1 UNIT

Introduction

CONTENTS

PART-1

VLSI Design Flow.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.1. Draw and explain the VLSI design flow (Y-chart).

AKTU 2018-19, Marks 05

OR

Draw the Y-chart and explain the VLSI design process. Mention its advantages.

AKTU 2020-21, Marks 07

Answer

A. Y-chart :

1. The chart was first introduced by D. Gajski. Y-chart illustrates a simplified design flow for most logic chips, using design activities on three different domains which resemble the letter Y.

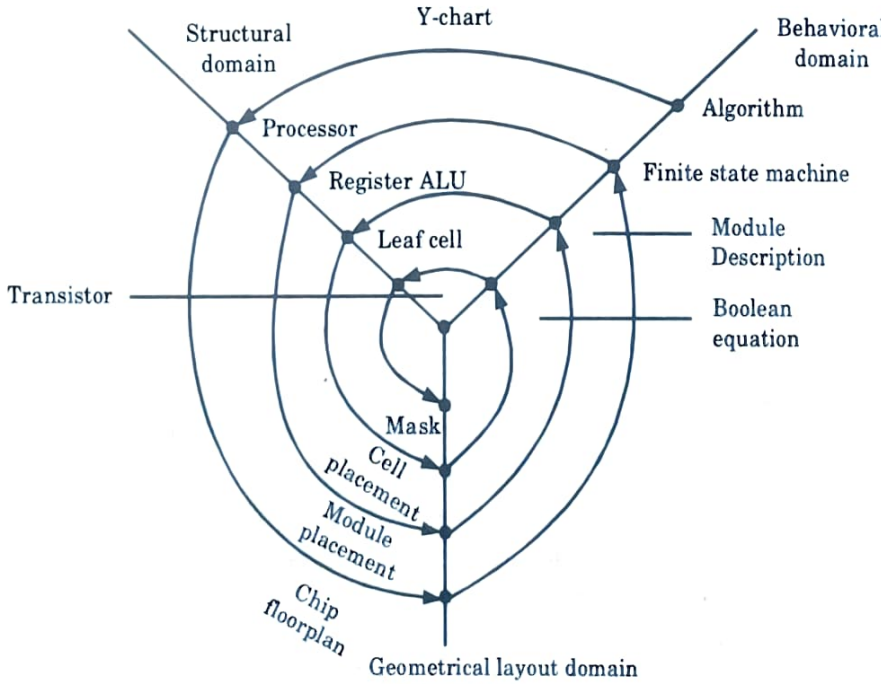


Fig. 1.1.1.

2. The Y-chart consists of three domains of representation namely behavioral, structural and geometrical layout domain.
3. The design flow starts from algorithm that describes the behavior of the target chip.
4. The corresponding architecture of the processor is then defined. It is mapped onto the chip surface by floor-planning as shown in Fig. 1.1.2.

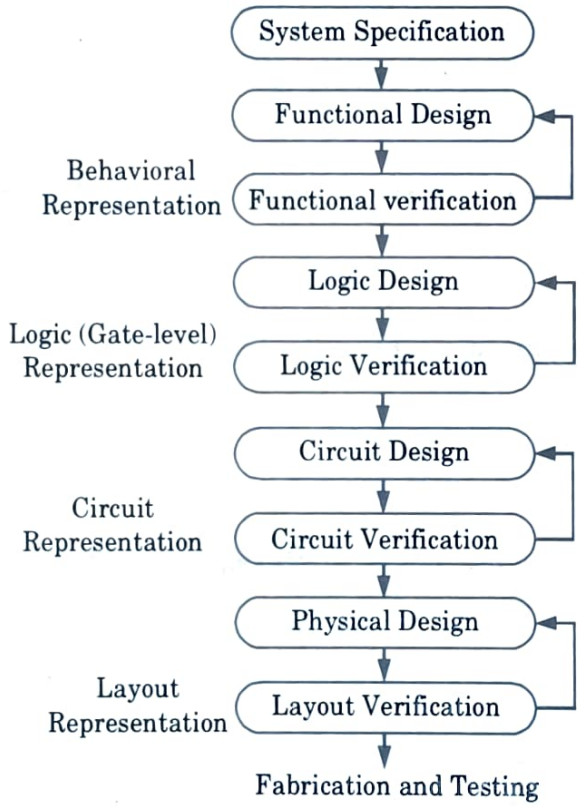


Fig. 1.1.2.

5. The next design evolution defines FSM implemented with registers and ALU. These modules are then geometrically placed onto chip surface using CAD tools.
6. The third evolution starts with behavioral module description. Individual modules are then implemented with leaf cells. Logic gates (leaf cells) can be placed by using a cell placement and routing program.
7. The last stage involves implementation of leaf and mask generation. Fig. 1.1.2 below provides a more simplified view of the VLSI design flow.
8. Note that the verification of design plays a very important role in every step during this process.
9. The failure of design during verification causes expensive re-design at latter stage and results in increased time to market.

10. Although top-down design flow appeals for design process, but in reality there is no uni-directional top-down design flow. Both top-down and bottom-up approaches have to be combined for a successful design.

B. Advantages :

1. Reduced size for circuits.
2. Increased cost-effectiveness for devices.
3. Improved performance in terms of the operating speed of circuits.
4. Requires less power than discrete components.
5. Higher device reliability.
6. Requires less space and promotes miniaturization.

PART-2

General Design Methodologies : Critical Path and Worst Case Timing Analysis.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.2. Explain critical path with example.

Answer

A. Critical path :

1. The critical path is defined as the path between an input and an output with the maximum delay. Once the circuit timing has been computed, the critical path can easily be found by using a traceback method.
2. Critical paths are timing-sensitive functional paths, because of the timing of these paths is critical, no additional gates are allowed to be added to the path, to prevent increasing the delay of the critical path.

B. Example :

1. The STA (Static Timing Analysis) tool will add the delay contributed from all the logic connecting the Q output of one flip-flop to the D input of the next and then compare it against the defined clock period of the CLK pins. This should be strictly less than the clock period defined for that clock.
2. If the delay is less than the clock period, then the path meets timing. If it is greater, then the path fails timing.
3. The critical path is the path out of all the possible paths that either exceeds its constraint by the largest amount, or if all paths pass, then the one that comes closest of failing.

Que 1.3. **Write a short note on worst case timing analysis.**

Answer

1. The worst-case timing analysis (WCTA) portion of WCCA analyzes the timing of digital devices and signal paths under worst-case conditions.
2. It usually accompanies many other assessments such as decoupling, signal integrity, and DC compatibility.
3. For WCTA, Timing Designer analyzes the results of the Printed-Circuit Board (PCB) propagation delay, extracted from HyperLynx and the static timing numbers of the FPGA or ASIC.
4. Through WCTA, setup and hold time violations can be analyzed.

PART-3

Overview of Design Hierarchy.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.4. **Write a short note on design hierarchy, regularity, modularity and locality.**

OR

Explain the concept of design hierarchy with the help of example.

AKTU 2017-18, Marks 10

Answer

A. Design hierarchy :

1. The use of hierarchy involves dividing module into sub-modules until the complexity of the smaller parts becomes manageable.
2. Similarly, the design of VLSI chip can be represented in three domains.
3. It is important for the simplicity of design that the hierarchies in different domains be mapped. Fig. 1.4.1 shows the structural decomposition of a CMOS 2 bit adder into its components.
4. Adder can be decomposed into 1 bit adders, separate carry and sum circuits and individual logic gates.
5. The above partitioning provides a valuable guidance for realization of these blocks on the chip. The approximate shape and area of each sub-module should be estimated in order to provide a useful floor place.

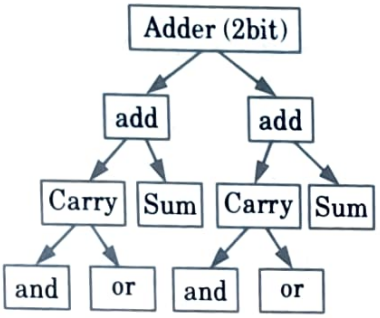


Fig. 1.4.1.

B. Regularity :

1. Regularity means that the hierarchical decomposition of a large system should result in simple as well as similar blocks as much as possible.
2. Regularity can exist at all level of abstraction. For example, design of array structures consisting of identical cells at transistor level and at gate level.
3. If the designer has a small library of basic building blocks, a number of different functions can be constructed by using this principle.
4. Regularity usually reduces the number of different modules that need to be designed and verified, at all levels of abstraction.

C. Modularity :

1. Modularity in design means that the various functional blocks which make up the large system must have well-defined functions and interfaces.
2. Each block or module can be designed independent from each other such flexibility is provided by modularity. All the blocks can be combined easily at the end of the design process to form the large system.
3. The concepts of modularity enable the parallel process during the design.

D. Locality :

1. The concept of locality ensures that connections between module to module are mostly between neighbouring modules, avoiding distance connections as much as possible.
2. During interfacing each module in the system, make sure that the internals of each module becomes unimportant to the exterior module.
3. Far-interconnections can be avoided as they provide long delays in the system. All the time-delays operations should be performed locally, without the need to access far placed modules.

PART-4

Layers of Abstraction, Integration Density and Moore's Law.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.5. **What are the levels of abstraction ? Explain.**

Answer

The levels (layers) of abstraction are :

i. **Behavioral or algorithmic level :** This is the highest level of abstraction. A module can be implemented in terms of the design algorithm. The designer no need to have any knowledge of hardware implementation.

ii. **Data flow level :** In this level, the module is designed by specifying the data flow. Designer must how data flows between various registers of the design.

iii. **Gate level :** The module is implemented in terms of logic gates and interconnections between these gates. Designer should know the gate-level diagram of the design.

iv. **Switch level :** This is the lowest level of abstraction. The design is implemented using switches/transistors. Designer requires the knowledge of switch-level implementation details.

Que 1.6. **What do you mean by high integration density? Also explain the different scales of integration ?**

Answer

A. **High integration density :** A high integration density means a large number of components in a small area.

B. **Different scales of integration :**

Depending upon the number of active devices per chip, there are different levels of integration :

i. **SSI :**

When the active devices per chip are less than 100, then it is referred as small scale integration (SSI). Most of the SSI chips use integrated resistors, diodes and bipolar transistors.

ii. **MSI :**

When the count of active devices per chip is between 100 and 1000, then it is referred as medium scale integration (MSI). In most of the MSI chips, BJTs and enhancement mode MOSFETs are integrated.

iii. LSI :

In large scale integration (LSI), the number of active devices per chip ranges between 1000 and 100,000. In general, LSI chips use MOS transistors; as it requires less number of steps for integration. Thus more number of components can be produced on the chip with MOS transistors than with the bipolar transistors.

iv. VLSI :

When the active devices per chip are over hundreds of thousands, then it is referred as very large scale integration (VLSI). Almost all modern chips employ VLSI technique.

v. ULSI :

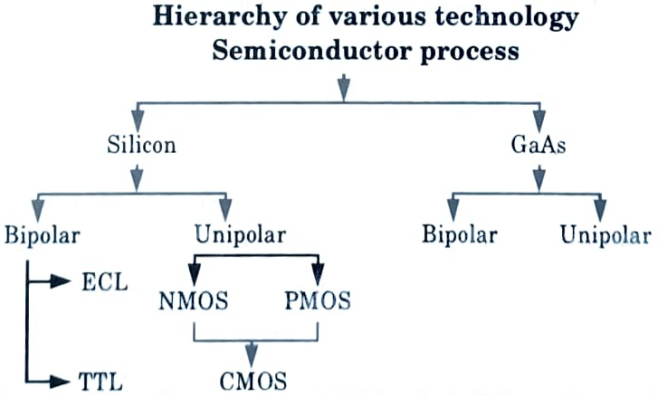
Recently a new level of integration has been introduced which is known as ultra large scale integration (ULSI). In ULSI technique, more than one million active devices are integrated on a single chip. Pentium microprocessors use ULSI technology.

Que 1.7. Discuss the hierarchy of various semiconductors with Moore's law. Draw the *Y*-chart and explain the VLSI design process.

AKTU 2019-20, Marks 07

Answer

A. Hierarchy of various semiconductors with Moore's law :



1. The IC was invented in February 1959 by Jack Kilby of Texas Instrument. The planar version of the IC was developed by Robert Noyce at Fairchild in July 1959. Since then the evolution of this technology has been extremely fast paced.
2. One way to gauge the process of the field is to look at the complexity of IC's as a function of time.
3. When we plot log of the component count as a function of time, we get a straight line, indicating that there has been an exponential growth in the complexity of chips over three decades.

4. The component count has roughly doubled every 18 months, as was noted early by Gordon Moore. This regular doubling is known as Moore's law.
5. The main factor that has enabled this increase of complexity is the ability to shrink or scale devices.
6. Clearly, one can pack a larger number of components with greater functionality on an IC if they are smaller; also being advantageous in terms of faster IC's which consume less power.
7. Fig. 1.7.1 shows the level of integration versus time for memory chips.

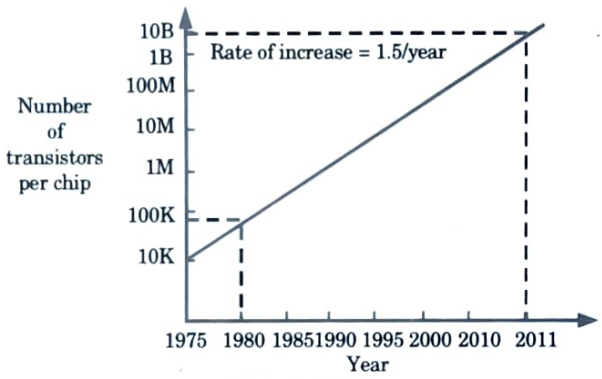


Fig. 1.7.1.

8. It can be observed that in terms of transistor count, logic chips contain significantly fewer transistors in any given years mainly due to large consumption of chip area for complex interconnects.
9. Memory circuits are highly regular and thus more cells can be integrated with much less area for interconnects.
10. The demand for digital CMOS IC's will continue to be strong due to salient features such as low power, reliable performance, circuit techniques for high speed such as using dynamic circuits.
11. It is now projected that the minimum feature size in CMOS ICs can decrease up to 35 nm within a decade.
12. Bipolar and gallium arsenide (GaAs) circuits have been used for very high speed circuits, and this practice may continue. MMIC's, GaAs MESFET technology has been highly successful.
13. As long as the downward scaling of CMOS technology remains strong, other technologies are likely to remain the technology of tomorrow.

B. *Y-chart and VLSI design process* : Refer Q. 1.1, Page 1–2F, Unit-1.

PART-5

VLSI Design Styles.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.8. What are the VLSI design styles ? Explain any one.

