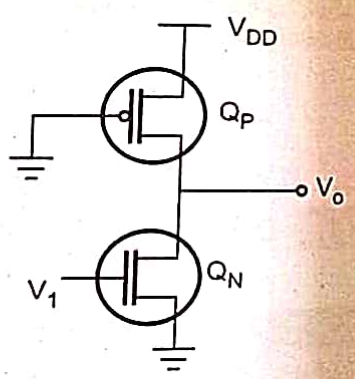
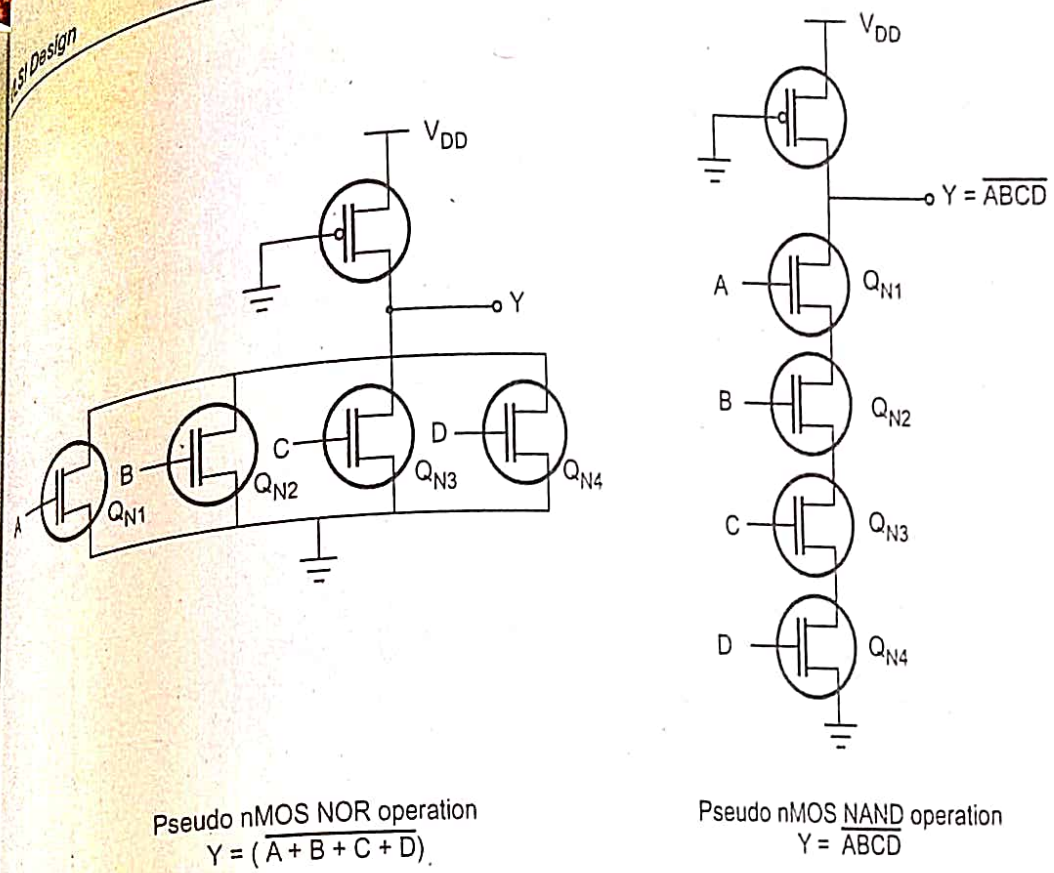


Fig. 8.4.2 Pseudo nMOS logic inverter

- The advantage of pseudo nMOS over CMOS is, each input must be connected to the gate of only on transistor or alternatively only one additional transistor (nMOS) will be needed for each additional gate input.
- Transistor is Q_N driven by input voltage whereas gate of transistor Q_P is grounded. Transistor Q_N acts as driver and Q_P act as load, hence the name pseudo nMOS. The number of transistors required in pseudo nMOS is less than complementry CMOS. Fig. 8.4.3 shows four input pseudo nMOS NOR and NAND operations.



Pseudo nMOS NOR operation
$Y = (\overline{A + B + C + D})$

Pseudo nMOS NAND operation
$Y = \overline{ABCD}$

Fig. 8.4.3

- The pseudo nMOS is preferred where high output is needed for most of the time. In such condition the static power dissipation is much less also the propagation delay becomes shorter. A common application of pseudo nMOS is in design of address decoders for memory chips and in ROMs.

8.4.4 Dynamic CMOS Logic

- A dynamic CMOS logic consists of an n-transistor logic structure of which output node is precharged to V_{DD} by a pass transistor and discharged conditionally by an n-transistor connected to V_{SS}.
- The precharge phase occurs when CLK = 0, the supply path is closed via n-transistor 'ground switch' during CLK = 1. The input capacitance of this gate is same as pseudo nMOS gate. The pull-up time is improved by virtue of active switch but pull-down time is increased due to ground switch.
- There are many advantages of dynamic CMOS logic over static CMOS logic or pseudo nMOS logic. The elimination of complementary pMOS transistors

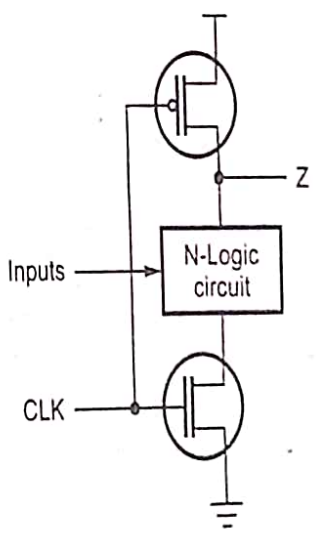


Fig. 8.4.4 Basic CMOS dynamic gate structure

significantly reduces the surface needed to implement the various logic functions. The switching speeds are increased using the dynamic logic.

8.4.5 Clocked CMOS Logic

- Clocked CMOS logic is used for low-power dissipation. This logic structure is meant to incorporate latches or interface with other form of logic. Fig. 8.4.5 shows a clocked CMOS gate.

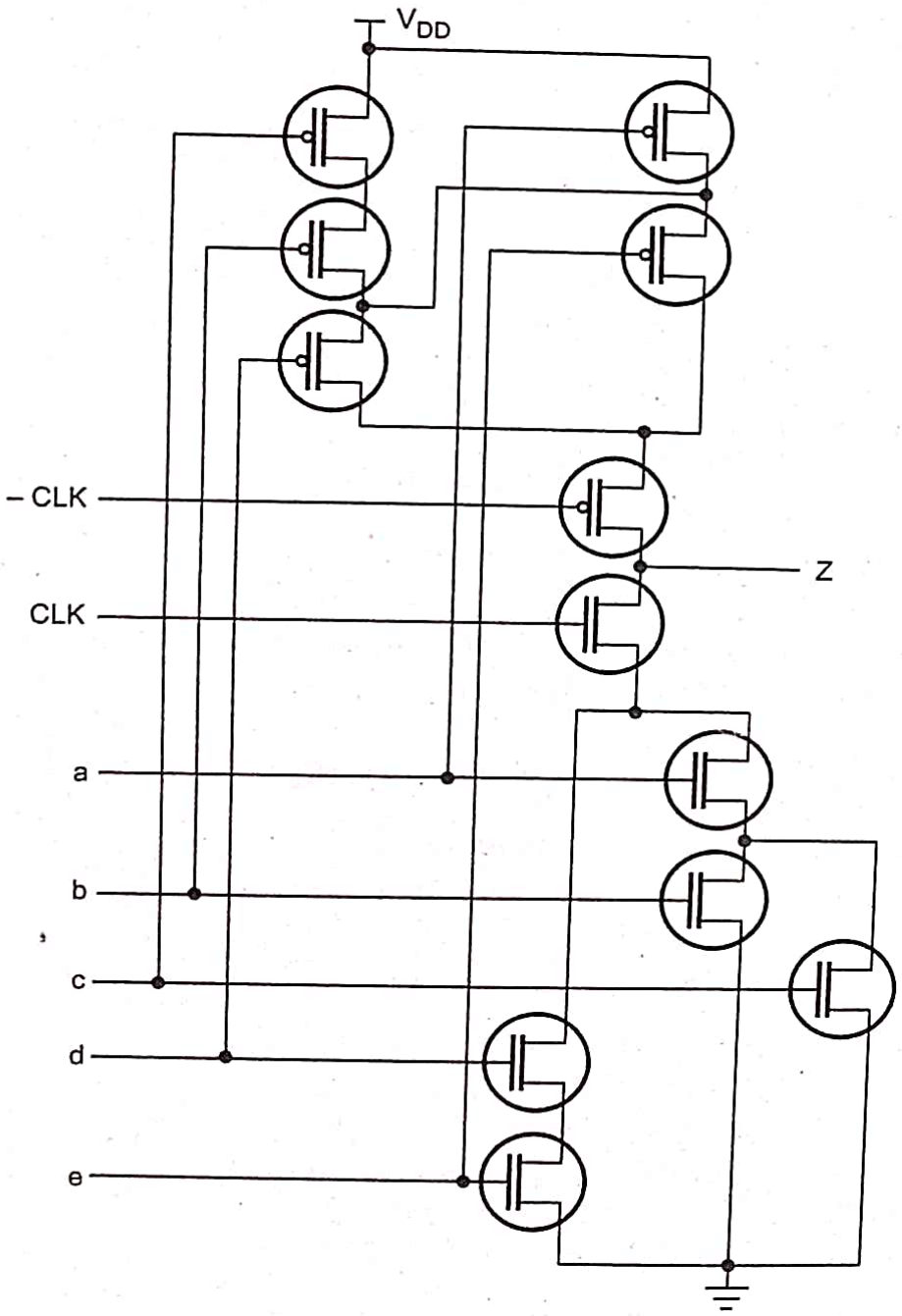


Fig. 8.4.5 Clocked CMOS gate

8.9 Design

- These logic structures are introduced to incorporate latches or interface with other dynamic forms of logic. The clocked CMOS gate have same input capacitance as other complementary gates but larger rise and fall times because of series clocking transistors.
- The series clock transistors can either be at the output of the gate or at the power supply ends. Since clocked CMOS circuit places an additional n-transistors in series with logic transistors it is recommended remedy for "hot electron effects".