Answer

The VLSI design styles are :

i. Gate array design

ii. Standard-cells based design

iii. Full custom design

iv. Field Programmable Gate Array (FPGA) :

1. Fully fabricated FPGA chips containing thousands of logic gates or even more, with programmable interconnects, are available to users for their custom hardware programming to realize desired functionality.
2. This design style provides a means for fast prototyping and also for cost-effective chip design, especially for low-volume applications.
3. A typical FPGA chip consists of I/O buffers, an array of configurable logic blocks (CLBs), and programmable interconnect structures.
4. The programming of interconnects is implemented by programming of RAM cells whose output terminals are connected to the gates of MOS pass transistors. A general architecture of FPGA from XILINX is shown in Fig. 1.8.1.

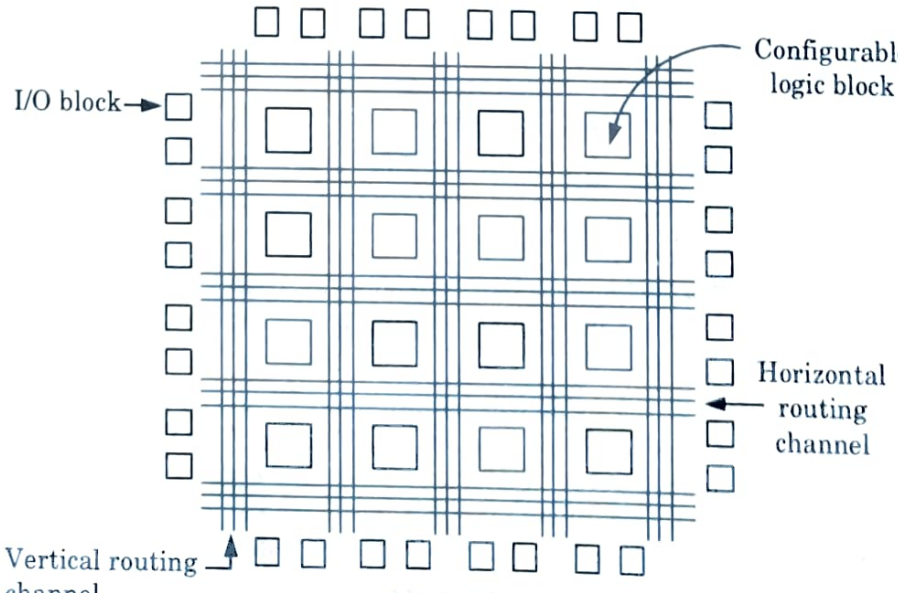


Fig. 1.8.1.

5. A simple CLB is shown in Fig. 1.8.2. It consists of four signal input terminals (A, B, C, D), a clock signal terminal, user-programmable multiplexers, an SR-latch, and a look-up table (LUT).

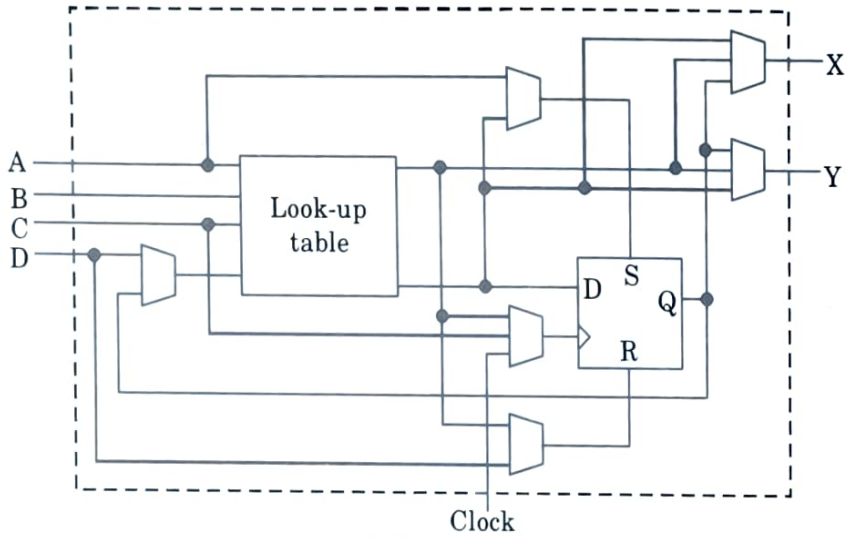


Fig. 1.8.2.

6. The LUT is a digital memory that stores the truth table of the boolean function. Thus, it can generate any function of upto four variables or any two functions of three variables.
7. The CLB is configured such that many different logic functions can be realized by programming its array.

Que 1.9. Write a short note on gate array design.

Answer

1. Gate array implementation requires a two-step manufacturing process. The first phase, which is based on generic (standard) masks, results in an array of uncommitted transistors on each GA chip.
2. These uncommitted chips can be stored for later customization, which is completed by defining the metal interconnects between the transistors of the array.
3. Since the patterning of metallic interconnects is done at the end of the chip fabrication, the turn-around time can be still short, a few days to a few weeks.
4. A corner of a gate array chip which contains bonding pads on its left and bottom edges, diodes for I/O protection, nMOS transistors and pMOS transistors for chip output driver circuits in the neighboring areas of bonding pads, arrays of nMOS transistors and pMOS transistors, underpass wire segments, and power and ground buses along with contact windows.

5. Typical gate array platforms allow dedicated areas, called channels, for intercell routing between rows or columns of MOS transistors.
6. The availability of these routing channels simplifies the interconnections, even using one metal layer only.
7. The interconnection patterns that perform basic logic gates can be stored in a library, which can then be used to customize rows of uncommitted transistors according to the netlist.
8. In general, the GA chip utilization factor, as measured by the used chip area divided by the total chip area, is higher than that of the FPGA and so is the chip speed, since more customized design can be achieved with metal mask designs.
9. The current gate array chips can implement as many as of thousands of logic gates.

Que 1.10. Explain standard-cells based design.

Answer

1. The standard-cells based design is one of the most prevalent full custom design styles which require development of a full custom mask set.
2. The standard cell is also called the polycell.
3. In this design style, all of the commonly used logic cells are developed, characterized, and stored in a standard cell library.
4. A typical library may contain a few hundred cells including inverters, NAND gates, NOR gates, complex AOI, OAI gates, *D*-latches, and flip-flops.
5. The characterization of each cell is done for several different categories. It consists of :

i. Delay time versus load capacitance
ii. Circuit simulation model
iii. Timing simulation model
iv. Fault simulation model
v. Cell data for place-and-route
vi. Mask data.

6. 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.
7. The power and ground rails typically run parallel to the upper and lower boundaries of the cell, thus, neighboring cells share a common power and ground bus.
8. The input and output pins are located on the upper and lower boundaries of the cell.

Que 1.11. **Discuss full custom design in brief.**

Answer

1. In a full custom design, the entire mask design is done anew without use of any library. However, the development cost of such a design style is becoming prohibitively high.
2. Thus, the concept of design reuse is becoming popular in order to reduce design cycle time and development cost.
3. The most rigorous 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.
4. For logic chip design, a good compromise can be achieved by using a combination of different design styles on the same chip, such as standard cells, data-path cells and PLAs.
5. 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.

PART-6

Packaging.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.12. **Define the following method :**

i. PTH packages
ii. SMT packages

OR

What are the types of common IC packages ? Explain.

Answer

A. **PTH packages :** The package pins can be introduced in holes drilled in the PCB which is called pin-through-hole (PTH) method.

B. **SMT packages :** The package pins can be directly soldered on the PCB which is called surface-mounted technology (SMT) method.

C. Some common IC package types are :

i. Dual in-line packages (DIP) :

1. Dual in line packages (DIP) are characterized by their high interconnect inductances, which can lead to significant noise problems in high-frequency applications.
2. The maximum pin count of DIP is typically limited to 64.
3. DIP has the advantage of low cost but their dimensions can be prohibitive, especially for small, portable product.

ii. Pin grid array (PGA) packages :

1. This PTH package type offers a higher pin count and higher thermal conductivity compared to DIPs, especially when a passive or active heat sink is attached on the package.
2. The PGA packages require a large PCB area, and the package cost is higher than DIP, especially for ceramic PGAs.

iii. Chip carrier packages (CCP) :

1. This SMT package type is available in two variations, the leadless chip carrier and the leaded chip carrier.
2. The leadless chip carrier is designed to be mounted directly on the PCB, and it can support a high pin count.
3. The leaded chip carrier package solves this problem since the added leads can accommodate small dimension variations caused by the difference in the thermal coefficients.

iv. Quad flat packs (QFP) :

1. This SMT package type is similar to leaded chip carrier packages, except that the leads extend outward rather than being bent under the package body.
2. Ceramic and plastic QFPs with very high pin counts are becoming popular package types in recent years.

v. Multi-chip modules (MCM) :

1. This IC package option can be used for special applications requiring very high performance, where multiple chips are assembled on a common substrate contained in a single package.
2. Thus, a large number of critical interconnections between the chips can be made within the package.

PART-7

CMOS Logic.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.13. **Draw and explain the working of CMOS inverter with its transfer characteristics.** **AKTU 2018-19, Marks 10**

Answer

CMOS inverter :

1. Fig. 1.13.1 shows the CMOS inverter. In the CMOS inverter, the PMOS and NMOS devices Q_P and Q_N are driven simultaneously by an input V_{in}.

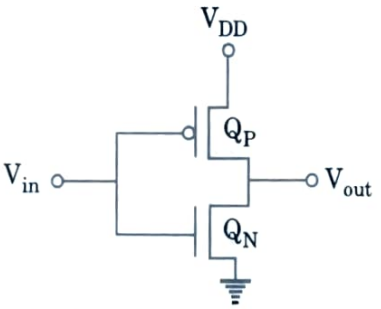


Fig. 1.13.1 CMOS inverter circuit.

Circuit operation :

1. When input is high ($\simeq V_{DD}$) Q_N is made to conduct, while Q_P is forced to cut-off. This causes the output becomes low ($V_o = 0$) as shown in Fig. 1.13.2.

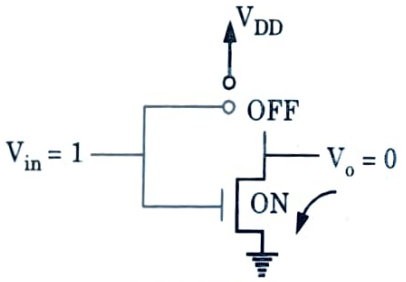


Fig. 1.13.2.

2. When input is low (≈ 0 V) Q_N is OFF and Q_P becomes ON, therefore the output becomes logic high ($V_o = V_{DD}$) as shown in Fig. 1.13.3.

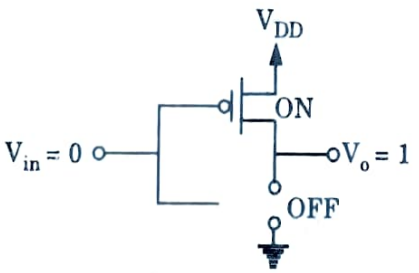


Fig. 1.13.3.

3. Table 1.13.1 shows the operation of COMS inverter circuit.

Table 1.13.1.

V_{in}	Q_P	Q_N	V_o
0	ON	OFF	1
1	OFF	ON	0

Voltage transfer characteristics :

1. As shown in Fig. 1.13.4, the more positive output voltage corresponds to a logic 1 is $V_{OH} = V_{DD}$, and the more negative output voltage corresponds to a logic 0 is $V_{OL} = 0$.
2. When output is in the logic 0 state, the PMOS transistor is cut-off and when the output is in the logic 1 state, the NMOS transistor is cut-off.

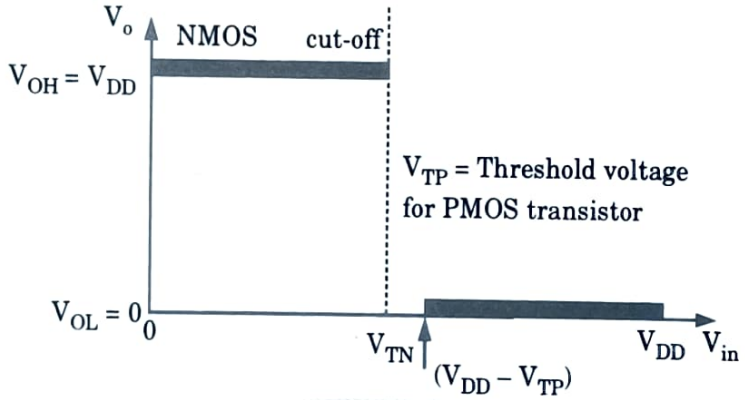


Fig. 1.13.4.

Que 1.14. Sketch a 2-input CMOS NAND gate and also discuss its working.

Answer

1. Fig. 1.14.1(*a*) shows a 2-input CMOS NAND gate. It consists of two series nMOS transistors between *Y* and GND and two parallel pMOS transistors between *Y* and V_{DD}.
2. If either input *A* or *B* is '0,' at least one of the nMOS transistors will be OFF, breaking the path from *Y* to GND. But at least one of the pMOS transistors will be ON, creating a path from *Y* to V_{DD}. Hence, the output *Y* will be '1'.
3. If both inputs are '1', both of the nMOS transistors will be ON and both of the pMOS transistors will be OFF. Hence, the output will be '0'.
4. The truth table is given below and the symbol is shown in Fig. 1.14.1(*b*).

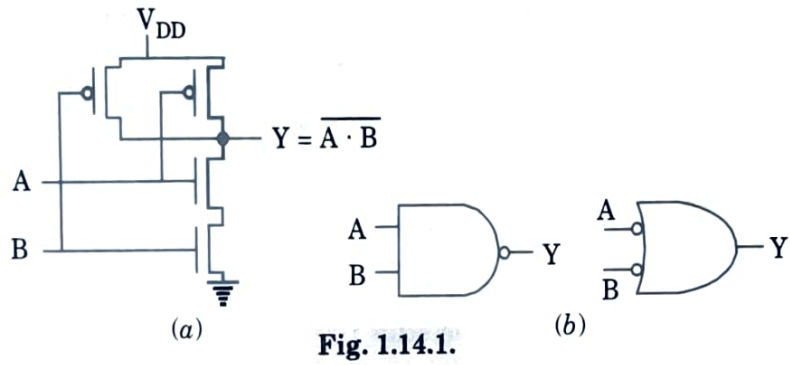


Fig. 1.14.1.

NAND gate truth table

A	B	pull-down network	pull-up network	Y
0	0	OFF	ON	1
0	1	OFF	ON	1
1	0	OFF	ON	1
1	1	ON	OFF	0

Que 1.15. Write a short note on combinational logic.

Answer

1. The inverter and NAND gates are examples of complementary CMOS logic gates, also called static CMOS gates.
2. In general, a fully complementary CMOS gate has an nMOS pull-down network to connect the output to '0' (GND) and pMOS pull-up network to connect the output to '1' (V_{DD}) as shown in Fig. 1.15.1. The networks are arranged such that one is ON and the other OFF for any input pattern.
3. The NAND gate used a series pull-down network and a parallel pull-up network.

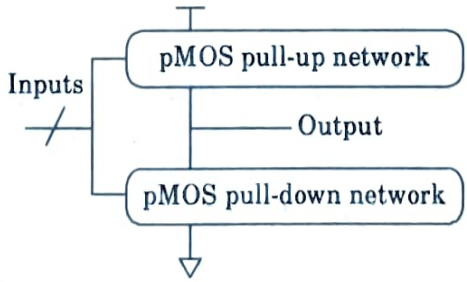


Fig. 1.15.1. General logic gate using pull-up and pull-down network.

4. Two or more transistors in series are ON only if all of the series transistors are ON. Two or more transistors in parallel are ON if any of the parallel transistors are ON. This is shown Fig. 1.15.2 for nMOS and pMOS transistor pairs.
5. By using combinations of these constructions, CMOS combinational gates can be constructed.

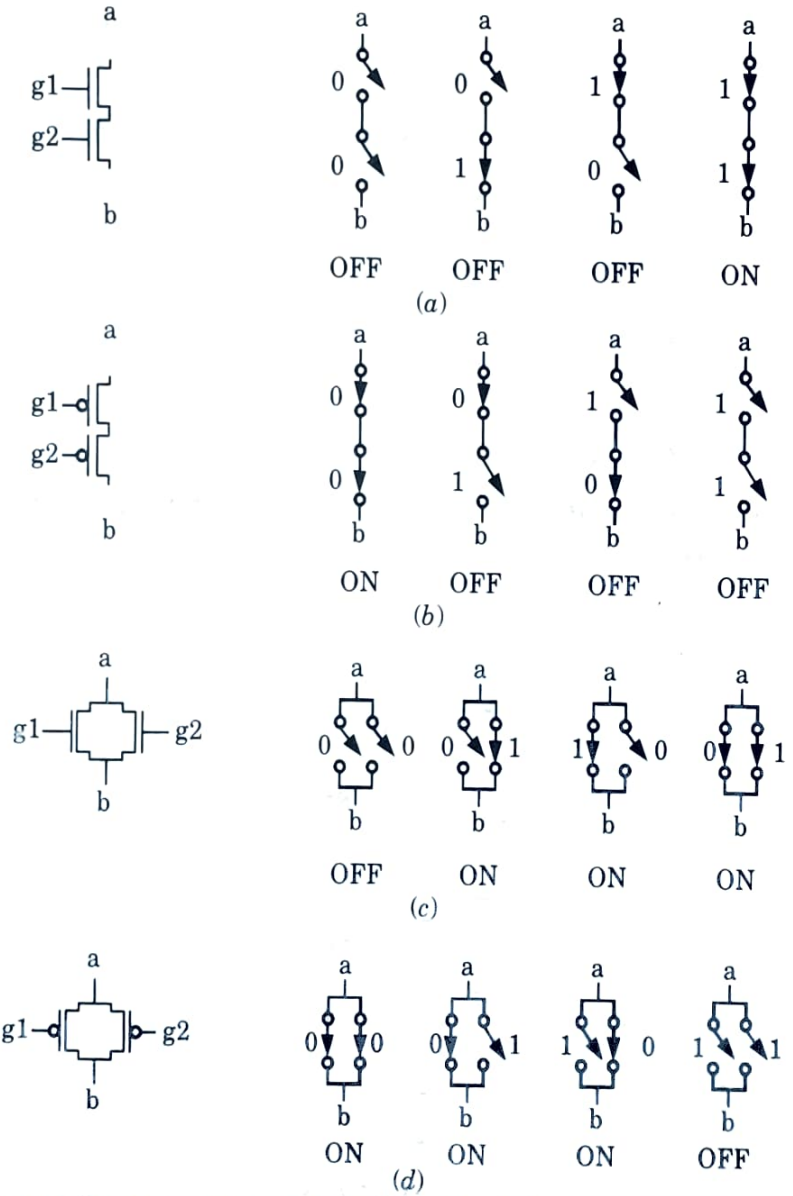


Fig. 1.15.2. Connection and behaviour of series and parallel transistors.

6. The possible levels at the output are shown in Table 1.15.1. The '1' and '0' levels have been encountered with the inverter and NAND gates, where either the pull-up or pull-down is OFF and the other structure is ON.

7. When both pull-up and pull-down are OFF, the high-impedance or floating Z output state results. This is of importance in multiplexers, memory elements and bus drivers.
8. The crowbarred X level exists when both pull-up and pull-down are simultaneously turned ON.
9. This causes an indeterminate level and also static power to be dissipated. It is usually an unwanted condition in any CMOS digital circuit.

Table 1.15.1. Output states of CMOS logic gate

	pull-up OFF	pull-up ON
pull-down OFF	Z	1
pull-down ON	0	crowbarred (X)

Que 1.16. Implement the CMOS logic for the following Boolean expression.

i. $Y = (A + B + C) \cdot D$

ii. $Y = (A + B + C)(D + E) \cdot F$

iii. **3 input NOR gate**

AKTU 2017-18, Marks 10

Answer

i. $Y = (A + B + C) \cdot D$

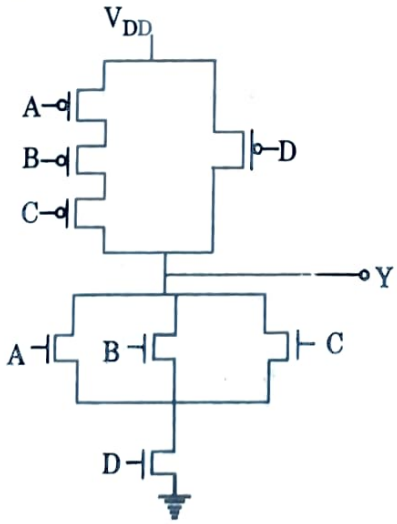


Fig. 1.16.1.

ii. $Y = (A + B + C)\,(D + E) \cdot F$

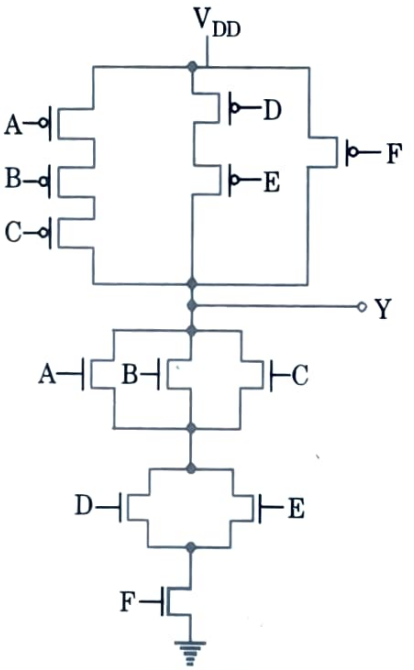


Fig. 1.16.2.

iii. 3 input NOR gate :

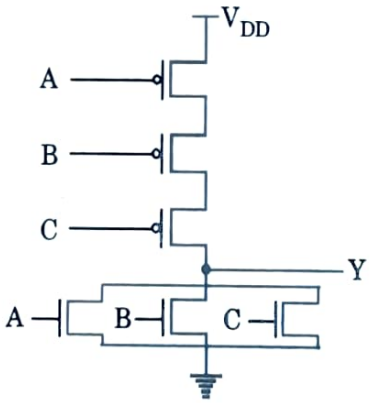


Fig. 1.16.3.

Que 1.17. **Implement the Boolean expression**

$$Y = AB + (C + D)\,(E + F) + (G + H).$$

AKTU 2020-21, Marks 07

Answer

Given, $Y = AB + (C + D)\,(E + F) + (G + H)$

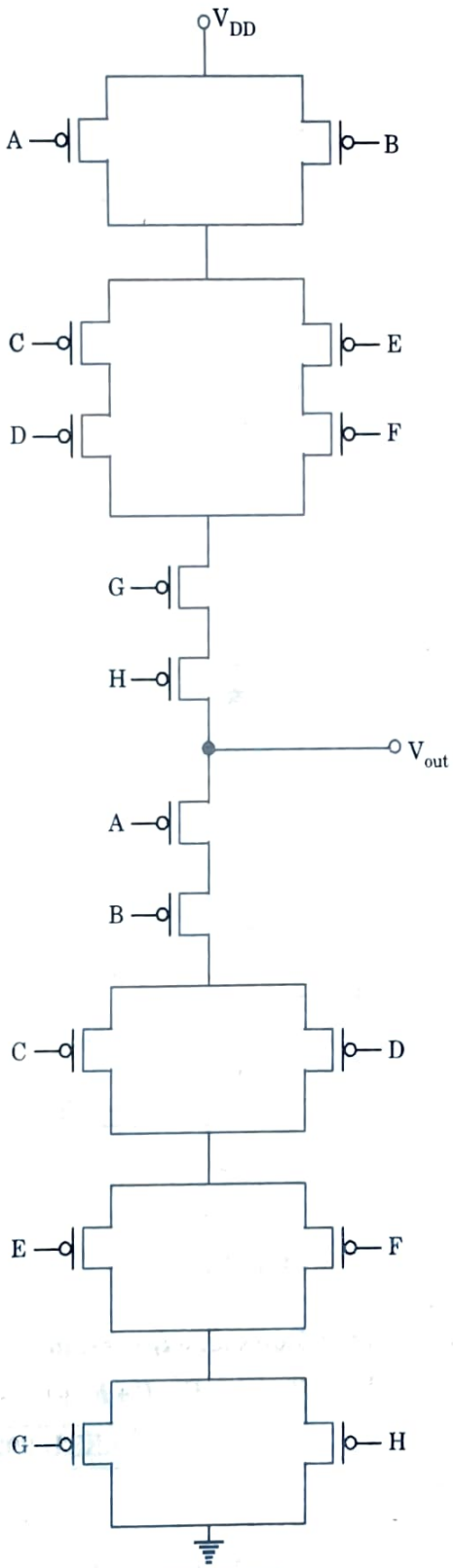


Fig. 1.17.1.

PART-8

Propagation Delay Definitions.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.18. Explain propagation delay time.

Answer

1. The propagation delay time τ_{PHL} and τ_{PLH} determine the input-to-output signal delay during the high-to-low and low-to-high transitions of the output, respectively.
2. By definition, τ_{PHL} is the time delay between the $V_{50\%}$-transition of the rising input voltage and the $V_{50\%}$-transition of the falling output voltage.
3. Similarly, τ_{PHL} is defined as the time delay between the $V_{50\%}$-transition of the falling input voltage and the $V_{50\%}$-transition of the rising output voltage.

PART-9

Sheet Resistance.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.19. Write a short note on sheet resistance.

Answer

1. Consider a uniform slab of conducting material of resistivity ρ, of width W, thickness t, and length between faces l. The arrangement is shown in Fig. 1.19.1.
2. Consider the resistance R_{AB} between two opposite faces.

$$R_{AB} = \frac{\rho l}{A}$$

where, A = Cross-section area

3. Thus $R_{AB} = \frac{\rho l}{tW}$ ohm $[\because A = tW]$

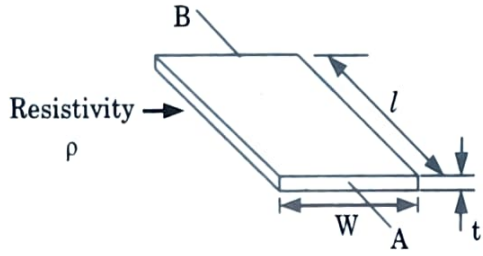


Fig. 1.19.1. Sheet resistance model.

4. If $l = W$ *i.e.*, a square of resistive material, then

$$R_{AB} = \frac{\rho}{t}$$

where R_s = Ohm per square of sheet resistance

Thus $R_s = \frac{\rho}{t}$ ohm per square

5. R_s is completely independent of the area of the square.

2
UNIT

Interconnect Parameters

CONTENTS

PART-1

Resistance, Inductance and Capacitance.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.1. How to compute the resistance of wire or interconnect ?

Answer

1. The resistance of a uniform slab of conducting material can be expressed as

$$R = \frac{\rho}{t}\frac{l}{w} \qquad ...(2.1.1)$$

where ρ is the resistivity. This expression can be rewritten as

$$R = R_s \frac{l}{w} \qquad ...(2.1.2)$$

where, $R_s = \frac{\rho}{t}$ is the sheet resistance and has units of Ω/square.

2. To obtain the resistance of a conductor on a layer, multiply the sheet resistance by the ratio of length to width of the conductor.
3. The resistivity of thin metal films used in wires tends to be higher because of scattering off the surfaces and grain boundaries.
4. As shown in Fig. 2.1.1, copper must be surrounded by a lower-conductivity diffusion barrier that effectively reduces the wire cross-sectional area and hence raises the resistance. Moreover, the polishing step can cause dishing that thins the metal.
5. If the average barrier thickness is $t_{barrier}$ and the height is reduced by t_{dish}, the resistance becomes,

$$R = \frac{\rho}{(t - t_{dish} - t_{barrier})}\frac{l}{(w - 2t_{barrier})} \qquad ...(2.1.3)$$

6. The resistivity of polysilicon, diffusion, and wells is significantly influenced by the doping levels.
7. Polysilicon and diffusion typically have sheet resistances under 10 Ω/square when silicided and up to several hundred Ω/square when unsilicided.
8. Wells have lower doping and thus even higher sheet resistance. Large resistors are often made from wells or unsilicided polysilicon.

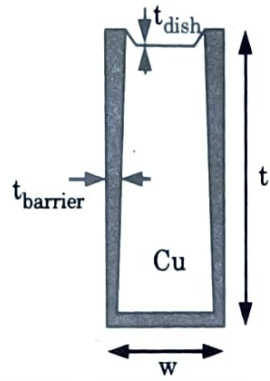


Fig. 2.1.1. Copper barrier layer and dishing.

9. Contacts and vias also have a resistance, which is dependent on the contacted material and size of the contact.
10. Multiple contacts should be used to form low-resistance connections, as shown in Fig. 2.1.2.
11. When current turns at a right angle or reverse, a square array of contacts is generally required, while fewer contacts can be used when the flow is in the same direction.

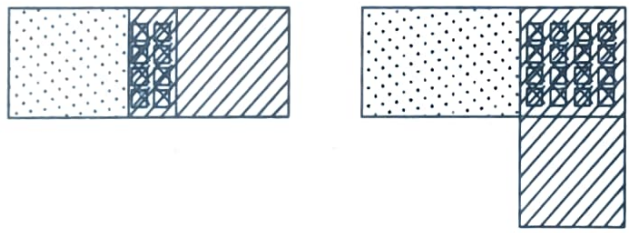


Fig. 2.1.2. Multiple bias for low-resistance connections.