Example 8.4.1 *In a logic design logic function is Z = (A+B+C+D)(E+F+G)(H+I) implemented with domino CMOS circuits diagram with implements Z.*

Solution :

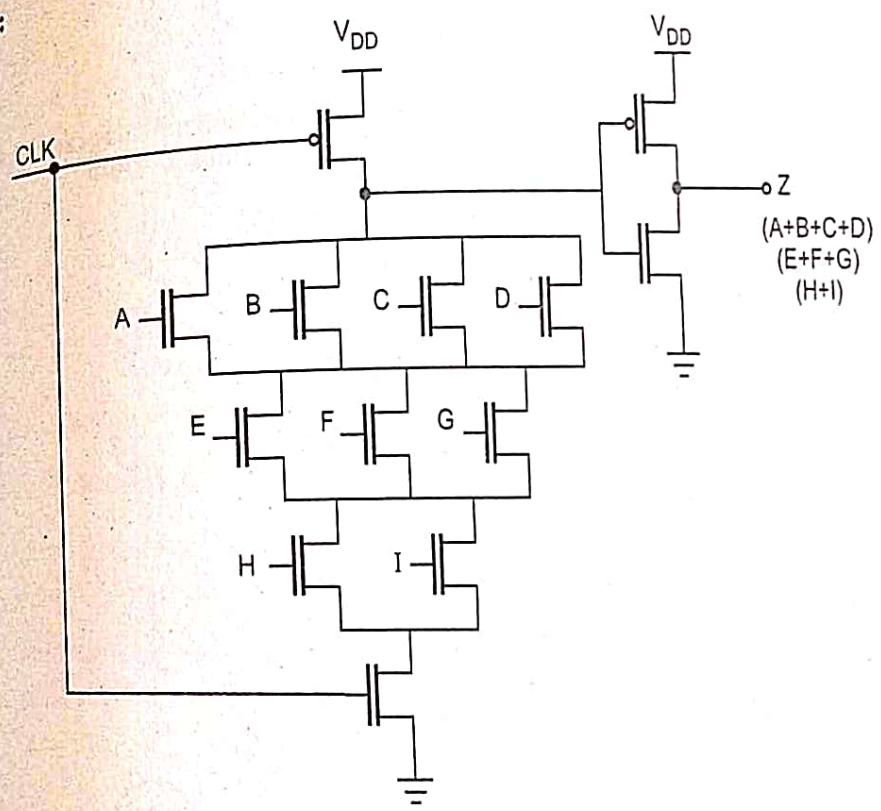


Fig. 8.4.6

University Questions

1. *With appropriate circuit example, illustrate the cascading problem in dynamic CMOS logic and the possible solutions for the same.* **GTU : Winter-14, Marks 7**
2. *Explain domino CMOS logic.* **GTU : Summer-15, Marks 7**
3. *Describe basic dynamic CMOS logic concept.* **GTU : Summer-21, Marks 7**
4. *Explain dynamic CMOS logic (Precharge-evaluate logic) in detail. Also illustrate the cascading problem in dynamic CMOS logic.* **GTU : Winter-21, Marks 7**
5. *Write advantages and disadvantages of dynamic logic circuit.* **GTU : Winter-21, Marks 3**

8.5 High-Performance Dynamic CMOS Circuits

GTU : Summer-21

8.5.1 CMOS Domino Logic

- A generalized diagram of CMOS domino logic gate is shown in Fig. 8.5.1.
- CMOS domino logic gate reduces significantly number of transistors require to realize any complex Boolean functions. This can be illustrated by implementation of a Boolean function. Consider as 8-input Boolean function
  $Z = AB + (C + D)(E + F) + GH.$

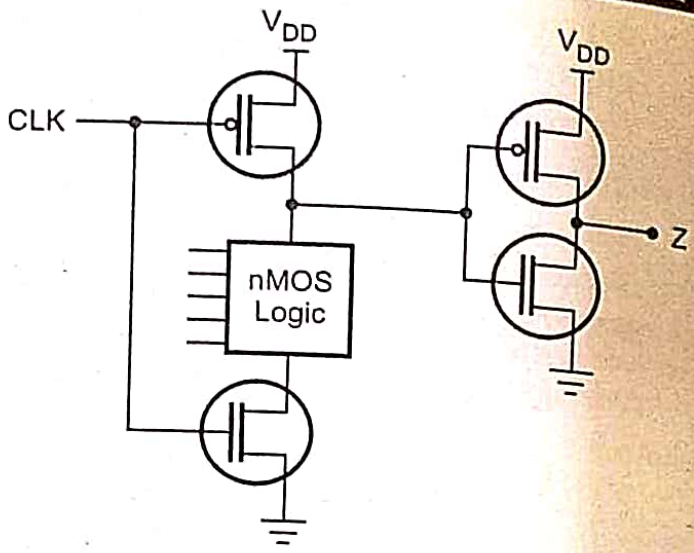


Fig. 8.5.1 Generalized diagram of CMOS domino logic

Implementation of Boolean function using standard CMOS

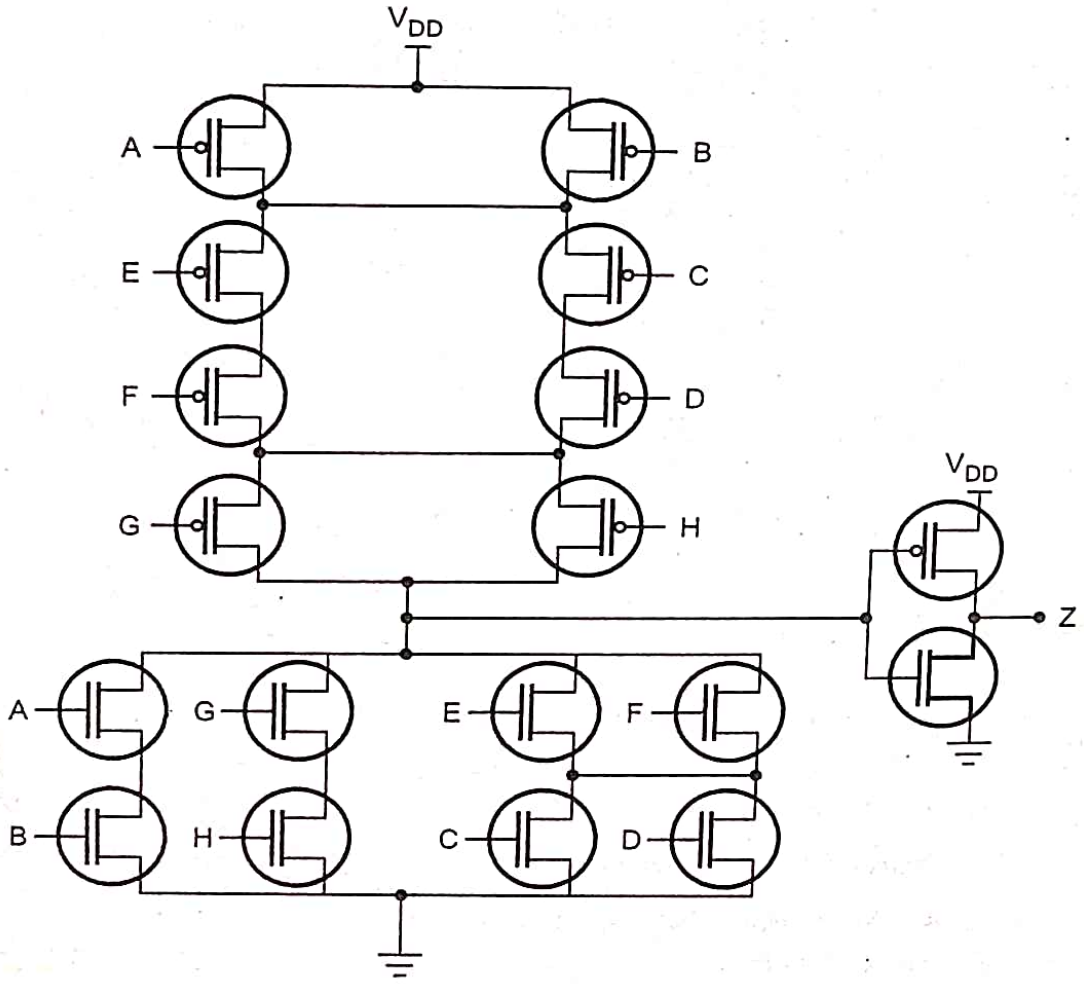


Fig. 8.5.2 Implementation of Z = AB + (C + D) (E + F) + GH using standrad CMOS

- When the some complex logic is implemented using domino CMOS logic there is reduction in circuit complexity Fig. 8.5.3 shows complex logic implementation using CMOS domino logic.