Que 2.2. **How to model the capacitance of wires ?**

OR

Mention the advantages of interconnect modeling. Describe the interconnect modeling to calculate *R* and *C*.

AKTU 2020-21, Marks 07

OR

What are interconnecting models ? Explain any two of them in brief.

AKTU 2017-18, Marks 10

Answer

A. Interconnect models : The wires linking the transistors together are called interconnect. Interconnect models are :

i. Resistor

ii. Capacitor

iii. Inductor.

B. Advantages of interconnect modeling : The interconnect modeling can be used efficiently during high-level design space exploration, interconnect-driven design planning, and synthesis- and timing-driven placement to ensure design convergence for deep sub-micrometer designs.

C. Capacitance model :

1. The wire capacitance has two major components: the parallel plate capacitance of the bottom of the wire to ground and the fringing capacitance arising from fringing fields along the edge of a conductor with finite thickness.
2. In addition, a wire adjacent to a second wire on the same layer can exhibit capacitance to that neighbour. These effects are illustrated in Fig. 2.2.1.
3. The parallel plate capacitance formula is

$$C = \frac{\varepsilon_{ox}}{h} wl \qquad ...(2.2.1)$$

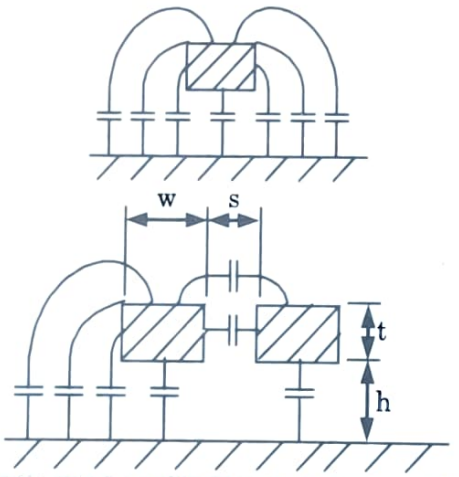


Fig. 2.2.1. Effect of fringing fields on capacitance.

4. The fringing capacitance is more complicated to compute. One intuitively appealing approximation treats a lone conductor above a ground plane as a rectangular middle section with two hemispherical end caps, as shown in Fig. 2.2.2.

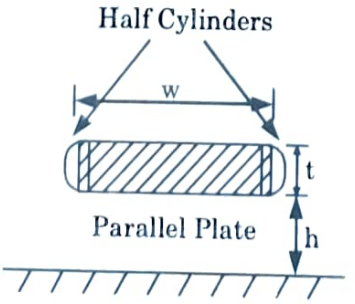


Fig. 2.2.2. Yuan and trick capacitance model including fringing fields.

5. The total capacitance is assumed to be the sum of a parallel plate capacitor of width $w - t/2$ and a cylindrical capacitor of radius $t/2$.
6. This results in an expression for the capacitance that is accurate within 10 % for aspect ratios less than 2 and $t = h$.

$$C = \varepsilon_{ox} l \left[\frac{w - \frac{t}{2}}{h} + \frac{2\pi}{\ln\left(1 + \frac{2h}{t} + \sqrt{\frac{2h}{t}\left(\frac{2h}{t} + 2\right)}\right)} \right] \quad ...(2.2.2)$$

7. An empirical formula that is computationally efficient and relatively accurate is

$$C = \varepsilon_{ox} l \left[\frac{w}{h} + 0.77 + 1.06\left(\frac{w}{g}\right)^{0.25} + 1.06\left(\frac{t}{h}\right)^{0.5} \right] \quad ...(2.2.3)$$

which is good to 6 % for aspect ratio less than 3.3.

8. A cross-section of the model used for capacitance upper bound calculations is shown in Fig. 2.2.3.

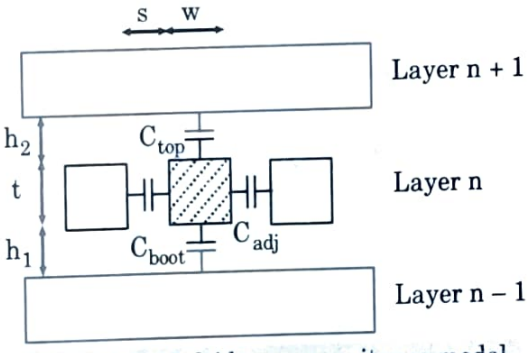


Fig. 2.2.3. Multi layer capacitance model.

9. The total capacitance of the conductor of interest is the sum of its capacitance to the layer above, the layer below, and the two adjacent conductors.

$$C_{total} = C_{top} + C_{boot} + 2C_{adj}$$

$$= \varepsilon_o l \left[2k_{vert} \frac{w}{h} + 2k_{horiz} \frac{t}{s} \right] + C_{fringe} \quad ...(2.2.4)$$

D. Resistance model : Refer Q. 2.1, Page 2–2F, Unit-2.

Que 2.3. How to model the inductance of wires ?

Answer

1. Inductance is difficult to extract and model, so engineers prefer to design in such a way that inductive effects are negligible. Nevertheless, inductance needs to be considered in high-speed designs for wide wires such as clocks and power busses.

2. Current flowing around a loop generates a magnetic field proportional to the area of the loop and the amount of current.
3. Changing the current requires supplying energy to change the magnetic field. This means that changing current induces a voltage proportional to the rate of change. The constant of proportionality is called the inductance.

$$V = L\frac{dI}{dt} \qquad ...(2.3.1)$$

4. Inductance and capacitance also set the speed of light in a medium. Even if the resistance of a wire is zero leading to zero RC delay, the speed of light flight-time along a wire of length with inductance and capacitance per unit length of L and C is

$$t_{pd} = l\sqrt{LC} \qquad ...(2.3.2)$$

5. If the current return paths are the same as the conducts on which electric field lines terminate, the signal velocity v is

$$v = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\varepsilon_{ox}\mu_0}} = \frac{c}{\sqrt{3.9}} \qquad ...(2.3.3)$$

where, μ_0 is the magnetic permeability of free space ($4\pi \times 10^{-7}$H/m) and c is the speed of light in free space (3×10^8 m/s).

6. Changing magnetic fields in turn produce currents in other loops. Hence, signals on one wire can inductively couple onto another, this is called inductive crosstalk.
7. The inductance of a conductor of length l and width w located a height h above a ground plane is approximately

$$L = l\frac{\mu_0}{2\pi}\ln\left(\frac{8h}{w} + \frac{w}{4h}\right) \qquad ...(2.3.4)$$

assuming $w < h$ and thickness is negligible.

8. Inductance depends on the entire loop and therefore cannot be simply decomposed into sections as with capacitance. It is therefore impractical to extract the inductance from a chip layout.

PART-2

Skin Effect and Its Influence.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.4. Write a short note on skin effect.

Answer

1. Current flows along the path of lowest impedance $Z = R + j\omega L$. At high frequency, ω, impedance becomes dominated by inductance.
2. The inductance is minimized if the current flows only near the surface of the conductor closer to the return path.
3. This skin effect can reduce the effective cross-sectional area of thick conductors and raise the effective resistance at high frequency.
4. The skin depth for a conductor is

$$\delta = \sqrt{\frac{2\rho}{\omega\mu}} \qquad ...(2.4.1)$$

where, μ is the magnetic permeability of the dielectric.

5. The frequency of importance is the highest frequency with significant power in the Fourier transform of the signal. This is not the chip operating frequency, but rather is associated with the faster edges.
6. A sine wave with the same 20-80 % rise/fall time as the signal has a period of 8.65 t_{rf}. Therefore, the frequency associated with the edge can be approximated as

$$\omega = \frac{2\pi}{8.65 t_{rf}} \qquad ...(2.4.2)$$

where, t_{rf} is the average 20-80 % rise/fall time.

PART-3

Lumped RC Model.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.5. **What do you mean by lumped *RC* model ? Also explain RC network.**

OR

Explain the Elmore delay model with suitable *RC* networks. Mention its merits. **AKTU 2020-21, Marks 07**

OR

Explain the Elmore delay model with suitable diagram.

AKTU 2018-19, Marks 10

Answer

A. Lumped *RC* model : A first approach lumps the total wire resistance of each wire segment into one single R and similarly combines the capacitance into a single capacitor C. This simple model, called the lumped RC model.

B. Register capacitor network (Elmore delay model) :

1. Consider the resistor-capacitor network of Fig. 2.5.1.
2. The result of this circuit topology is that there exists a unique resistive path between the source node s and any node i of the network. The total resistance along this path is called the path resistance R_{ii}.
3. For example, the path resistance between the source node s and node 4 is,

$$R_{44} = R_1 + R_3 + R_4$$

4. The shared path resistance R_{ik}, which represents the resistance shared among the paths from the root node s to nodes k and i:

$$R_{ik} = R_i \, (R_i \ [\text{path}(s \ i) \ldots \text{path}(s \ k)])$$

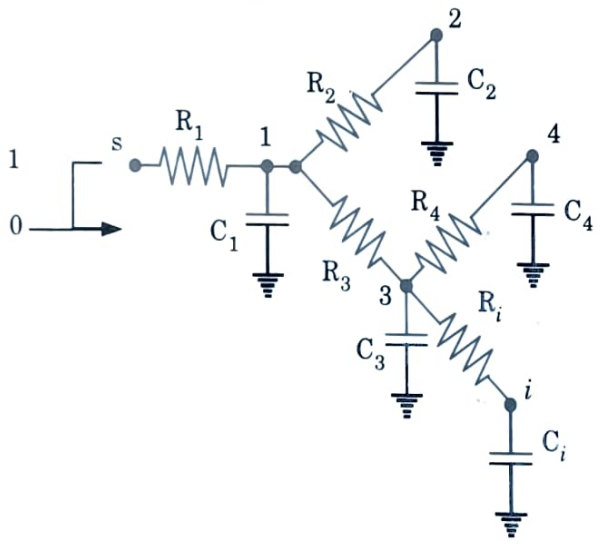


Fig. 2.5.1. Tree structure RC network.

5. Assume now that each of the N nodes of the network is initially discharged to GND, and that a step input is applied at node s at time $t = 0$.
6. The Elmore delay at node i is then given by the following expression :

$$T_{Di} = \sum_{k=1}^{N} C_k R_{tk}$$

C. Merits of Elmore delay model :

i. Simple closed-form expression.

ii. Useful for interconnect optimization.
iii. Upper bound of 50 % delay.
iv. High fidelity.

Que 2.6. Draw and explain the working of Lumped *RC*-model for interconnects. **AKTU 2018-19, Marks 10**

Answer

1. When the physical dimension, in particular, the length, of a wire channel is much smaller as compared with the wavelength of the signal passing through the channel, the channel can be treated as a lumped element with its characteristics depicted by a low-pass RC network, as shown in Fig. 2.6.1, with

$$C \approx C_a(wl) + 2C_f l$$

where C_a is the area capacitance per unit area and

C_f is the fringe capacitance per unit length of the wire channel.

2. A more rigorous criterion on whether lumped models should be used to characterize the behavior of channels is determined by comparing the round-trip time that the signal travels along the channels, denoted by 2τ where τ is the time for the signal to travel from one end of the channel to the other, and the rise time of the signal t_r.

3. If $2\tau < t_r$, lumped models should be used. Otherwise, distributed models should be used.

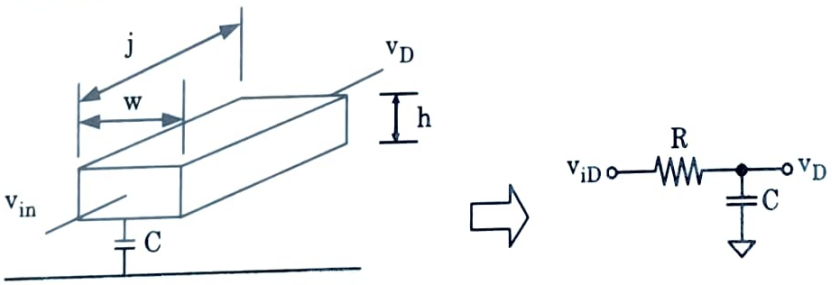


Fig. 2.6.1. Lumped RC model of wire channels.

PART-4

The Distributed RC Model.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.7. Explain distributed *RC* model.

Answer

1. A wire is a distributed circuit with a resistance and capacitance per unit length.
2. Its behavior can be approximated with a number of lumped elements.
3. Three standard approximations are the *L*-model, π-model, and *T*-model, so named because of their shapes.
4. Fig. 2.7.1 shows how a distributed *RC* circuit is equivalent to *N* distributed *RC* segments of proportionally smaller resistance and capacitance, and how these segments can be modeled with lumped elements.

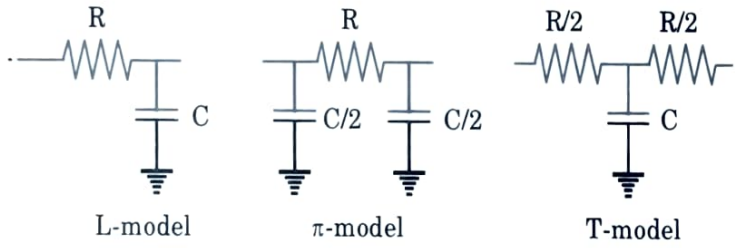


Fig. 2.7.1. Lumped approximation to distributed RC circuit.

5. As the number of segments approaches infinity, the lumped approximation will converge with true distributed circuit.
6. The *L*-model is a poor choice because a large number of segments are required for accurate results.
7. The π-model is much better, three segments are sufficient to give results accurate to 3 %.
8. The *T*-model is comparable to the π-model, but produces a circuit with one more node that is slower to solve by hand or with a circuit simulator.
9. Therefore, it is common practice to model long wires with a 3-5 segment π-model for simulation.

PART-5

Transient Response.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.8. Compute the step response (or transient response) of an inverter.

Answer

1. Fig. 2.8.1(*a*) shows an inverter X_1 driving another inverter X_2 at the end of a wire.
2. Suppose a voltage step from 0 to V_{DD} is applied to node A and to compute the propagation delay, t_{pd}, through X_1, *i.e.*, the delay from the input step until node B crosses $V_{DD}/2$.
3. These capacitances are shown in Fig. 2.8.1(*b*). Fig. 2.8.1(*c*) shows the equivalent circuit diagram in which all the capacitances are lumped into a single C_{out}.
4. Before the voltage step is applied, $A = 0$. N_1 is OFF, P_1 is ON, and $B = V_{DD}$.
5. After the step, $A = 1$. N_1 turns ON and P_1 turns OFF and B drops toward 0.
6. The rate of change of the voltage V_B at node B depends on the output capacitance and on the current through N_1:

$$C_{out} \frac{dV_B}{dt} = I_{dsn1} \qquad ...(2.8.1)$$

7. Suppose the transistors obey the long-channel models. The current depends on whether N_1 is in the linear or saturation region. The gate is at V_{DD}, the source is at 0, and the drain is at V_B. Thus, $V_{gs} = V_{DD}$ and $V_{ds} = V_B$. Initially, $V_{ds} = V_{DD} > (V_{gs} - V_t)$, so N_1 is in saturation.

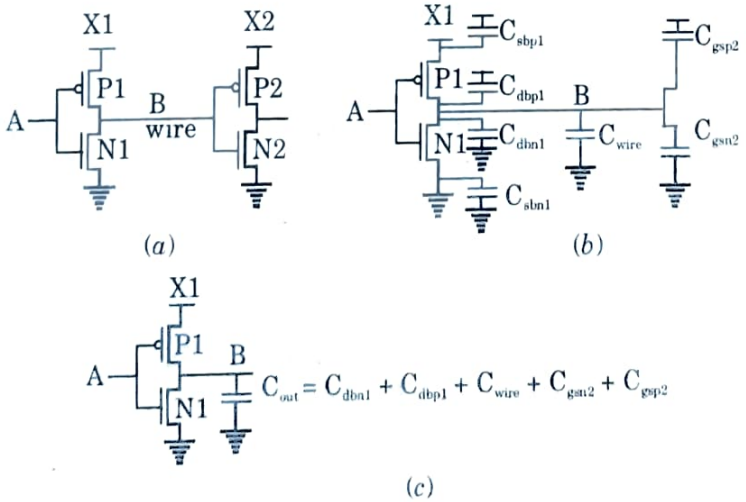


Fig. 2.8.1. Capacitances for inverter delay calculations

8. As V_B falls below $V_{DD} - V_t$, N_1 enters the linear region. The differential equation governing V_B is given by,

$$\frac{dV_B}{dt} = \frac{\beta}{C_{out}} \begin{cases} \frac{(V_{DD} - V_t)^2}{2} & ; \quad V_B > V_{DD} - V_t \\ \left(V_{DD} - V_t - \frac{V_B}{2}\right) V_B & ; \quad V_B < V_{DD} - V_t \end{cases} \quad ...(2.8.2)$$

9. During saturation, the current is constant and V_B drops linearly until it reaches $V_{DD} - V_t$. Thereafter, the differential equation becomes non-linear. The response can be computed numerically.

10. Assuming $V_{tn} + |V_{tp}| < V_{DD}$, the ramp response includes three phases, as shown in Table 2.8.1. Thus, we can write the differential equations for V_B in each phase :

Phase 1 : $V_B = V_{DD}$

Phase 2 : $$\frac{dV_B}{dt} = \frac{I_{dsp1} - I_{dsn1}}{C_{out}} \quad ...(2.8.3)$$

Phase 3 : $$\frac{dV_B}{dt} = \frac{I_{dsn1}}{C_{out}}$$

Table 2.8.1. Phases of inverter ramp response :

Phase	V_A	N_1	P_1	V_B
1	$0 < V_A < V_{tn}$	OFF	ON	V_{DD}
2	$V_{tn} < V_A < V_{DD} - \|V_{tp}\|$	ON	ON	Intermediate
3	$V_{DD} - \|V_{tp}\| < V_A < V_{DD}$	ON	OFF	Falling toward 0

PART-6

RC Delay Model.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.9. **Discuss briefly about effective resistance. Also explain gate and diffusion capacitance of *RC* delay model.**

OR

Explain *RC* delay model for interconnects.

AKTU 2020-21, Marks 07

OR

Explain the concept of *RC* delay model. **AKTU 2019-20, Marks 3.5**

Answer

RC delay models approximate the non-linear transistor I-V and C-V characteristics with an average resistance and capacitance over the switching range of the gate.

A. Effective resistance :

1. The *RC* delay model treats a transistor as a switch in series with a resistor. The effective resistance is the ratio of V_{ds} to I_{ds} averaged across the switching interval of interest.
2. A unit nMOS transistor is defined to have effective resistance R.
3. An nMOS transistor of K times unit width has resistance R/K because it delivers K times as much current.
4. A unit pMOS transistor has greater resistance, generally in the range of $2R$–$3R$, because of its lower mobility.
5. According to the long-channel model, current decreases linearly with channel length and hence resistance is proportional to L.
6. Moreover, the resistance of two transistors in series is the sum of the resistances of each transistor.
7. However, if a transistor is fully velocity-saturated, current and resistance become independent of channel length.
8. The resistance of transistors in series is somewhat lower than the sum of the resistances, because series transistors have smaller V_{ds} and are less velocity-saturated.
9. The effect is more pronounced for nMOS transistors than pMOS because of the higher mobility and greater degree of velocity saturation.

B. Gate and diffusion capacitance :

1. Each transistor has gate and diffusion capacitance.
2. A transistor of k times unit width has capacitance kC.
3. Diffusion capacitance depends on the size of the source/drain region. Wider transistors have proportionally greater diffusion capacitance.
4. Increasing channel length increases gate capacitance proportionally but does not affect diffusion capacitance.

Que 2.10. Draw and explain equivalent *RC* circuit models for nMOS and pMOS transistors. Also draw the equivalent circuit for an inverter.

Answer

1. Fig. 2.10.1 shows equivalent *RC* circuit models for nMOS and pMOS transistors of width k with contacted diffusion on both source and drain.
2. The pMOS transistor has approximately twice the resistance of the nMOS transistor because holes have lower mobility than electrons.
3. The pMOS capacitors are shown with V_{DD} as their second terminal because the n-well is usually tied high. Fig. 2.10.2 shows the equivalent circuit for a fanout-of-1 inverter with negligible wire capacitance.
4. The unit inverters of Fig. 2.10.2(a) are composed from an nMOS transistor of unit size and a pMOS transistor of twice unit width to achieve equal rise and fall resistance.
5. Fig. 2.10.2(b) gives an equivalent circuit, showing the first inverter driving the second inverter's gate.
6. If the input A rises, the nMOS transistor will be ON and the pMOS OFF. Fig. 2.10.2(c) illustrates this case with the switches removed.
7. The capacitors shorted between two constant supplies are also removed because they are not charged or discharged. The total capacitance on the output Y is $6C$.

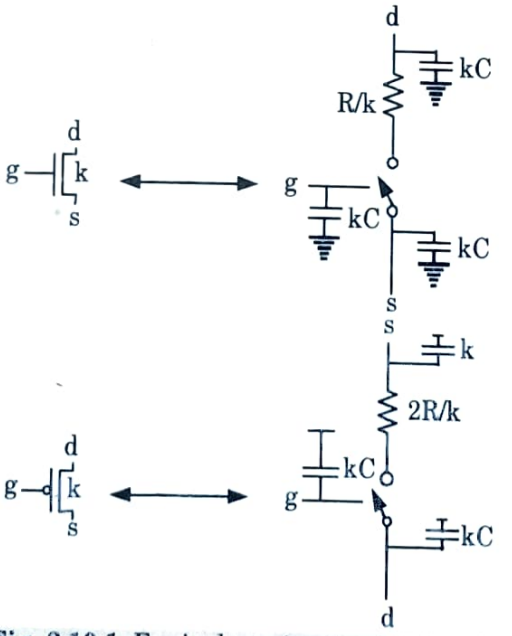


Fig. 2.10.1. Equivalent circuits of transistors.

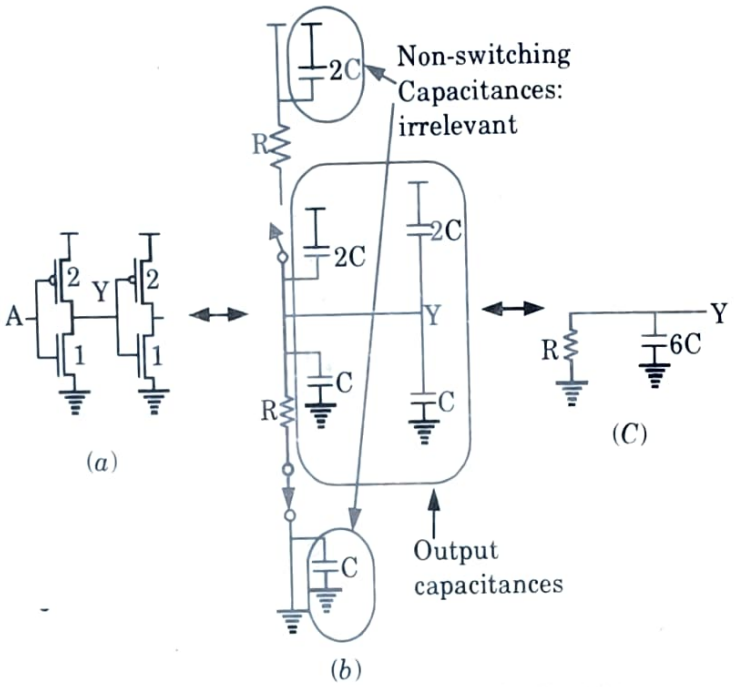


Fig. 2.10.2. Equivalent circuit for an inverter

Que 2.11. **Estimate t_{pd} (or step response) for first-order and second-order *RC* system.**

Answer

1. Consider applying the *RC* model to estimate the step response of the first-order system shown in Fig. 2.11.1.

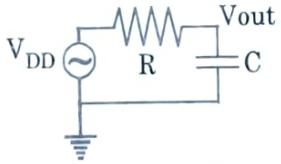


Fig. 2.11.1.

This system is a good model of an inverter sized for equal rise and fall delays.

2. The system has a transfer function

$$H(S) = \frac{1}{1 + sRC} \qquad ...(2.11.1)$$

and a step response

$$V_{out}(t) = V_{DD}\, e^{t/\tau} \qquad ...(2.11.2)$$

where, $\tau = RC$

3. The propagation delay is the time at which V_{out} reaches $V_{DD}/2$, as shown in Fig. 2.11.2.

$$t_{pd} = RC \ln 2 \qquad ...(2.11.3)$$

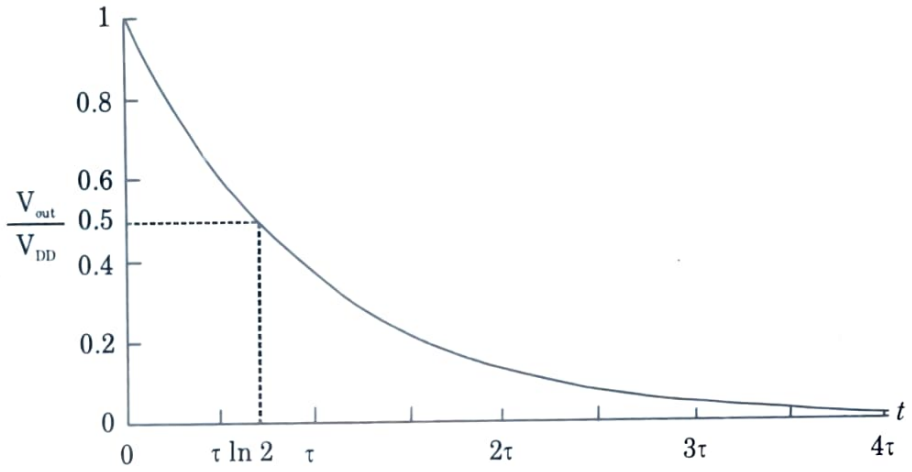


Fig. 2.11.2.

4. Here we define a new effective resistance $R' = R\ln 2$. Now the propagation delay is simply R-C. For the sake of convenience,

$$t_{pd} = RC \qquad ...(2.11.4)$$

where, the effective resistance R is chosen to give the correct delay.

6. Fig. 2.11.3 shows a second-order system. R_1 and R_2 might model the two series nMOS transistors in a NAND gate or an inverter driving a long wire with non-negligible resistance.

6. The transfer function is

$$H(s) = \frac{1}{1 + s[R_1C_1 + (R_1 + R_2)C_2] + s^2R_1C_1R_2C_2} \qquad ...(2.11.5)$$

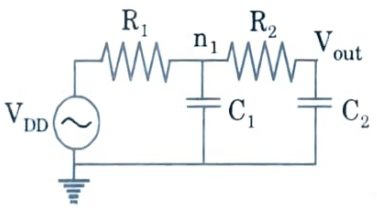


Fig. 2.11.3.

7. The function has two real poles and the step response is

$$V_{out}(t) = V_{DD}\frac{\tau_1 e^{t/\tau_1}.\tau_2 e^{t/\tau_2}}{\tau_1\ \tau_2} \qquad ...(2.11.6)$$

with
$$\tau_{1,2} = \frac{R_1C_1 + (R_1 + R_2)C_2}{2}\left(1 \pm \sqrt{\frac{4R^*C^*}{[1 + (1 + R^*)C^*]^2}}\right) \qquad ...(2.11.7)$$

$$R^* = \frac{R_2}{R_1}; C^* = \frac{C_2}{C_1}$$

8. Eq. (2.11.7) is so complicated that it defeats the purpose of simplifying a CMOS circuit into an equivalent RC network. However, it can be further approximated as a first-order system with a single time constant :

$$\tau = \tau_1 + \tau_2 = R_1C_1 + (R_1 + R_2)C_2 \quad ...(2.11.8)$$

9. This approximation works best when one time constant is significantly bigger than the other. For example, if $R_1 = R_2 = R$ and $C_1 = C_2$ then $\tau_1 = 2.6\,RC$, $\tau_2 = 0.4\,RC$, $\tau = 3\,RC$ and the second-order response and its first-order approximation are shown in Fig. 2.11.4.
10. The error in estimated propagation delay from the first-order approximation is less than 7 %.

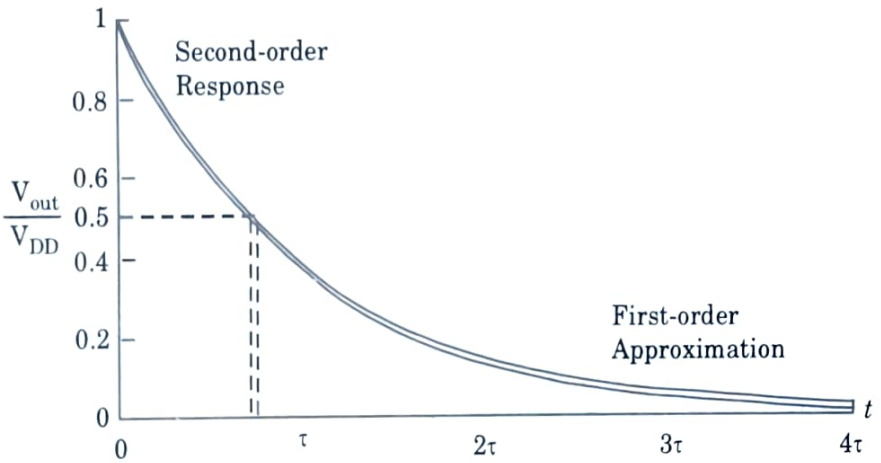


Fig. 2.11.4.