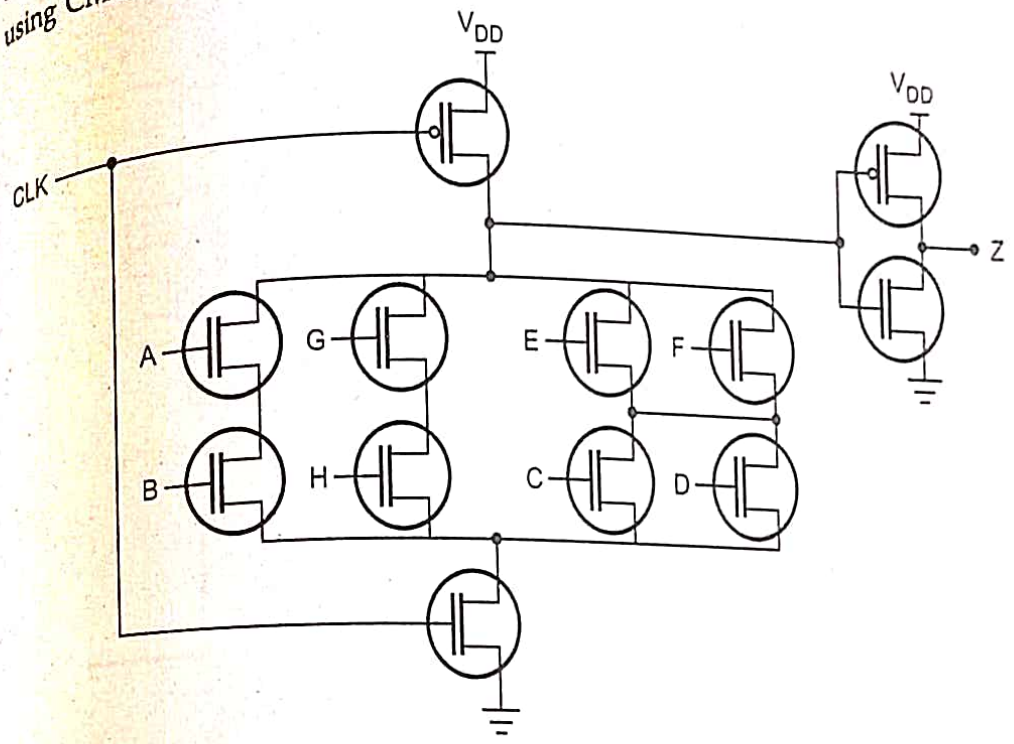


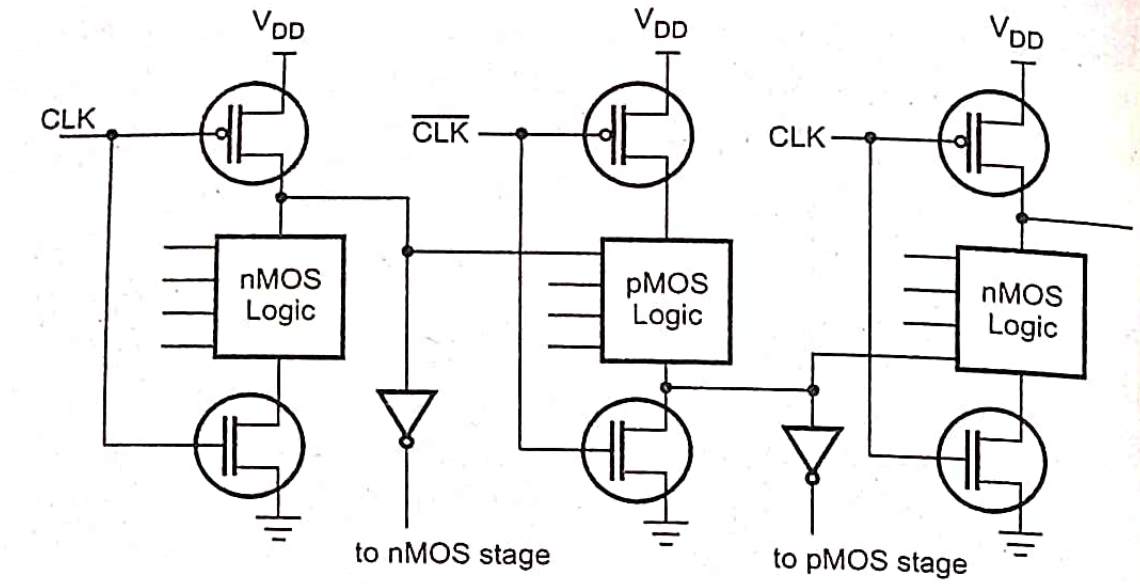
Fig. 8.5.3 Implementation of Z = AB + (C + D) (E + F) + GH using domino CMOS

Limitations associated with domino CMOS

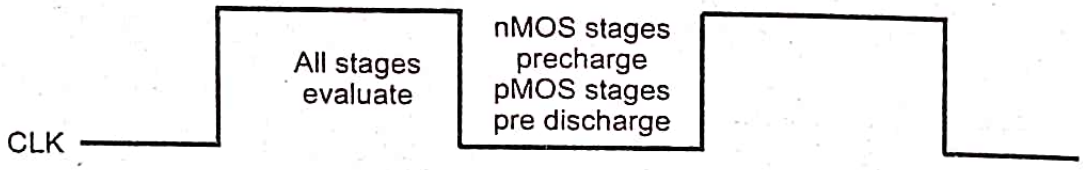
1. Only non-inverting structures can be implemented using domino CMOS. For inversions conventional CMOS is used.
2. The charge sharing between dynamic stage output node and intermediate nodes of nMOS logic block during evaluation phase may cause error in outputs.

8.5.2 NORA CMOS Logic

- In domino CMOS logic all logic operations are performed by nMOS transistors acting as pull down networks. Where as role of pMOS transistor is limited to precharging the dynamic nodes. NORA CMOS logic consists of alternating nMOS and pMOS stages. Fig. 8.5.4 shows NORA CMOS logic and the scheduling of precharge/evaluation phases.



(a) NORA CMOS (alternate nMOS and and pMOS logic)



(b) Precharge and evaluation phases

Fig. 8.5.4 NORA CMOS Logic Construction

Advantages of NORA CMOS

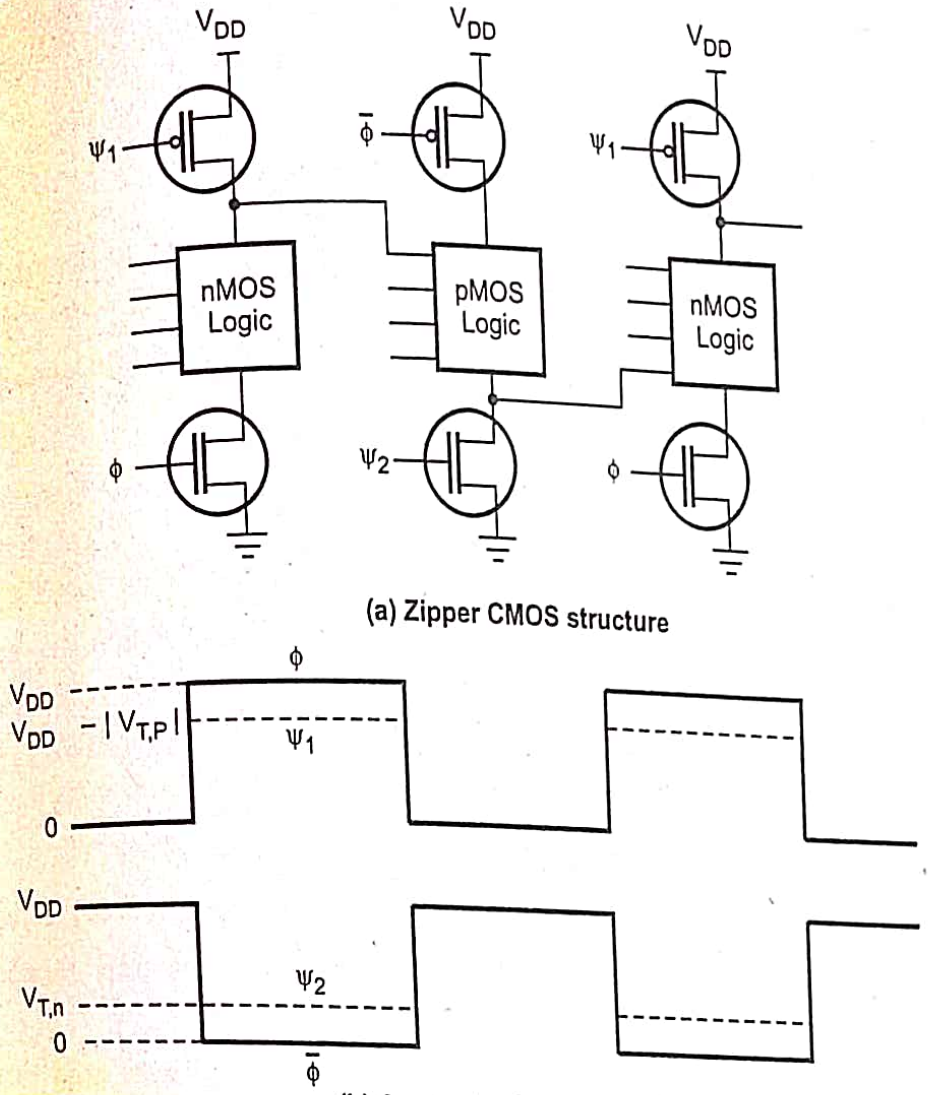
1. A static CMOS inverter is not needed at output of every dynamic logic stage.
2. NORA logic is compatible with domino CMOS logic.
3. NOR CMOS logic allows pipelined architecture.

Disadvantages of NORA CMOS

1. NORA CMOS logic suffers from charge sharing and leakage similar to dynamic CMOS structures.

8.5.3 Zipper CMOS Logic

- Zipper CMOS is identical to NORA CMOS except the clock signals. The zipper CMOS clock requires the generation of different clock signals for precharge and for pull down transistors. General circuit structure of zipper CMOS circuit and clock signals are shown in Fig. 8.5.5.



(a) Zipper CMOS structure

(b) Clock signals for zipper CMOS

Fig. 8.5.5 Zipper CMOS circuit and clock signals

University Question

1. *Explain high performance domino CMS logic circuit.* **GTU : Summer-21, Marks 7**

8.6 Short Questions and Answers

Q.1 Write about the pass transistor.

Ans. : Pass transistor

- MOS transistors can be used as switches by virtue of the isolated nature of the gate. When the transistor is operating in the linear region, the device acts as a linear resistance under gate voltage control. In this mode the transistor can be used as an on-off switch.
- The output node V_{out} is either held at its previous logic value of the switch is open or its capacitance C is charged up to V_{in} if the gate voltage allows the

transistor path to be closed, forcing a current for a short period of time. The transistor used in this fashion is called a **pass transistor.**

Q.2 List the advantages of pass transistor ?

Ans. : Advantages of pass transistor

1. Pass Transistor Logic (PTL) circuits are often superior to standard CMOS circuits in terms of layout density, circuit delay and power consumption.
2. They do not have path VDD to GND and do not dissipate standby power (static power dissipation).

Q.3 Write about the clocked CMOS logic and its usage.

Ans. : Clocked CMOS logic

- Clocked CMOS logic is used for low-power dissipation. This logic structure is meant to incorporate latches or interface with other form of logic.
- The clocked CMOS gate have same input capacitance as other complementary gates but larger rise and fall times because of series clocking transistors.

Q.4 Enlist the advantage of dynamic logic circuit over static logic circuit.

Ans. : Advantages of dynamic logic gates :

1. Allow implementation of simple sequential circuits with memory functions.
2. Use of common clock signals throughout the system enables the synchronization of various circuit blocks.
3. Implementation of complex circuits requires a smaller silicon area than static circuits.
4. Often consumes less dynamic power than static designs, due to smaller parasitic capacitances.
5. Smaller area than fully static gates.
6. Higher speed : Smaller parasitic capacitances.
7. Glitch free operation if design carefully.

9 Chip I/P and O/P Circuits

Syllabus

On chip Clock Generation and Distribution, Latch - Up and its Prevention.

Contents

9.1 On-Chip Clock Generation and Distribution . . . Winter-14, 21, ... Summer-15, 21, ... Marks 7

9.2 Latch-up and its Prevention . . . . . Summer-15, 21, ... Marks 7

9.3 Short Questions and Answers

9.1 On-Chip Clock Generation and Distribution

GTU : Winter-14, 21, Summer-15, 21

- Most VLSI systems are a combination of - pipelined system and Finite State Machines (FSM).
- Choosing the right clocking scheme affects the functionality, speed, and power of a circuit.

9.1.1 Clock Skew

- Clock is necessary in synchronous systems to distinguish one step in a computation from the previous or next step. Clock should be applied at all clocked elements in a system simultaneously. So that the system shares a common time reference. Practically, the arrival time differs from one point to another, this difference is called as **clock skew.**
- A system may need one or more logical clocks. A logical clock is a signal with no skew. The clock skew occurs due to mismatched clock network paths, processing and environmental variations. An acceptable compromise among skew, power consumption and metal resource usage is expected. Distributing a single clock across the entire chip in a low-skew manner is challenging. The **clock-skew** is defined as difference between nominal and actual interarrival time of a pair of physical clocks.
- High speed digital signals usually include ringing, crosstalk and noise. At high speed of clock these effects are superimposed and it becomes difficult to study the individual effect. A sufficiently slow clock allows all signal transients to decay before starting the next clock cycle.
- The clock speed is determined by settling time of the clock signal. The circuit elements like memory, logic circuit is then decided by clock speed. Clocking strategy is very significant in SoC design.