PART-7

Linear Delay Model.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.12. **Explain linear delay model.**

OR

Explain the delay estimation with different optimization techniques. **AKTU 2018-19, Marks 10**

Answer

Delay estimation with different optimization techniques are :

i. **Elmore delay :** Refer Q. 2.5, Page 2–7F, Unit-2.

ii. **Linear delay :**

1. The *RC* delay model showed that delay is a linear function of the fanout of a gate. Based on this observation, designers further simplify delay analysis by characterizing a gate by the slope and *y*-intercept of this function.

2. In general, the normalized delay of a gate can be expressed in units of τ as

$$d = f + p \qquad ...(2.12.1)$$

where p is the parasitic delay inherent to the gate when no load is attached and f is the effort delay or stage effort that depends on the complexity and fanout of the gate :

$$f = gh \qquad ...(2.12.2)$$

3. The complexity is represented by the logical effort, g. An inverter is defined to have a logical effort of 1. More complex gates have greater logical efforts, indicating that they take longer to drive a given fanout.

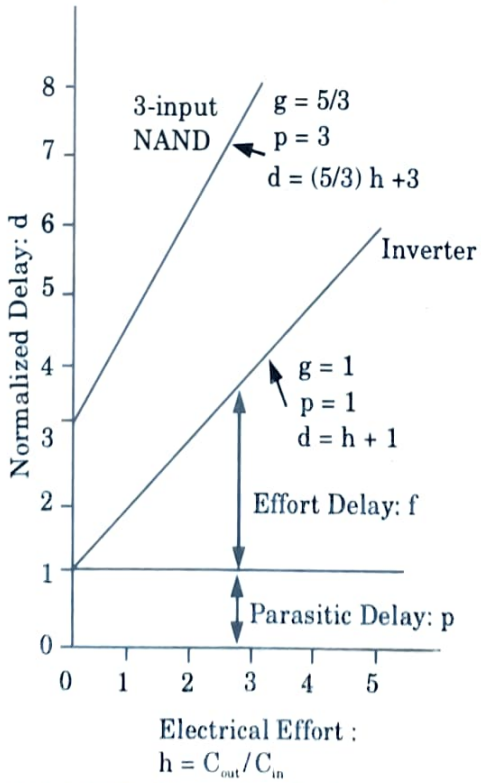


Fig. 2.12.1. Normalized delay vs fanout.

4. A gate driving h identical copies of itself is said to have a fanout or electrical effort of h. If the load does not contain identical copies of the gate, the electrical effort can be computed as

$$h = \frac{C_{out}}{C_{in}} \quad ...(2.12.3)$$

where, C_{out} is the capacitance of the external load being driven and C_{in} is the input capacitance of the gate.

5. Fig. 2.12.1 plots normalized delay v/s electrical for an idealized inverter and 3-input NAND gate.
6. The y-intercepts indicate the parasitic delay, *i.e.*, the delay when the gate drives no load.
7. The slope of the lines is the logical effort.
8. The inverter has a slope of 1 by definition. The NAND has a slope of 5/3.

Que 2.13. Analyze the linear delay model with its different limitations. **AKTU 2018-19, Marks 10**

Answer

A. Linear delay model : Refer Q. 2.12, Page 2–17F, Unit-2.

B. Limitations :

a. Input and output slope :

1. The largest source of error in the linear delay model is the input slope effect.
2. As the rise time of the input increases, the delay also increases because the active transistor is not turned fully ON at once.

b. Input arrival times :

1. Another source of error in the linear delay model is the assumption that one input of a multiple input gate switches while the others are completely stable.
2. When two inputs to a series stack turn ON simultaneously, the delay will be slightly longer than predicted because both transistors are only partially ON during the initial part of the transition.
3. When two inputs to a parallel stack turn ON simultaneously, the delay will be shorter than predicted because both transistors deliver current to the output.

Que 2.14. How to estimate the logical effort and parasitic delay ?

OR

Write short notes on :

i. Logical effort

ii. Parasitic delay. **AKTU 2018-19, Marks 10**

OR

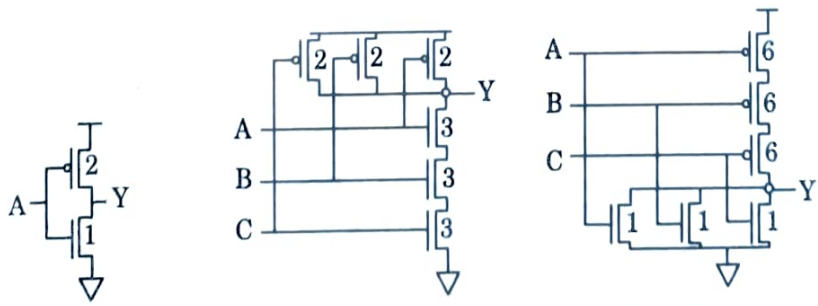
Define the term logical effort and electrical effort. Explain linear *RC* delay model. Mention the limitation of this model.

AKTU 2020-21, Marks 07

Answer

A. Logical effort :

1. Logical effort of a gate is defined as the ratio of the input capacitance of the gate to the input capacitance of an inverter that can deliver the same output current.
2. Logical effort indicates how much worse a gate is at producing output current as compared to an inverter, given that each input of the gate may only present as much input capacitance as the inverter.
3. Fig. 2.14.1 shows inverter, 3-input NAND, and 3-input NOR gates with transistors widths chosen to achieve unit resistance assuming pMOS transistors have twice the resistance of nMOS transistors.
4. The inverter presents three units of input capacitance. The NAND presents five units of capacitance on each input, so the logical effort is 5/3.
5. Similarly, the NOR presents seven units of capacitance, so the logical effort is 7/3. This matches our expectation that NANDs are better than NORs because NORs have slow pMOS transistors in series.



(a) $C_{in} = 3$, $g = 3/3$ (b) $C_{in} = 5$, $g = 5/3$ (c) $C_{in} = 7$, $g = 7/3$

Fig. 2.14.1. Logical gates sized for unit resistance.

Table 2.14.1 the logical effort of common gates

Gate type	Number of inputs				
	1	2	3	4	*n*
Inverter	1				
NAND		4/3	5/3	6/3	$(n + 2)/3$
NOR		5/3	7/3	9/3	$(2n + 1)/3$
Tristate, Multiplexer	2	2	2	2	2
XOR, XNOR		4,4	6,12,6	8,16,16,8	

B. Parasitic delay :

1. The parasitic delay of a gate is the delay of the gate when it drives zero load. It can be estimated with *RC* delay models.
2. A crude method good for hand calculations is to count only diffusion capacitance on the output node.
3. For example, consider the gates in Fig. 2.14.1; assuming each transistor on the output node has its own drain diffusion contact.
4. Transistor widths were chosen to give a resistance of *R* in each gate.
5. The inverter has three units of diffusion capacitance on the output, so the parasitic delay is $3RC = \tau$. In other words, the normalized parasitic delay is 1.
6. In general, we will call the normalized parasitic delay p_{inv}.p_{inv} is the ratio of diffusion capacitance to gate capacitance in a particular process.

Table 2.14.2 the parasitic delay of common gates

Gate type	Number of inputs				
	1	2	3	4	*n*
Inverter	1				
NAND		2	3	4	*n*
NOR		2	3	4	*n*
Tristate, Multiplexer	2	4	6	8	2*n*

7. The parasitic delay also depends on the ratio of diffusion capacitance to gate capacitance.

C. Electrical effort : The effort delay of $4(b'/k)C = 4bC$ depends on the ratio (b) of external load capacitance to input capacitance and thus charges with transistor widths. The factor 4 is by the complexity of the gate. The capacitance ratio is called the electrical effort.

PART-8

Logical Effort of Paths, Scaling.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.15. What do you understand by logical effort of paths ? Also give the limitations of logical effort.

OR

What are the limitations of logical effort ?

AKTU 2019-20, Marks 3.5

Answer

A. Logical effort of paths :

1. Designers often need to choose the fastest circuit topology and gate sizes for a particular logic function and to estimate the delay of the design.
2. Logical effort provides a simple method "on the back of an envelope" to choose the best topology and number of stages of logic for a function.
3. Based on the linear delay model, it allows the designer to quickly estimate the best number of stages for a path, the minimum possible delay for the given topology, and the gate sizes that achieve this delay.

B. Limitations of logical effort :

1. Logical effort does not account for interconnect. Logical effort is most applicable to high-speed circuits with regular layouts where routing delay does not dominate. Such structures include adders, multipliers, memories, and other datapaths and arrays.
2. Logical effort explains how to design a critical path for maximum speed, but not how to design an entire circuit for minimum area or power given a fixed speed constraint.
3. Paths with non-uniform branching or reconvergent fanout are difficult to analyze by hand.
4. The linear delay model fails to capture the effect of input slope. Fortunately, edge rates tend to be about equal in well-designed circuits with equal effort delay per stage.

Que 2.16. Why transistor scaling is of great importance in VLSI ? Write down comparison between constant-field scaling and constant-voltage scaling.

AKTU 2017-18, Marks 10

AKTU 2019-20, Marks 07

Answer

1. The design of high-density chips in MOS VLSI 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.
2. Scaling of MOS transistors is concerned with systematic reduction of overall dimensions of the devices as allowed by the available technology, while preserving the geometric ratios found in the larger devices.
3. The proportional scaling of all devices in a circuit would certainly result in a reduction of the total silicon area occupied by the circuit, thereby increasing the overall functional density of the chip.
4. There are two basic types of size-reduction strategies :

i. Constant-field scaling :

1. This scaling option attempts to preserve the magnitude of internal electric fields in the MOSFET, while the dimensions are scaled down by a factor of S. It is also known as full scaling.
2. To achieve this goal, all potentials must be scaled down proportionally, by the same scaling factor.
3. The full scaling reduces both the drain current and the drain-to-source voltage by a factor of S; hence, the power dissipation of the transistor will be reduced by the factor S^2.

$$P' = I'_D \cdot V'_{DS} = \frac{1}{S^2} I_D \cdot V_{DS} = \frac{P}{S^2} \qquad ...(2.16.1)$$

The significant reduction of the power dissipation is one of the most attractive features of full scaling.

4. The proportional reduction of all dimensions on chip will lead to a reduction of various parasitic capacitances and resistances as well, contributing to the overall performance improvement.

ii. Constant-voltage scaling :

1. In particular, the peripheral and interface circuitry may require certain voltage levels for all input and output voltages, which in turn would necessitate multiple power supply voltages and complicated level-shifter

arrangements. Here, constant-voltages scaling is usually preferred over full scaling.

2. In constant-voltage scaling, all dimensions of the MOSFET are reduced by a factor of S, as in full scaling. The power supply voltage and the terminal voltages, on the other hand, remain unchanged.

3. The drain current is increased by a factor of S while the drain-to-source voltage remains unchanged, the power dissipation of the MOSFET increases by a factor of S.

$$P' = I'_D \cdot V'_{DS} = (S \cdot I_D) \cdot V_{DS} = S \cdot P$$

4. Constant-voltage scaling increases the drain current density and the power density by a factor of S^3. The large increase in current and power densities may eventually cause serious reliability problems for the scaled transistor, such as electromigration, hot-carrier degradation, oxide breakdown, and electrical over-stress.

3 UNIT

Dynamic CMOS Design

CONTENTS

PART-1

Dynamic CMOS Design : Steady-State of Dynamic Gate Circuits.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.1. Explain the dynamic CMOS design.

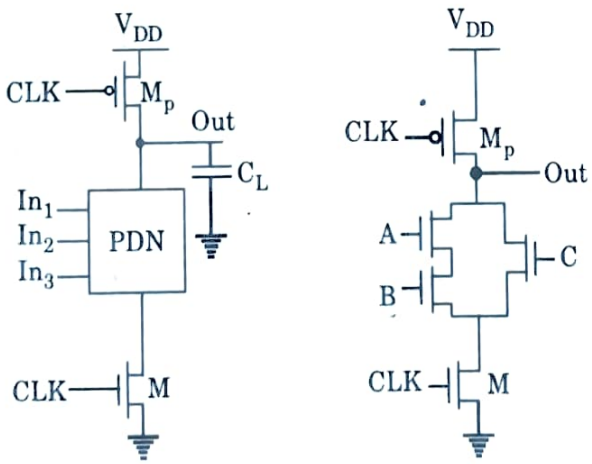
Answer

A. Dynamic CMOS design :

1. Dynamic circuit, which relies on temporary storage of signal values on the capacitance of high impedance circuit nodes.
2. In this, an alternate logic style called dynamic logic is presented that obtains a similar result, while avoiding static power consumption.
3. With the addition of a clock input, it uses a sequence of precharge and conditional evaluation phases.

B. Construction and working :

1. The basic construction of an (n-type) dynamic logic gate is shown in Fig. 3.1.1. The PDN (pull-down network) is constructed exactly as in complementary CMOS.



(a) n-type network (b) Example

Fig. 3.1.1. Basic concepts of dynamic gate.

2. The operation of this circuit is divided into two major phases :

a. Precharge :

i. When CLK = 0, the output node, Out is precharged to V_{DD} by the pMOS transistor M_p. During that time, the evaluate nMOS transistor M_e is OFF, so that the pull-down path is disabled.

ii. The evaluation FET eliminates any static power that would be consumed during the precharge period (this is, static current would flow between the supplies if both the pull down and the precharge device were turned ON simultaneously).

b. Evaluation :

i. For CLK = 1, the precharge transistor M_p is OFF, and the evaluation transistor M_e is turned ON.

ii. The output is conditionally discharged based on the input values and the pull-down topology.

iii. If the inputs are such that the PDN conducts, then a low resistance path exists between Out and GND and the output is discharged to GND.

Que 3.2. What are the properties of dynamic circuit ?

Answer

The properties of dynamic circuit are :

1. Dynamic logic has higher speed than equivalent static family.
2. It occupies less area. The number of transistors is lower than in the static case.
3. It is non-ratioed. The noise margin does not depend on transistor ratios, as is the case in the pseudo-nMOS family.
4. It has low power dissipation. It only consumes dynamic power. No static current path ever exists between V_{DD} and GND.
5. Dynamic logic always require clock.

Que 3.3. Explain the behavior of pass transistor in dynamic CMOS logic implementation. **AKTU 2020-21, Marks 07**

OR

Explain the working of pass transistor circuit. Also explain how the charge stored affects the transfer of logic '1' and logic '0' in nMOS pass transistor circuits.