Example 9.1.1 *Explain with waveform that how much clock skew between* CLK1 *and* CLK2 *can be tolerated in following circuit shown in Fig. 9.1.1 when*
i) CLK 1 delayed after CLK 2 ii) CLK 2 delayed after CLK 1

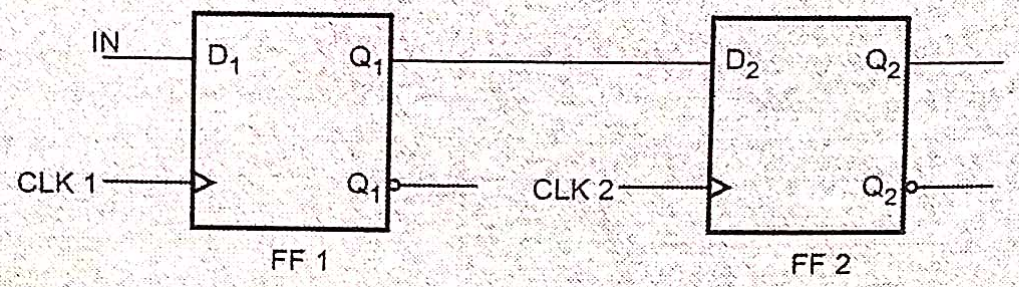


Fig. 9.1.1

Solution : If FF1's propagation delay from CLK1 to Q_1 is short, and if physical connection of Q_1 to FF2 is short, then change in Q_1 caused by CLK1 edge may actually reach FF2 before the corresponding CLK2 edge. In this case, FF2 may go to an incorrect next state determined by the next state of FF2 instead of current state.

If the change in Q_1 arrives at FF2 only slightly earliy relative to CLK2, then FF2 hold time specification may be violated, in which case FF2 may become metastable and produce an unpredictable output.

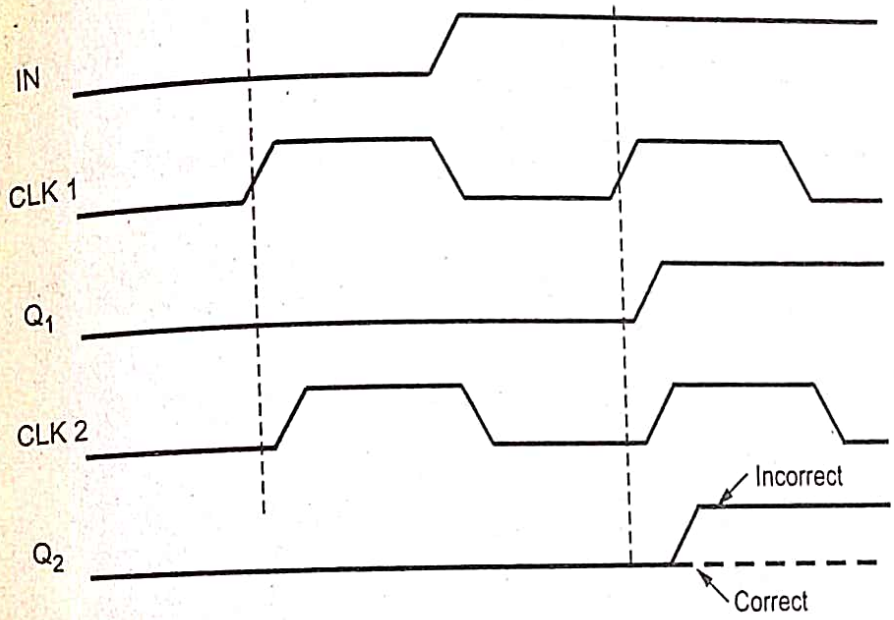


Fig. 9.1.2

For proper operation,

$$t_{ffpd\,(min)} + t_{comb\,(min)} - t_{hold} - t_{skew\,(max)} > 0$$

9.1.1.1 Clock Skew Computation

- Ideally Φ_n and Φ_{n+1} are synchronized → zero clock skew.
- Clock skew is the difference in clock signal arrival time between two sequentially adjacent register (blocks).

$$\tau_{skew} = \tau_n - \tau_{n+1}$$

Where

τ_n = Delay from the clock source to stage n

- Clock skew is due to the unbalanced of the data paths (i.e. different propagation delay). Although the layout of two data paths is identical, the different

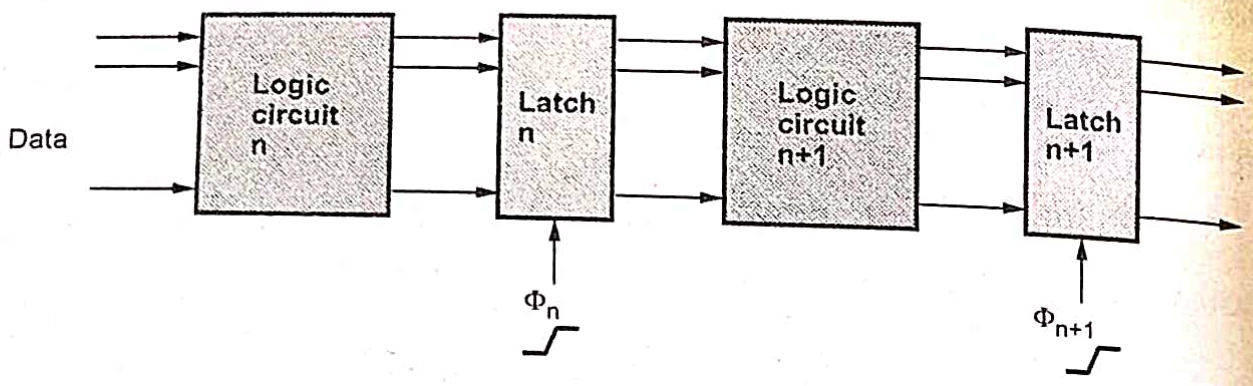


Fig. 9.1.3 Clock skew

neighbouring devices will also results in different propagation delay, subsequently clock skew.

- The minimum clock period between two registers must be greater than the sum of propagation delay and the clock skew.

$$T_{min} = \tau_{PD} + \tau_{skew}$$

Where

τ_{PD} = Propagation delay between stages n and n + 1,

τ_{skew} = Clock skew

9.1.1.2 Positive and Negative Clock Skew

- Positive clock skew increases the minimum clock period. Subsequently reduces the maximum operation frequency.
- Positive clock skew does not create race conditions because the input of the combinational circuits (stage n + 1) is not available yet.
- Negative clock skew reduces the minimum clock period. Subsequently increase the maximum operation frequency.
- Negative clock skew may create race conditions.
- τ_{skew} of negative clock skew cases must be less than the time required for the data to leave Latch (n), propagate through the interconnects and combinational logic circuits (n+1) in between, and allow Latch (n+1) to latch up.

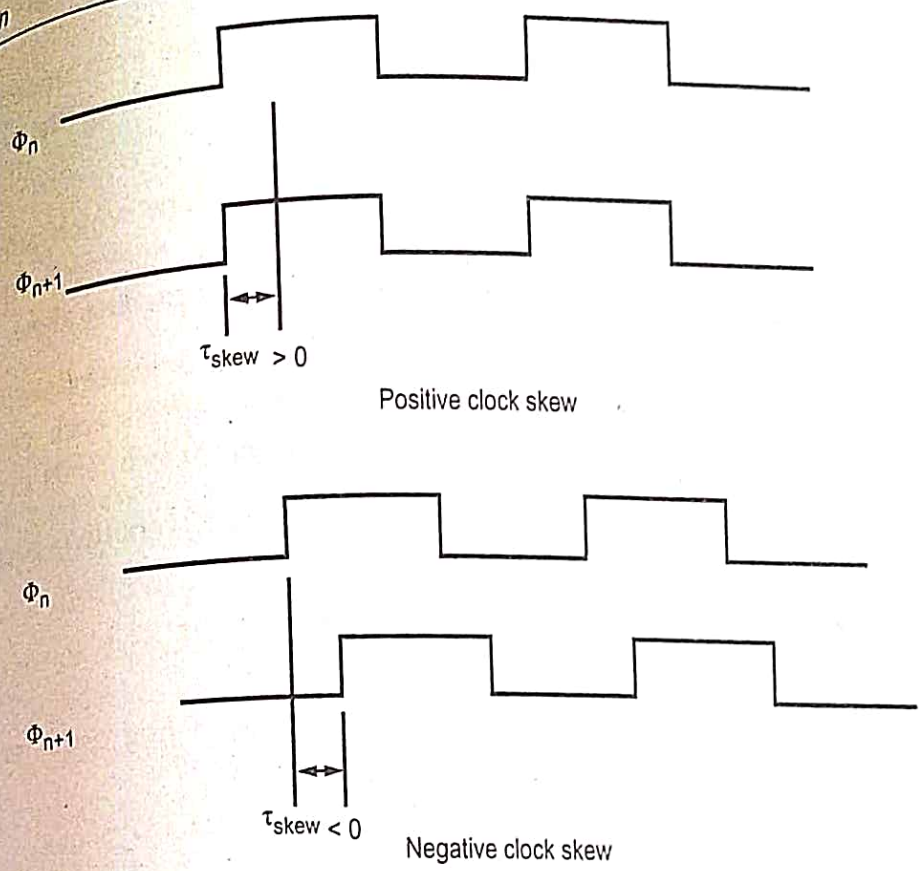


Fig. 9.1.4 Positive and negative clock skew

9.1.1.3 System Level Clock Skew

- The Φ_1 and Φ_2 are synchronized ideally. In reality, however Φ_1 and Φ_2 are not synchronized due to clock skew.
- The outputs generated by chips 1 and 2 are not be synchronized.
- Solution - use Phase-Lock Loop (PLL) to synchronize Φ_1 and Φ_2.

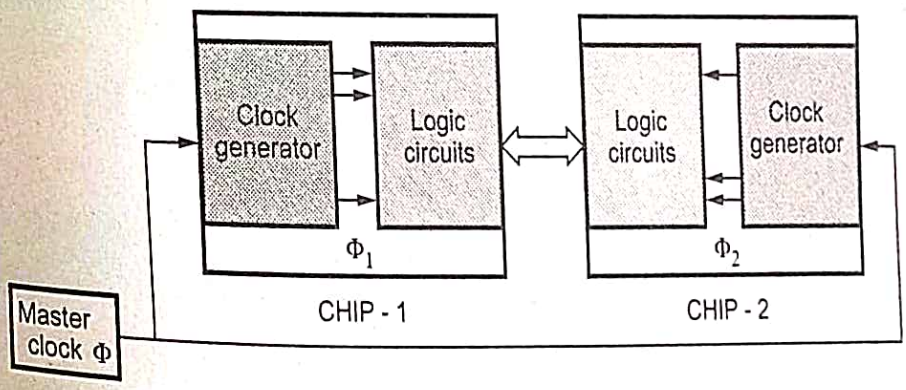


Fig. 9.1.5 Clock skew at system levels

9.1.1.4 Sources of Clock Skew

- Clock skew sources can be classified as
  1. Systematic
  2. Random
  3. Drift
  4. Jitter

1. Systematic : Systematic clock skew is that portion which exists even under nominal conditions. Systematic skew always exists because of flight time along the wire after gater.