Answer

1. The fundamental building block of nMOS dynamic logic circuits consisting of a nMOS pass transistor driving the gate of another nMOS as shown in Fig. 3.3.1.

2. The pass transistor M_P is controlled by periodic clock pulse and acts as a switch to charge up or charge down the parasitic capacitance C_x.

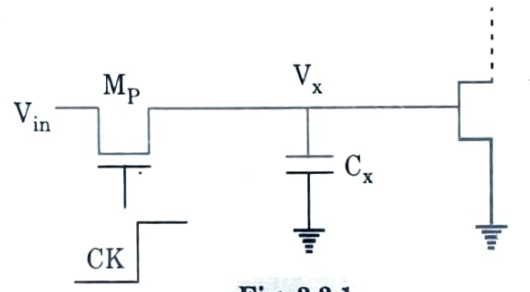


Fig. 3.3.1.

3. There are two possible operation when $CK = 1$, they are logic '1' transfer *i.e.*, charging up the capacitance C_x and logic '0' transfer *i.e.*, charging down the capacitance C_x.
4. The pass transistor M_P provides the only current path to the capacitive node (soft node) V_x.
5. When $CK = 0$, the pass transistor M_P ceases the path and charge stored in parasitic capacitor C_x is used to determine the output of the inverter.

A. Logic '1' transfer :

1. Assume that the soft node voltage is equal to 0 initially, *i.e.*, $V_x(t = 0) = 0$ V. A logic '1' level is applied to the input terminal which corresponds to $V_{in} = V_{OH} = V_{DD}$.
2. Now the clock signal at the gate of the pass transistor goes from 0 to V_{DD} at $t = 0$.
3. The pass transistor M_P starts to conduct as soon as the clock signal becomes active. With $V_{DS} = V_{GS}$, the M_P will operate in saturation region throughout this cycle, since $V_{DS} > V_{GS} - V_{T,n}$.

B. Logic '0' transfer :

1. Assume that the scft node voltage V_x is at logic '1' level initially *i.e.*, $V_x(t = 0) = V_{max} = (V_{DD} - V_{T,n})$. A logic '0' level is applied to the input terminal, which corresponds to $V_{in} = 0$ V.
2. Now the clock signal at the gate of the pass transistor goes from 0 to V_{DD} at $t = 0$.
3. The pass transistor M_P starts to conduct as soon as the clock signal becomes active. In this case, the direction of drain current flowing through M_P will be opposite to that in the logic '1' transfer.
4. With $V_{GS} = V_{DD}$ and $V_{DS} = V_{max}$, the pass transistor operates in the linear region throughout this cycle, since $V_{DS} < V_{GS} - V_{T,n}$.

Que 3.4. Estimate the equation for the charge storage and charge leakage at the soft node capacitance C_x.

Answer

1. The preservation of correct logic at soft node during inactive clock pulse depends on sufficient amount of charge in C_x.
2. Let us assume that high logic has been transferred to the soft node during active clock pulse, and both the input voltage V_{in} and clock are equal to 0 V as shown in Fig. 3.4.1.

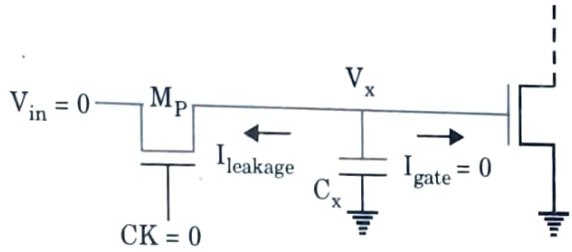


Fig. 3.4.1.

3. The charge in C_x will gradually leak away due to leakage currents associated with pass transistor M_P. The gate current of inverter transistor is negligible.
4. The leakage current responsible for draining the soft node capacitance over time has two responsible components. They are subthreshold channel current and reverse conduction current.

$$I_{leakage} = I_{subthreshold\ (MP)} + I_{reverse\ (MP)} \qquad ...(3.4.1)$$

Other component of C_x, like oxide related parasitic can be considered constants, and represented by C_{in}.

5. The sum of two main components are used to express the total charge stored in the soft node

$$Q = Q_j(V_x) + Q_{in}$$

where

$$Q_{in} = C_{in} \cdot V_x \qquad ...(3.4.2)$$

$$C_{in} = C_{gb} + C_{poly} + C_{metal}$$

6. The time derivation of the total soft node charge Q will give the total leakage current as

$$I_{leakage} = \frac{dQ}{dt} = \frac{dQ_j(V_x)}{dt} + \frac{dQ_{in}}{dt}$$

$$= \frac{dQ_j(V_x)}{dV_x} + \frac{dV_x}{dt} + C_{in}\frac{dV_x}{dt} \qquad ...(3.4.3)$$

Que 3.5. Explain the term voltage bootstrapping in CMOS logic with suitable example. **AKTU 2018-19, Marks 05**

Answer

1. Voltage bootstrapping is a very useful dynamic circuit technique for overcoming threshold voltage drops in digital circuits.

2. Dynamic voltage bootstrapping techniques offer a simple yet effective way to overcome threshold voltage drops which occurs in most situations.

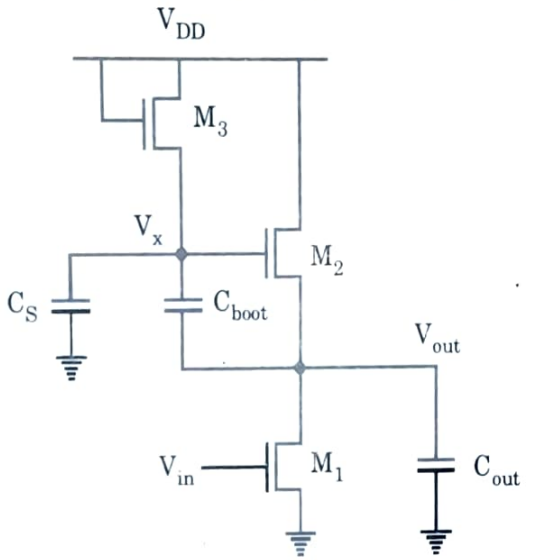


Fig. 3.5.1. Dynamic bootstrapping arrangement to boost V_x during switching.

3. When input voltage V_{in} is low, maximum value of output voltage can be attain, is limited by

$$V_{out}\,(\max) = V_x - V_{T2}\,(V_{out}) \qquad ...(3.5.1)$$

4. To overcome threshold voltage drop and to obtain full logic high level (V_{DD}) at output node, the voltage V_x must be increased.

5. As shown in figure M_3 is added to circuit. The two capacitors C_S and C_{boot} represent capacitances which dynamically couple voltage V_x to ground and to output respectively.

6. This circuit produces a high V_x during switching, so that threshold voltage drop is overcome at output.

$$V_x \geq V_{DD} + V_{T2}\,(V_{out}) \qquad ...(3.5.2)$$

According to equation,

$$i_{CS} \approx i_{Cboot} \Rightarrow C_S \frac{dV_x}{dt} \approx C_{boot} \frac{d(V_{out} - V_x)}{dt}$$

$$\frac{dV_x}{dt} = \frac{C_{boot}}{C_S + C_{boot}} \cdot \frac{dV_{out}}{dt} \qquad ...(3.5.3)$$

7. It is seen from equation that increase in output voltage V_{out} during this switching event will proportionally increase voltage level V_x. Integrating eq. (3.5.3) on both sides, we get

$$\int_{V_{DD}-V_{T3}}^{V_x} dV_x = \frac{C_{boot}}{C_S + C_{boot}} \int_{V_{OL}}^{V_{DD}} dV_{out}$$

$$V_x = (V_{DD} - V_{T3}) + \frac{C_{boot}}{C_S + C_{boot}} (V_{DD} - V_{OL}) \qquad ...(3.5.4)$$

8. If capacitor $C_{boot} >> C_S$, maximum value of V_x is

$$V_x(\max) = 2V_{DD} - V_{T3} - V_{OL} \quad ...(3.5.4)$$

which proves that voltage bootstrapping can significantly boost the voltage level V_x.

Que 3.6. Differentiate between dynamic CMOS logic circuit and static CMOS logic circuit.

Answer

S. No.	Parameter	Static CMOS logic circuit	Dynamic CMOS logic circuit
1.	Glitches	30 % energy increase	Intrinsically does not have this problem
2.	Switching activity	Depends on previous state	Does not depend on previous state.
3.	Power down models	Effectively used	Generally, higher activity factor.
4.	Clock Power	No clock	Not well suited due to gate capacitance of precharge MOS transistor

PART-2

Noise Consideration in Dynamic Design, Charge Sharing.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.7. What are the noise considerations in dynamic design ?

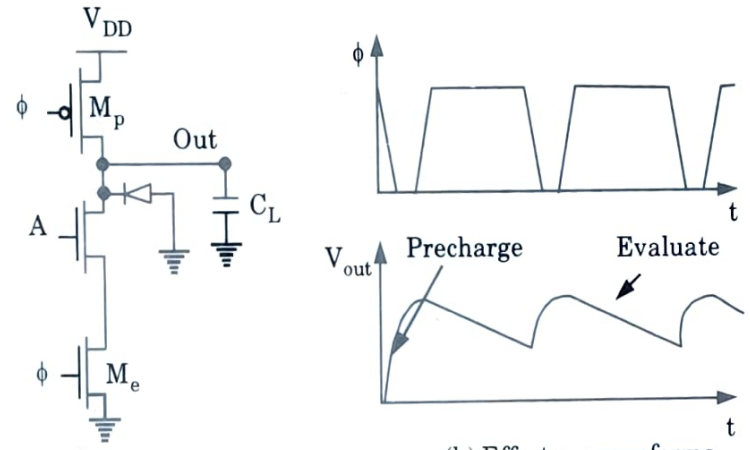
Answer

Noise considerations in dynamic design are :

i. Charge leakage :

1. The operation of the dynamic logic depends on the principles of dynamically storing a charge on the output node (capacitor).
2. Due to leakage currents, this charge gradually leaks away, resulting eventually in malfunctioning of the gate.

3. Charge leakage causes degradation in the logic high level.



(a) Leakage source (b) Effect on waveforms

Fig. 3.7.1. Charge leakage.

4. Dynamic circuits require a minimal clock rate which is typically between 250 Hz and 1 kHz.

ii. Charge sharing :

1. During the precharge phase, the output node is precharged to V_{DD}.

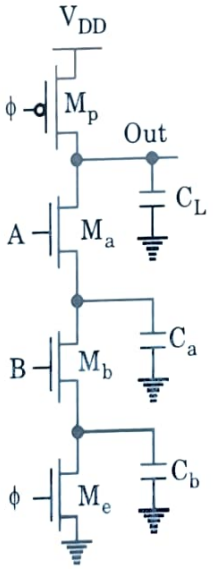


Fig. 3.7.2.

2. Capacitors C_a and C_b represent the parasitic capacitances of the internal nodes of the circuit.
3. Assume now that during precharge all inputs are set 0 and the capacitance C_a is discharged.
4. Assume further that input B remains at 0 during evaluation, while input A makes a 0 to 1 transition.

5. Turning transistor M_a ON, the charge stored originally on capacitor C_L is redistributed over C_L and C_a. This causes a drop in the output voltage, which cannot be recovered due to the dynamic nature of the circuit.
6. One way of alleviating the problem of charge sharing is to add a small pMOS transistor in parallel with the precharge transistor as shown in Fig. 3.7.3 to supply the extra current needed.

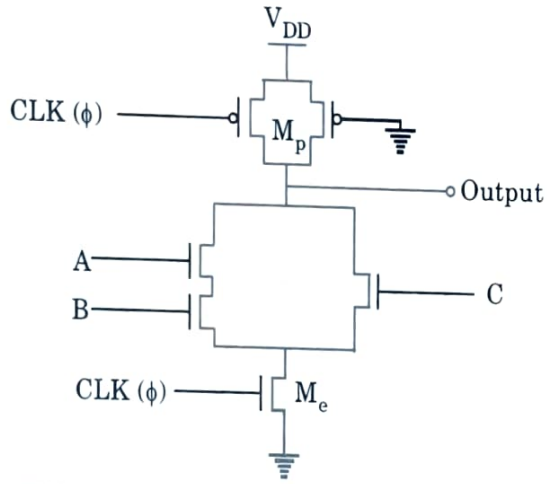


Fig. 3.7.3. Addition of level restoring transistor.

iii. Clock feed-through :

1. The clock signal is coupled to the storage node by the gate-source capacitance and the gate-overlap capacitance of the precharge device as shown in Fig. 3.7.4.

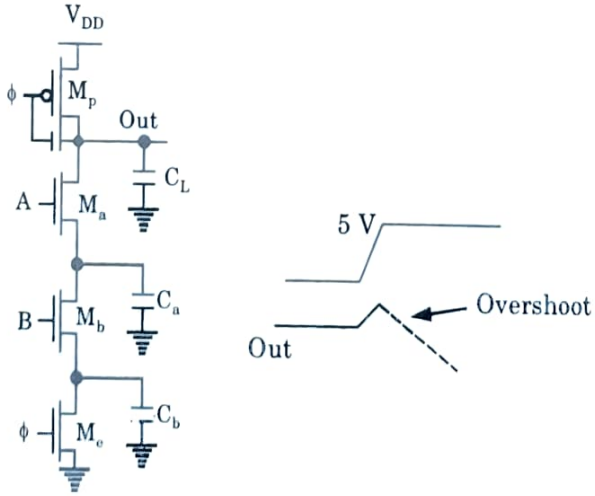


Fig. 3.7.4. Clock feed through.

2. The fast rising and falling edges of the clock couple into the signal node.
3. The disadvantage of clock feed-through is that it causes the signal level to rise sufficiently above the supply voltage as to forward bias the junction diode.

4. This causes electron injection into the substrate and eventually resulting in faulty operation.
5. This can be avoided by providing a sufficient number of well contacts close to the precharge device to collect the injected current.

iv. Cascading dynamic gate :

1. Cascade dynamic gate is a combination of two simple inverters connected in series, as shown in Fig. 3.7.5.

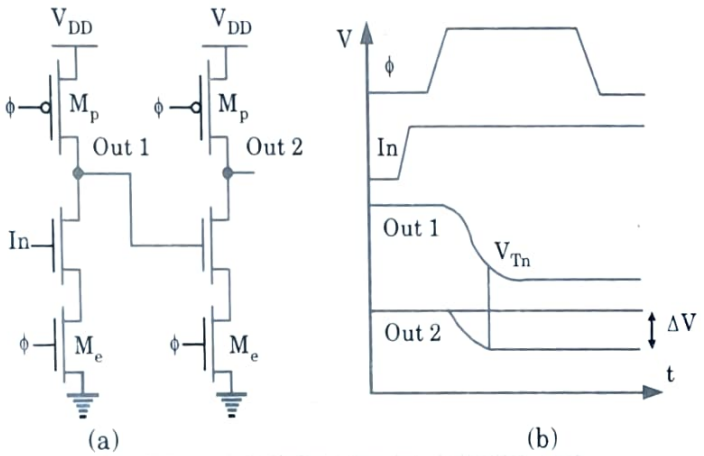


Fig. 3.7.5. Cascading dynamic gates.