2. Random : Random skew is caused by manufacturing variations which affects wire width, thickness or spacing, threshold voltage and oxide thickness. It causes unpredictable changes in resistance, capacitance, transistor current causing skew.

3. Drift : Drift skew is caused by environmental variations which occurs very slowly. e.g. after the chip turns on, it will heat up. The temperature affects each parameters differently also there can be temperature gradient across the chip which leads to skew.

4. Jitter : Jitter skew is caused by high-frequency environmental variation, particularly power supply noise.

9.1.2 On-Chip Clock Generation

- A simple technique for on-chip clock generation is to use a ring oscillator circuit as shown in Fig. 9.1.6.

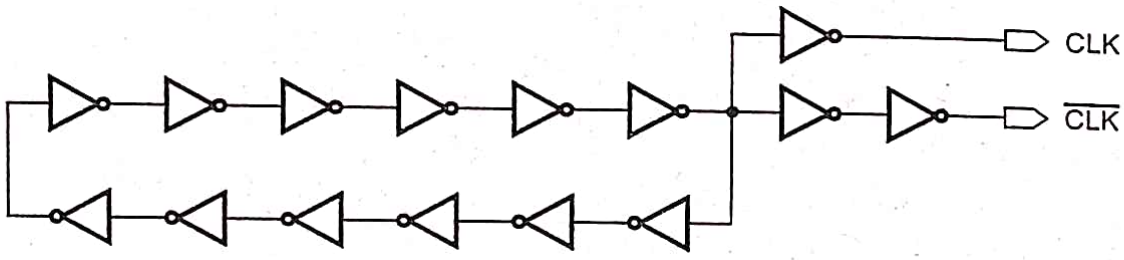


Fig. 9.1.6 Ring oscillator

- The clock generated by ring oscillator is quite unstable and process dependent hence to generate highly stable clock signal crystal oscillator is used. Fig. 9.1.7 shows circuit diagram of pierce crystal oscillator.

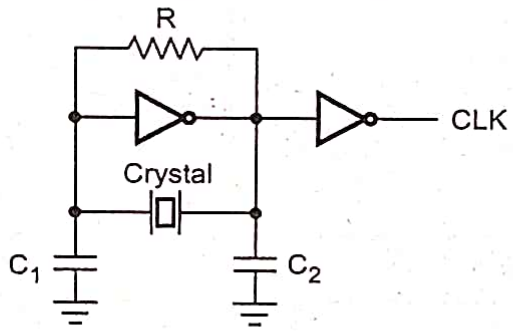


Fig. 9.1.7 Pierce crystal oscillator

- The inverter across the crystal provides the required potential difference and output inverter provides required amplification to drive the load.
- VLSI circuit requires two non-overlapping clock signals. The logical product of two clock signal should be zero all the times. A simple circuit that generates a pair of non-overlapping clock signals from a single clock is shown in Fig. 9.1.8.

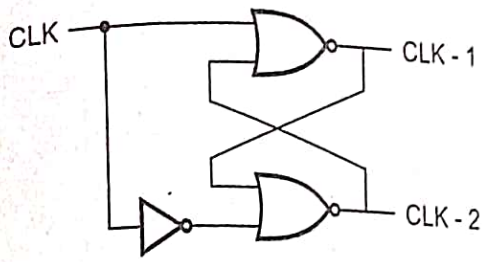


Fig. 9.1.8 Two non-overlapping clock signal generator

9.1.3 Clock Distribution

- In synchronous digital system, clock signal is used to define a time reference for the movement of data within the system. The clock distribution network distributes the clock signal from a common point to all the elements that need it. Clock distribution is regarded as one of the simple control signals.
- Stages of clock distribution involves :
  1. Impedance matching at every block.
  2. Local routing within every block.
  3. Global routing (matched impedance).
- The global clock must be distributed across the chip such that it reaches all the clocked elements at same time. In modern well designed process, the RC delay of resistive clock wire driving its own capacitance and the clock load capacitance tends to be close to 1 nsec. If the clock is routed randomly, it leads to a clock skew of about 1 nsec between physical clocks near and far from the clock generator, which is several times the cycle time of the system. Therefore, the clock distribution system must be carefully designed to equalize the flight time between clock generator and the clock receivers.
- The clock distribution networks can be classified as -
  1. Grids
  2. H-trees
  3. Spines
  4. Ad-hoc or hybrid.

9.1.3.1 Grids

- A clock grid consists of a mesh of horizontal and vertical wires driven from edges. The mesh has to deliver the clock to clocked elements.
- The resistance between any nearby points in the mesh is very low hence the clock skew is also low. It reduces the problems associated with hold time, as such problems occurs between nearby elements where the propagation delay between elements is also small.

- Grids also compensates the random skew. It can be routed in the early stage of design without placing the components.
- Grids have systematic skew between the closest point and farthest point. It consumes large amount of metal and has high switching capacitance and power consumption.

9.1.3.2 H-Trees

- An H-tree is a structure by drawing an H-shape then recursively drawing H shapes on each of the vertices (recursive structure). The width of the wires in H-tree are adjusted to account for variations in load capacitance to equalize skew in the H-tree. Buffers are added to increase drive capability with low input capacitance. The clock is routed to a central point on the chip and balanced paths. When the path is balanced or distributed uniformly the clock skew is zero. Fig. 9.1.9 shows clock distribution by H-tree.

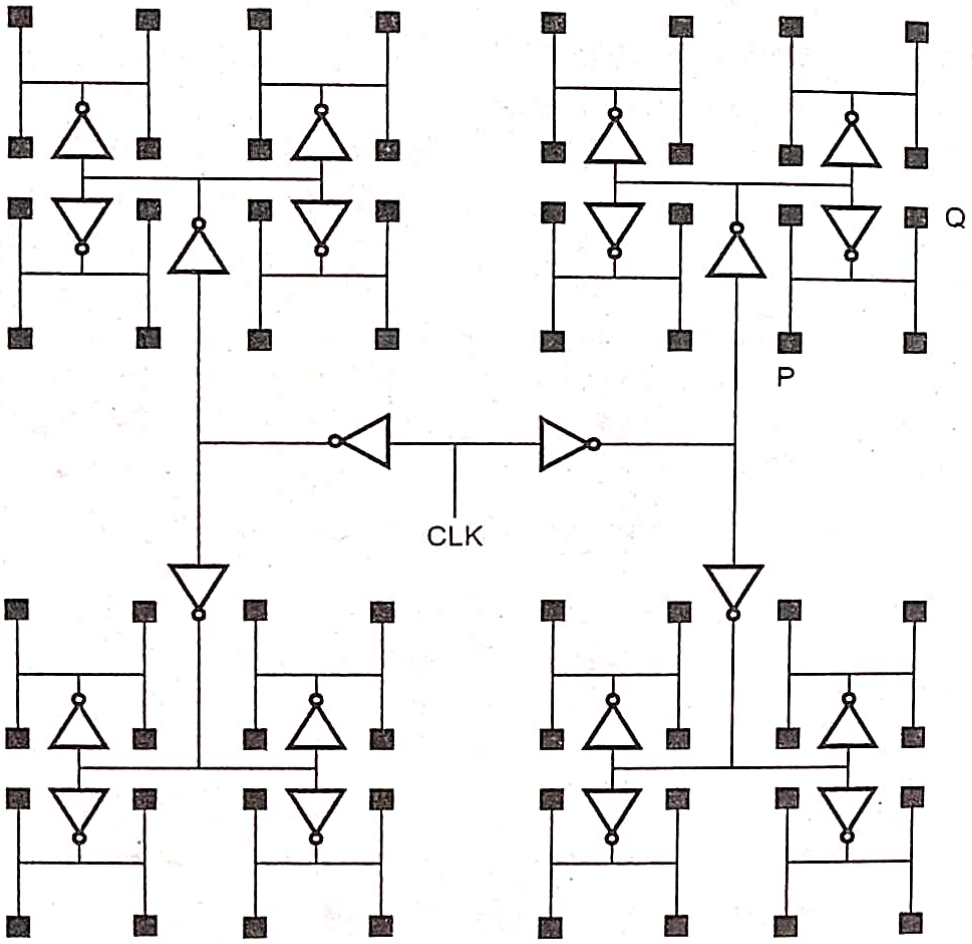


Fig. 9.1.9 H-tree

- H-tree can distribute a clock from the center to within an arbitrarily short distance on chip and maintains an equal wire lengths. The trees tend to use less wire and thus have lower capacitance than grids.

- Practically, H-tree shows some skew because clock loads are not uniform i.e. loading some leaves of tree more than others. Another drawback of H-tree is that they may have high random skew, drift and jitter between two nearby points that are leaves of different legs of tree, e.g.- points P and Q in Fig. 9.1.9 might experience large skews, as the points are close, this is a particular problem for hold times.
- H-tree structure finds its application particularly in regular-array networks in which all elements are identical also where the clock can be distributed as binary tree.

9.1.3.3 Spines

- Clock distribution using spines with serpentine routing is shown in Fig. 9.1.10.

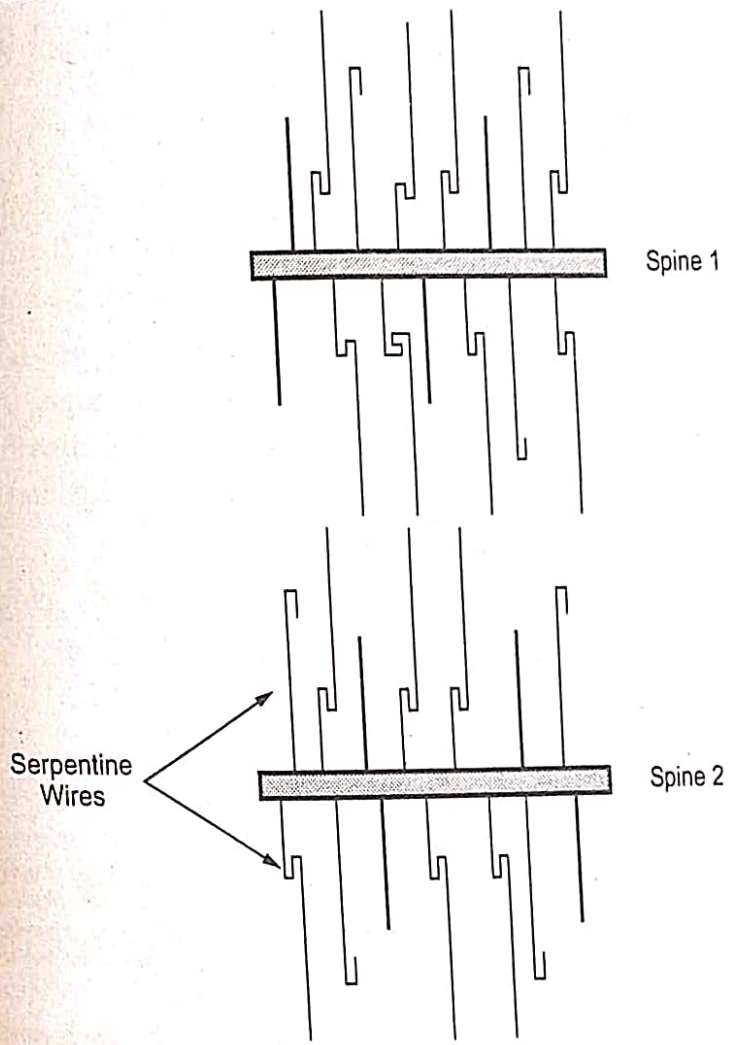


Fig. 9.1.10 Spine

- The spines drive length matched serpentine wires to a small group of clocked elements. For uniform loads, the spine avoids the systematic skew by matching the length of clock wires. Since each serpentine is driven individually any certain serpentine can be switched off, it saves power.

- The serpentine is easy to design and each load can be tuned separately. Use of multiple spine reduces the length of final clock wires e.g. Pentium-IV uses three clock spines.

9.1.3.4 Ad-hoc

- In ad-hoc clock distribution network, the clock is routed haphazardly by any attempt to equalize wire lengths or delay. The ad-hoc distribution network has reasonably low systematic skews because the buffer sizes can be adjusted until the nominal delays are nearly equal.
- Ad-hoc networks are subjected to severe random skew due to different wire and gate delays. Also the analysis of random skew, jitter and drift are difficult to analyze.
- Ad-hoc clock distribution is useful in many ASICs running at frequencies of 100's of MHz.

9.1.3.5 Hybrid

- A hybrid combination of H-tree and grid offers lower skew than either an H-tree or grid alone. In hybrid network, an H-tree is used to distribute the clock at multiple points over the chip. A grid is used to short these points.
- Hybrid approach has lower systematic skew compared to grid because the grid is driven from many points instead of just middle or edge. Also, the hybrid approach is less susceptible to skew from nonuniform load distributions compared to H-tree. Therefore, the hybrid approach is regular, making layout of well controlled transmission line structures easier.
- Hybrid distribution network is used by IBM in various chips like Power4, PowerPC and S/390.

University Questions

1. *What do yout understand by clock skew ? Explai various clock distribution schemes with diagrams.* **GTU : Winter-14, Marks 7**
2. *Draw general layout of an H-tree clock distribution network.* **GTU : Summer-15, Mark 1**
3. *What is clock-skew. Explain on-chip clock generation and distribution.* **GTU : Summer-21, Marks 4**
4. *Describe methods for on-chip clock generation.* **GTU : Summer-21, Winter-21, Marks 4**
5. *Discuss the on-chip clock generation and distribution.* **GTU : Winter-21, Marks 7**

9.2 Latch-up and its Prevention

GTU : Summer-15, 21

- A low impedance path in CMOS chips between power supply rail and ground rail because of interaction of parasitic pnp and npn transistors is called as **latch-up**.
- The latch-up susceptibility is inversely proportional to the product of substrate doping level and square of spacing i.e. if spacing is reduced by half and substrate doping is increased by two times, then latch-up susceptibility increases by factor of 2. Epitaxial layer and retrograde (well and channel formation) reduces severity of latchup problem.
- In a CMOS inverter due to latch-up parasitic npn and pnp transistors are formed. This parasitic npn and pnp transistors forms equivalent circuit of SCR.

9.2.1 Causes of Latch-up

- Following are the major causes of latch-up in CMOS circuits.
  1. Slewing of V_{DD} during start up. It cause enough displacement currents due to well junction capacitance in the substrate and well. The latch-up is induced if slew rate is large enough.
  2. Large currents in parasitic SCR in CMOS chip injects a triggering current which causes disturbance due to impedance mismatch in transmission lines of high speed circuits.
  3. Electro Static Discharge (ESD) can cause latch-up due to injection of minority carriers from the clamping device into either substrate or well.
  4. Sudden transients in power or ground buses due to simultaneous switching of many devices.
  5. Leakage currents in well junctions may cause lateral currents.
  6. Radiations due to X-rays, cosmic rays or alpha particles generates electron-hole pairs in both substrate and well regions.

9.2.2 Guidelines for Avoiding Latch-up

- The latch-up problem can be reduced or eliminated by employing several modifications or improvements.
  1. Reducing the gain of parasitic BJTs formed.
  2. Capture the injected minority carriers before they reach the base of parasitic BJTs.
  3. Place substrate and well contacts as close as possible to the source connections of MOS transistors.
  4. Use minimum area p-wells to minimize p-well photocurrent during transient pulses.

5. Avoid forward biasing of source / drain junctions to prevent injection of high current.
6. Maintain sufficient spacings between pMOS and nMOS transistors.
7. Layout n and p-channel transistors such that all nMOS transistors are close to V_{SS} and pMOS transistors are placed closed to V_{DD} rails.

University Questions

1. *Discuss guideline for avoiding of CMOS lath-up.* **GTU : Summer-15, Marks 7**
2. *Explain the latch - up problem observed in CMOS circuits and mention various techniques to prevent it.* **GTU : Summer-21, Marks 4**
3. *Describe latch-up problem in IC.* **GTU : Summer-21, Marks 4**

9.3 Short Questions and Answers

Q.1 Define clock skew.

Ans. : Clock skew

- **Clock skew** is a phenomenon in synchronous circuits in which the clock signal (sent from the clock circuit or source or clock definition point) arrives at different components at different times.
- **Causes of clock skew are** due to -
  1. Wire-interconnect length
  2. Temperature variations
  3. Capacitive coupling
  4. Material imperfections and
  5. Differences in input capacitance on the clock inputs
- These factors became more critical for high frequency.

Q.2 With respect to Flip-flop timings, explain the following :
1. Set-up timing 2. Hold time 3. Resolve time

Ans. : Flip-flop timings

1. Set-up time :

- The set-up time is time required for the input data signal at a flip flop to be valid before the incoming clock edge arrives.
- The time interval between the application of a signal that is maintained at a specified input terminal and a consecutive active transition at another specified input terminal.

2. Hold time :

- The hold time is the time required for the input data signal to remain valid after the clock edge transitioned.
- The interval during which a signal is retained at a specified input terminal after an active transition occurs at another specified input terminal.
- The combination of the two setup and hold time values determine the width of the metastability window. The larger the window, the greater the chance the device will go metastable.

3. Resolve time :

- The resolve time is the amount of time the flip flop's output must return to a valid level before it's used.

Resolve time = 1/{Clock frequency} – Path delay

- The output must be valid by the next clock, minus any chip or routing delay.

Path Delay = Tcko + Troute + Tsu

Where,

Tcko = Clock to output time of the flip flop,

Troute = Any trace delay between the the Q of the flip flop and the next device reading that data,

Tsu = Any set-up time required by the next device reading the data.

❑❑❑

Notes

10 Design for Testability

Syllabus

Introduction, Fault types and models, Controllability and observability, AdHoc Testable design techniques, Scan -based techniques, built-in SelfTest (BIST) techniques, current monitoring I_{DDQ} test.

Contents

10.1 Introduction

10.2 Fault Types and Models ..... Summer-21 ..... Marks 3

10.3 Observability ..... Winter-14,21, ... Marks 2

10.4 Controllability Winter-14, 21,

..... Summer-15 ..... Marks 3

10.5 Ad Hoc Testable Design Techniques ..... Summer-15, ... Marks 7

10.6 Scan-Based Techniques ..... Winter-14, ... Marks 3

10.7 Built-In Self Test (BIST) ..... Summer-15,21, Winter-21 · Marks 7

10.8 Current Monitoring I_{DDQ} Test

10.9 Short Questions and Answers

- Stuck-at-faults mostly occur due to shorting of gate oxide (NMOS gate to GND or PMOS gate to V_{DD}) or metal-to-metal shorts.

10.2.2 Stuck-Open and Stuck-Short Faults

- Stuck-open and stuck-short faults are usually referred as transistor faults. Physical faults which occurs at manufacturing level are called as **defects.** The electrical or logic-level faults by physical defects are referred as **defect oriented faults** such as open links, improper semiconductor doping, bridging faults etc.

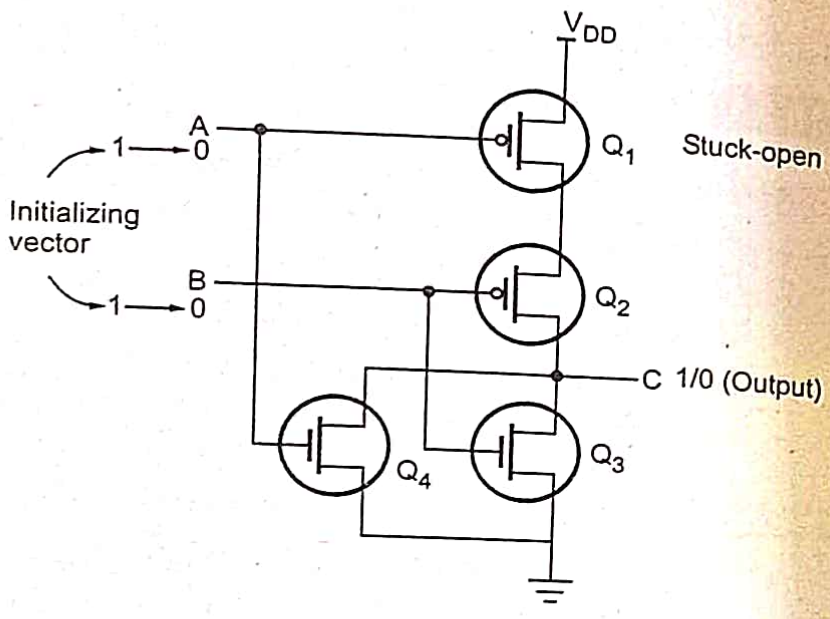


Fig. 10.2.2 CMOS NOR logic

- Consider a MOS transistor to be modelled as ideal switch. The defect may be switch being permanently in open state or short state. This fault model involves just one transistor as stuck-open or stuck-short. Fig. 10.2.2 shows as CMOS NOR logic circuit to illustrate this model.

10.2.3 Delay Faults

- The delay fault causes timing failures at target speed. Factors causing delay faults are :
  1. Improper estimation of on-chip interconnect delays and other timing considerations,
  2. Excessive variations in the fabrication process which cause significant variations in circuit delays and clock skews,
  3. Opens in metal lines connecting parallel transistors which make the effective transistor size much smaller,
  4. Aging effects such as hot-carrier induced delay increase.

10.2.4 Difference between Single Struct-at Fault and Multiple Struct-at Fault

Sr. No.	Single struct-at fault	Multiple struct-at fault
1.	Single Stuck-Fault model (SSF) is the classical or standard fault model.	Multiple Stuck-Fault (MSF) model is a straight forward extension of the SSF model.
2.	There is only two fault per line i.e. struct-at-1 and struct-at-0.	More than two faults per line.
3.	If n - is the number of possible SSF sites, there are 2n possible SSFs.	There are $3^n - 1$ possible MSFs.

University Questions

1. *Explain various types of faults observed during the chip testing.* **GTU : Summer-21, Marks 3**
2. *Describe different types of fault.* **GTU : Summer-21, Marks 7**

10.3 Observability

GTU : Winter-14,21

- **Observability** is a degree to which it can be observed that node at output operates correctly. Observability is useful, when a test engineer has to measure the output of a gate/chip within a larger circuit. To check its correct operation. Higher observability indicates less number of cycles required to measure output node value. Circuit having poor observability includes sequential circuits with long feedback 100 ps.

University Question

1. *Define observability.* **GTU : Winter-14,21, Marks 2**

10.4 Controllability

GTU : Winter-14, 21, Summer-15

- **Controllability** is the ability of an internal circuit within a chip the ease of setting the node to logic 1 or 0 state.
- Controllability is important while accessing the degree of difficulty of testing a particular signal in a circuit.
- A node with little controllability may take hundreds of cycles to get it to the right state.

University Questions

1. *Define controllability.* **GTU : Winter-14, Marks 2**
2. *Discuss controllability in brief.* **GTU : Summer-15, Mark 1**
3. *Explain controllability and observabality in brief.* **GTU : Winter-21, Marks 3**

10.5 Ad Hoc Testable Design Techniques

GTU : Summer-15

- Testability can be increased by making nodes more accessible at some cost by physically inserting more access circuits to the original design.
- Some of the ad hoc testable design techniques are :
  1. Partition-and-Mux Technique
  2. Initialize Sequential Circuit
  3. Disable Internal Oscillators and Clocks
  4. Avoid Asynchronous Logic and Redundant Logic
  5. Avoid Delay-Dependent Logic.

University Question

1. *Discuss ad hoc testable design techniques.* **GTU : Summer-15, 21, Marks 7**

10.6 Scan-Based Techniques

GTU : Winter-14

- The scan technique is a structured approach to designing sequential circuit for testability. The storage cells in registers are used as observation points, control points or both.
- The storage elements are connected to form a long, serial shift register therefore called scan path by using multiplexer and a mode control signal as shown in Fig. 10.6.1. (See Fig. 10.6.1 on next page)
- The scan based technique works in two modes.
  1. **Test mode :** The scan-in signal is clocked into scan path and output of last stage is scanned out.
  2. **Normal mode :** The scan-in path is disabled and the circuit functions as a sequential circuit.

Testing sequence

- The testing sequence is as follows :
  1. Set the mode to test mode and let latches accept data from scan-in input.
  2. Verify the scan path by shifting in and out the test data.

3. Scan in (shift in) the desired state vector into the shift register.
4. Apply the test pattern to primary input.
5. Set mode to normal mode and observe the primary outputs of the circuit after sufficient time for propagation.
6. Assert the circuit clock for one machine cycle to capture the outputs of the combinational logic into the registers.
7. Return to test mode ; scan out the contents of registers, and scan in the next pattern.
8. Repeat testing steps 3 to 7 until all test patterns are applied.

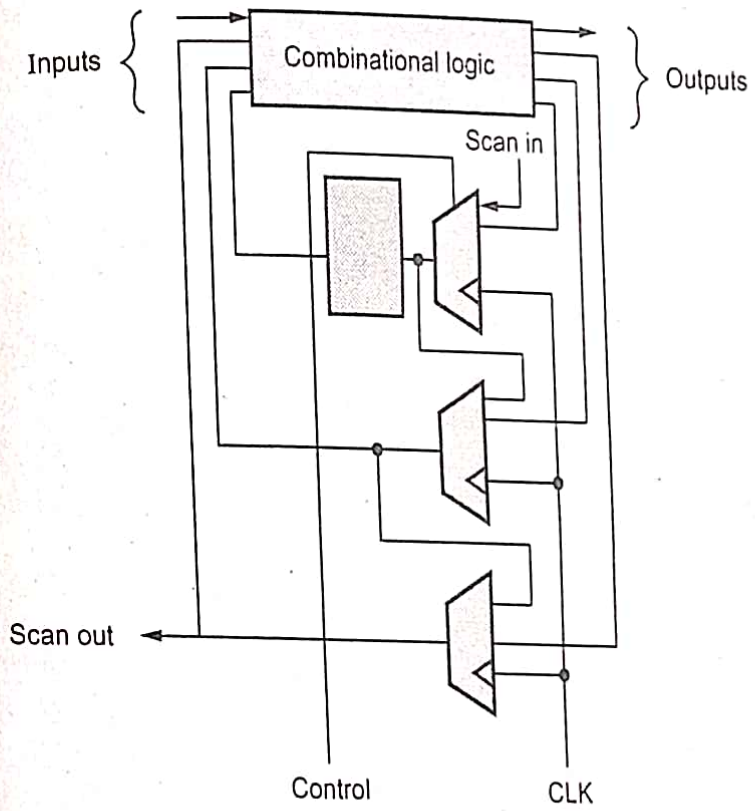


Fig. 10.6.1 Scan based technique structure

10.6.1 Boundary Scan Check

- **Boundary scan check** is a test technique which uses scan methodology involving shift registers. The shift register control monitors signal at each input and output pins that are connected in serial fashion to form a chain of data register called Boundary Scan Registers.
- Printed circuit boards are becoming more dense and complex with use of Surface Mount Devices (SMD). Hence most test system does not guarantee good fault coverage. Boundary scan check technique involves scan path and self tests which resolves above problems.

10.6.2 Merits of Boundary Scan Check

- Boundary scan check has following advantages over other testing methods.
  1. Increased fault coverage in system.
  2. The boundary scan check is much more time efficient. i.e. time required is much less than other systems.
  3. The process of boundary scan check is very simple.
  4. In built scan path and self test makes it more accurate and efficient.

10.6.3 Boundary Scan Standards

- Several standards are specified for boundary scan check testing. Their prime objective is to ensure testing and development less costly and efficient. Some boundary scan standards are mentioned -
  1. Joint Test Action Group [JTAG 1988]
  2. Element Test and Maintenance [ETM VHSIC]
  3. VHSIC Test and Maintenance [TM] Bus Standard.
  4. Testability Bus Standard [IEEE 1989J] / [IEEE 1149.1]

10.6.4 Boundary Scan Test Methodology

- The boundary scan shift register prevents output from rippling as data is shifted through the shift register during scan operation.

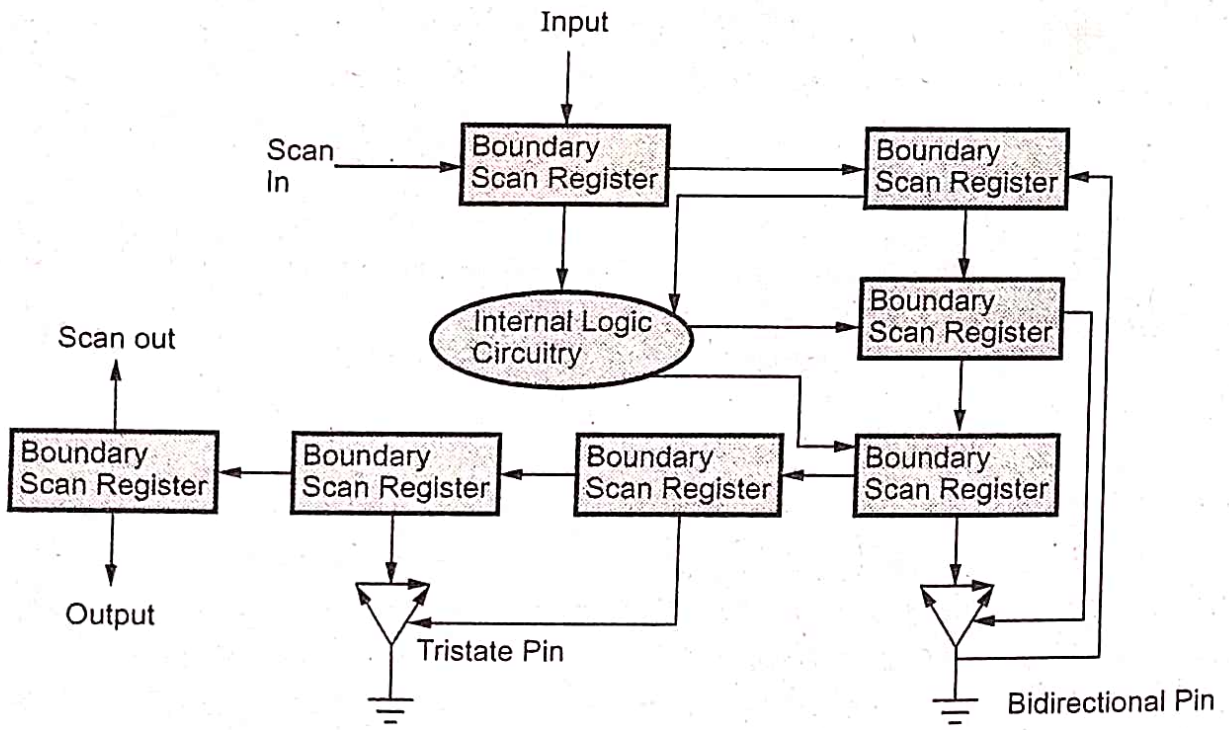


Fig. 10.6.2 Boundary scan test methodology

- The test set is scanned into boundary scan registers (BSR) by using a scan in port with test stimuli input. The response of the circuit is captured in parallel by BSRs in series and scanned out through scan out port.

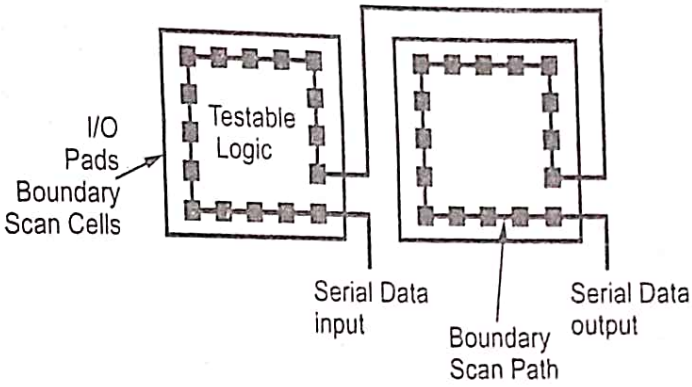


Fig. 10.6.3 Boundary scan architecture

- All the BSRs are interconnected to form a chain around the border of the circuit so as to form as single path as shown in Fig. 10.6.3. It indicates the architecture of IEEE 1149 boundary scan.
- The boundary scan path is having serial input output cells/pads and has appropriate clock pads. The cells or pads are provided for -
  a) Interconnections between chips.
  b) Internal self test.
- Various tests that carried out by the architecture IEEE 1149 are -
  a) Sampling and setting chip Input/Outputs.
  b) Connectivity test between components.
  c) Distribution and collection of self test and built-in test results.

University Question

1. *Draw and explain the general structure of scan-based design technique to improve testability of sequential circuits.* **GTU : Winter-14, Marks 3**

10.7 Built-In Self Test (BIST)

GTU : Summer-15,21, Winter-21

- With the increasing complexity of VLSI systems, test generation and application becomes an expensive and may not be an effective testing.
- Further the high speeds at which newer VLSI systems are designed to operate may not be possible to be simulated and this may create problems. These aspects can be well handled by incorporating BIST which is mainly focused at reducing -
  a. The volume of test data.
  b. Costs involved in test pattern generation. c. Test time.
- All above points can be covered by integrating an automatic test system into the design of chip which is possible by different techniques.

These test include,

1. Linear Feedback Shift Register (LFSR)
2. Built-in Logic Block Observer (BILBO)
3. Signature Analysis.

Advantages of BIST

1. Low cost.
2. High quality testing.
3. Faster fault detection.
4. Ease of diagnostics.
5. Reduced maintenance and repair costs.

- The essential circuit modules required for BIST are :
  1. Pseudo random pattern generator (PRPG)
  2. Output response analyzer (ORA)
- Fig. 10.7.1 shows general procedure for BIST.

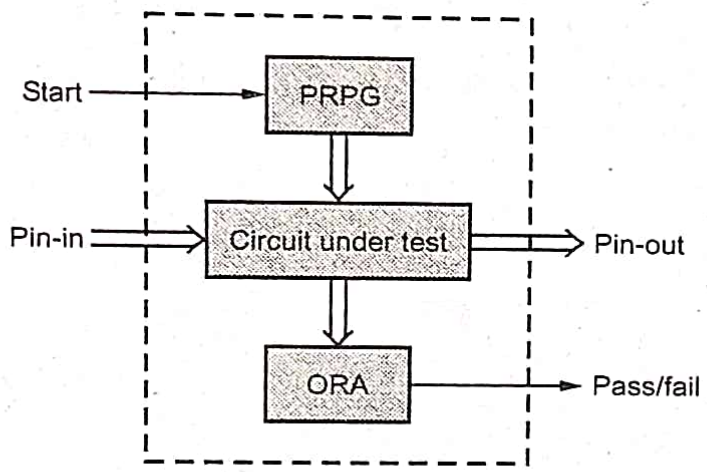


Fig. 10.7.1 BIST process

10.7.1 Pseudo Random Pattern Generator (PRPG)

- A pseudorandom bit generator (PRBG) is defined as a deterministic algorithm which, given a truly random binary sequence of length k, outputs a binary sequence of length 1_k that "appears" to be random.
- The input to the PRBG is called the seed, while the output of the PRBG is called a pseudo random bit sequence.
- A (pseudo) random bit sequence can be constructed by concatenation of (pseudo) random bits. A (pseudo) random number in interval <0, n> can be converted from (pseudo) random bit sequence of length $[\log_2 n] + 1$ (if n is exceeded, new sequence should always be used).

University Questions

1. *Explain Built-In Self Test (BIST) techniques with necessry diagrams.* **GTU : Summer-15, Marks 7**
2. *Explain Built-in Self Test (BIST) technique for circuit testing.* **GTU : Summer-21, Marks 7**
3. *Explain Built in self test (BIST) technique in brief.* **GTU : Winter-21, Marks 4**
4. *Draw block diagram of Built in Self Test (BIST).* **GTU : Winter-21, Marks 3**

10.8 Current Monitoring I_{DDQ} Test

- The I_{DDQ} test is usually used for testing fabrication defects. In bridging fault CMOS circuit draws very high current from power supply.
- The I_{DDQ} test is performed by applying the test vector and then monitoring the current drawn from power supply. This test requires more time but fault detection capability is greatly improves.
- While stuck at tests requires both fault sensitization and fault effect propagation, the I_{DDQ} test requires only fault sensitization. The fault coverage is easy to obtain and offers a full chip coverage capability for large design.

Design guidelines for I_{DDQ} testability

1. Low static current states.
2. No active pull-ups or pull-downs.
3. No internal drive conflicts (drivers share a bus)
4. No floating nodes.
5. No degraded voltages.

10.9 Short Questions and Answers

Q.1 Explain stuck-at-fault.

Ans. : Stuck-at-Fault

- The Stuck-at fault is a logical fault model in which faults are fixed (0 or 1) value to a net. A fault can exist at the input or at the output of a logic gate or a flip-flop in the circuit.
- If the net is stuck to 0, it is called stuck-at-0 (s-a-0) fault and if the net is stuck to 1, it is called stuck-at-1 (s-a-1) fault.

Q.2 Explain about the principle of built in self test.

Ans. : Principle of Built-In-Self-Test (BIST)

- The basic principle of Built-In-Self-Test (BIST) involves pseudo-random sequence generator (PRSG) that produces the input signals for a section of combinational circuitry and a signature analyzer to observe the output signals.

Q.3 Explain about test principles used for testing.

Ans. : Test principles

- Different test principles are incorporated in design and manufacturing of VLSI circuits. For testing a circuit exhaustively large number of test vectors are needed, which is not feasible even at very high speed. Fortunately the number of non-functional nodes on a chip is smaller than the number of states.
- A manufacturing test engineer must device test vectors for clearly detecting any defective node without requiring so many patterns.
- Fig. 10.9.1 illustrates testing of combinational circuit with 2^n test vector (n-inputs).

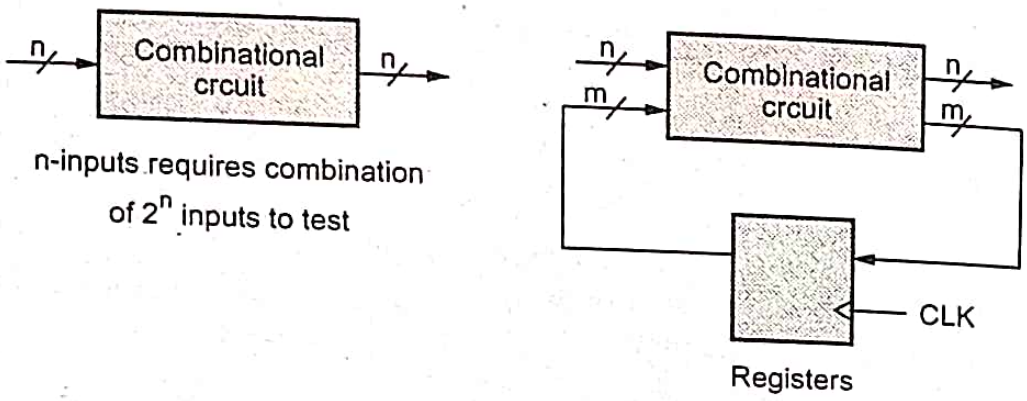


Fig. 10.9.1

- The combinational circuit is converted to sequential circuit by adding m-storage registers.
- The state of sequential circuit is determined by inputs and previous input condition. Minimum test vectors $2^{(m+n)}$ are required to be applied to test this circuit.

Q.4 Differentiate functional simulation and timing simulation.

Ans. :

- Functional simulation simply tests the logic "functional" operation of the circuit. There is no consideration for delay through the internal logic or the routing dealy paths associated with where the placer and the router interconnect things.
- It simply allows you to see that what you think you are coding provides the results you are intending.
- With timing simulation, the delay associated with the logic elements and the interconnect routing are taken into consideration (based on the speed grade of the chip selected).
- Delays may result in signals not meeting setup time, etc. which could explain why your results do not look as expected.

Q.5 What type of faults can be reduced by improving layout design ?

Ans. : Designing of system layout can be well carried out with a view to supporting testing.

- The layout may be designed inclusive of key point test pads or pads to energize special test modes. Layout should be designed with a view to reducing likelihood of short and open circuits.
- Experience on MOS circuits has shown frequent occurrence of short circuits in the diffusion layer and that of open circuits in the metal layer.
- This calls for special precautions in layout designing keeping these aspects in mind. The occurrence of such and other faults can be avoided by careful observance of design rules and ground rules.

Q.6 Define controllability and observability with respect to testing.

Ans. : Controllability

- Controllability is the ability of an internal circuit within a chip the ease of setting the node to logic 1 or 0 state.
- Controllability is important while accessing the degree of difficulty of testing a particular signal in a circuit.

Observability

- Observability is a degree to which it can be observed that node at output operates correctly. Observability is useful, when a test engineer has to measure the output of a gate/chip within a larger circuit.
- To check its correct operation. Higher observability indicates less number of cycles required to measure output node value.

Q.7 What is the difference between verification and validation ?

Ans. :

Sr. No.	Verification	Validation
1.	Verification is a static practice of verifying documents, design code and program.	Verification is a dynamic mechanism of validating and testing the actual product.
2.	It does not involve executing the code.	It always involves executing the code.
3.	It is human based checking of documents and files.	It is computer based execution of program.

4.	Verification uses methods like inspections, reviews, walkthroughs, and desk checking etc.	Validation uses methods like black box (functional) testing, gray box testing and white box (structural) testing etc.
5.	Verification is to check whether the software confirms to specifications.	Validation is to check whether software meets the customer expectations and requirements.
6.	It can catch errors that validation cannot catch. It is low level exercise.	It can catch errors that verification cannot catch. It is high level exercise.

Q.8 What are the different chip - level test techniques ?

Ans. : Chip level test techniques

1. Fault models : Stuck at faults, Stuck-open or Stuck-Shut-Fault
2. Scan path : Full pcan, Partial scan
3. Boundray scan check-JTAG, TAP controller, BIST

Q.9 Explain VLSI verification and testing.

Ans. :

- **VLSI Verification** is done before manufacturing. Before even tapeout. This is done for verifying if the chip design is working as expected.
- Example: If we have a counter design in verilog, We can simulatethe verilog file and verify if the sequence is correct. This is functionality check.
- **VLSI Testing** is done after manufacturing. After the chips are made, we will look for any structural damages or mistakes in the chip. At this stage, we will check if the chip passes the test. If it does not, we throw away the chip. If it passes, we can use(sell) the chip.
- To do testing, we have to put some extra special logic into the chip before it is taped out. This is called Design For Testing. (DFT).

❑❑❑