2. The problem is that during precharge all outputs are being precharged to 1.
3. The PDN of the second gate is thus in a conducting state at the onset of the evaluation phase.
4. Suppose now that "In" makes a 0 to 1 transition. At the onset of the evaluation period ($\phi = 1$), output "Out 1" starts to discharge.
5. When "Out 1" exceeds the switching threshold of the second gate, a conducting path exists between "Out 2" and GND. "Out 2" therefore discharges as well, and the correct output of the gate equals 1.
6. This conducting path is only turned OFF when "Out 1" reaches V_{Tn} and shuts OFF the nMOS pull-down transistor. This leaves "Out 2" at an intermediate voltage level. The charge loss leads to reduce noise margins and eventual malfunction.
7. It is obvious that the cascading problem arise because the output is precharged to "1", so correct operation is guaranteed as long as the inputs can only make a signal 0 to 1 transition during the evaluation period.

PART-3

Domino Logic.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.8. **Explain the domino CMOS logic. Also discuss the cascaded domino CMOS logic structures.**

OR

Explain CMOS domino circuit along with its features. How it can be cascaded in VLSI circuits ? **AKTU 2017-18, Marks 10**

Answer

1. The generalized circuit diagrams of domino CMOS logic gate is shown in Fig. 3.8.1. The additional inverter allows us to operate a number of such structures in cascade.
2. During precharge phase ($CK = 0$), the output node of dynamic CMOS is at logic high, and output of CMOS inverter become low.

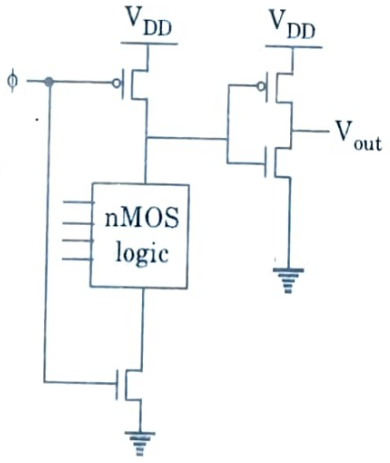


Fig. 3.8.1.

3. When ($CK = 1$), at the beginning of evaluation phase there are two possibilities :

a. The output node of dynamic CMOS stage either discharge to low or remains high.

b. Similarly inverter output voltage can also make one transition, from 0 to 1.

4. Now regardless of input voltages applied to dynamic CMOS stage, it is not possible for buffer output to make 1 to 0 transitions during evaluation phase.

5. Now consider the case for cascading domino CMOS logic gates as shown in Fig. 3.8.2.

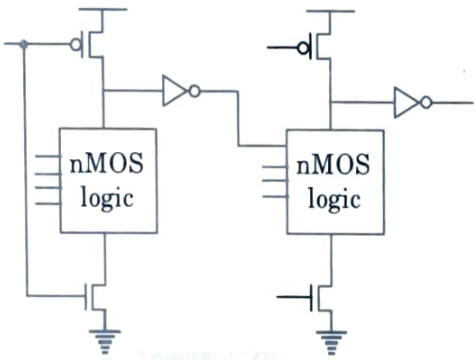


Fig. 3.8.2.

6. All input transistors will be OFF during the precharge phase, since all buffer output is equal to '0'.

7. During evaluation phase each buffer output can make 0 to 1 transition and thus each input of all subsequent logic stages can also make 0 to 1 transition.

8. In cascade structure consisting of several such stages, the evaluation of each stage ripples the next stage evaluation, similar to a chain of dominos falling one after the other. The structure is hence called domino CMOS logic.

Que 3.9. **Implement the Boolean function $Y = AB + (C + D)(F + E) + GH$ using domino CMOS logic.** **AKTU 2018-19, Marks 05**

OR

Implement the Boolean function $Z = AB + (C + D)(E + F) + GH$ using standard CMOS and domino CMOS logic. **AKTU 2019-20, Marks 07**

Answer

1. Fig. 3.9.1 and Fig. 3.9.2 show the realization of the logic by using conventional CMOS and domino logic styles.

2. From implementation, it is clear that the number of transistors used in domino logic is less than the number of transistors in conventional CMOS logic.

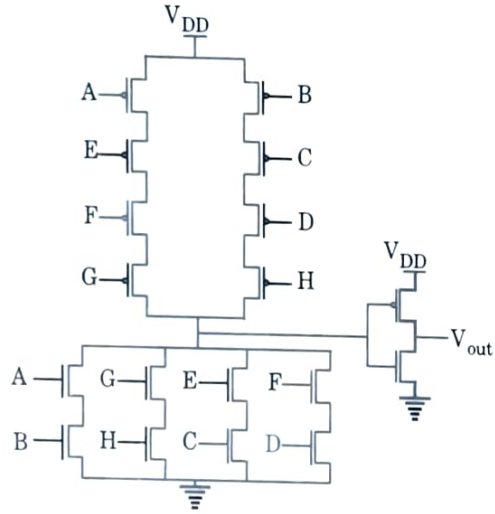


Fig. 3.9.1. Conventional CMOS logic gate.

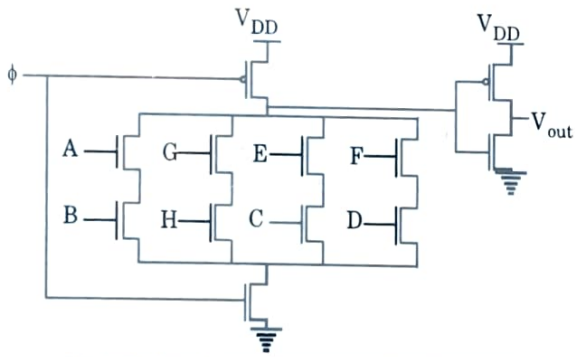


Fig. 3.9.2. Domino CMOS logic style.

Que 3.10. **Write the properties of domino CMOS logic.**

Answer

The properties of domino CMOS logic are :

1. Each gate requires $N + 4$ transistors.
2. Logic evaluation propagates as falling dominoes hence minimum evaluation period is determined by the logic depth.
3. The nodes must be precharged during the precharge period. Total precharge time depends on size of pMOS.
4. Inputs must be stable during the evaluation period.
5. Gates are ratioless and non-inverting.

PART-4

NP-CMOS logic.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.11. **Explain NORA CMOS (NP-Domino) logic circuit. Also write its advantages and disadvantages.**

OR

Explain Domino and NORA CMOS logic circuit with suitable example. **AKTU 2019-20, Marks 07**

OR

Enlist the advantages of dynamic logic circuit over static logic circuit. Explain NORA CMOS logic circuit with suitable example.

AKTU 2018-19, Marks 10

Answer

A. Domino CMOS logic : Refer Q. 3.8, Page 3–11F, Unit-3.

B. Advantages of dynamic logic circuit over static logic circuit : The dynamic logic implementation of complex functions generally requires a smaller silicon area than does the static logic implementation.

C. NORA CMOS logic circuit :

1. This design alternative to domino CMOS logic eliminates the output buffer without causing race problems between clock and data that arise when cascading dynamic gates.
2. NORA CMOS (No-Race CMOS) avoids these race problems by cascading alternate nMOS and pMOS blocks for logic evaluation. The cost is routing two complemented clock signals.
3. The cascaded NORA gate structure is shown in Fig. 3.11.1.

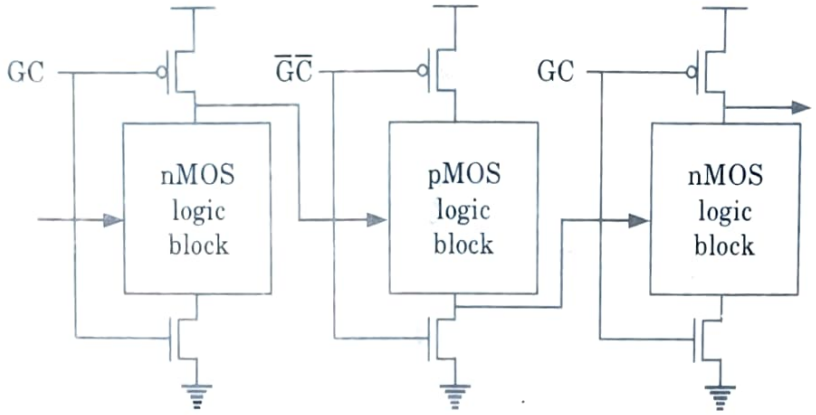


Fig. 3.11.1. NORA CMOS cascaded gates.

4. When the global clock (GC) is low ($\overline{GC}$ high), the nMOS logic block output nodes are precharged high, while outputs of gates with pMOS logic blocks are precharged low. When the clock changes, gates are in the evaluate state.

D. Advantages of NORA CMOS logic circuit :

1. Static CMOS inverter is not required at the output of every dynamic logic state.
2. Compatible with domino CMOS logic.
3. It allows pipelined system architecture.

E. Disadvantage of NORA CMOS logic circuit : NORA CMOS also suffers from charge sharing and leakage. To overcome dynamic charge sharing and soft node leakage problem a circuit technique called Zipper CMOS can be used.

Que 3.12. Write the difference between dynamic CMOS logic circuit and static CMOS logic circuit. Explain the classification of dynamic CMOS logic circuit and design a 2 input EXOR logic gate using domino logic. **AKTU 2019-20, Marks 07**

Answer

A. Difference : Refer Q. 3.6, Page 3–7F, Unit-3.

B. Classification of dynamic CMOS logic circuit :

i. Domino logic : Refer Q. 3.8, Page 3–11F, Unit-3.

ii. NORA logic : Refer Q. 3.11, Page 3–14F, Unit-3.

C. 2 input EXOR logic gate using domino logic gate : The 2 input EXOR logic gate is shown in Fig. 3.12.1.

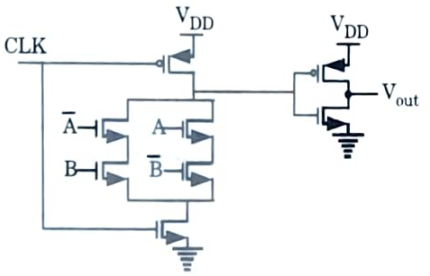


Fig. 3.12.1.

PART-5

Problems in Single-Phase Clocking.

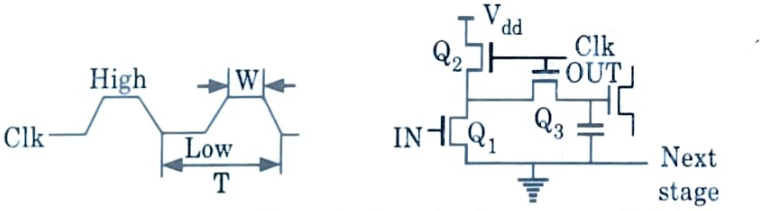
Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.13. Explain single-phase clock. What are the problems in single-phase clock ?

Answer

A. Single-phase clock :



(a) Single-phase clock (b) Single-phase n-type MOS inverter

Fig. 3.13.1.

1. Single-phase clock consists of a sequence of pulses having high and low levels with width W and time period T as shown in Fig. 3.13.1(a).
2. A single-phase clock has two states (low and high) and two edges per period.
3. The schematic diagram of a single-phase dynamic nMOS inverter is shown in Fig. 3.13.1(b).

4. When the clock is in the high state, both transistors Q_2 and Q_3 are ON. Depending on the input, Q_1 is either ON or OFF.
5. If the input voltage is low, Q_1 is OFF and the output capacitor charges to V_{DD} through Q_2 and Q_3.
6. When the input voltage is high, Q_1 is ON and the output is discharged through it to a low level.
7. When the clock is in the low state, the transistors Q_2 and Q_3 are OFF, isolating the output capacitor. This voltage is maintained during the OFF period of the clock, provided the period is not too long.
8. During the period, the power supply is also disconnected and no current flow through the circuit. As current flows only when the clock is high, the power consumption is small, and it depends on the duty cycle.
9. It may be noted that the output of the circuit is also ratioed, because the low output voltage depends on the ratio of the ON resistance of Q_1 to that of Q_2 (ratio of high time to the time period T).
10. As we know that, this ratio is related to the physical dimensions of Q_1 to Q_2 (Low ratio) and is often referred to as the inverter ratio.

B. Problems in single-phase clock :

1. The circuits realized using the single-phase clocking scheme has the disadvantages that the output voltage level is dependent on the inverter ratio and the number of transistors in the current path of GND.
2. In other words, single-phase dynamic circuits are ratioed logic. Moreover, the circuit dissipates power when the output is low and the clock is high.
3. Another problem arising out of single-phase clocked logic is known as clock skew problem.
4. This is due to a delay in a clock signal during its journey through a number of circuit stages. This results in undesired signals like glitch, hazards, etc. Some of the problems can be overcome using two-phase clocking scheme.

PART-6

Two-Phase Non-Overlapping Clocking.

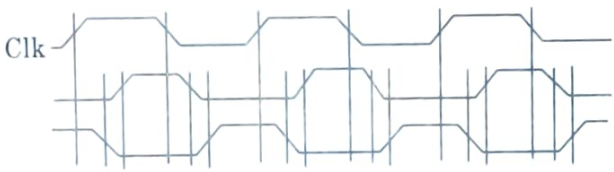
Questions-Answers

Long Answer Type and Medium Answer Type Questions

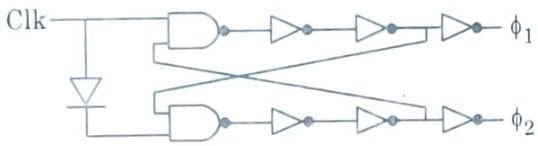
Que 3.14. Briefly discuss about two-phase non-overlapping clock.

Answer

1. A two-phase non-overlapping clock is shown in Fig. 3.14.1(a). As the two-phases (ϕ_1 and ϕ_2) are never high simultaneously, the clock has three states and four edges and satisfies the property $\phi_1 \phi'_2 = 0$.
2. There is a dead time, Δ, between transitions of the clock signals as shown in Fig. 3.14.1(a).
3. The schematic diagram of a circuit that generates two-phase clock is shown in Fig. 3.14.1(b).



(a) Two-phase clock



(b) Two-phase clock generator circuit

Fig. 3.14.1.

4. An inverter based on two-phase clock generator is shown in Fig. 3.14.2.
5. When the clock ϕ_2 is high, the intrinsic capacitor charges to V_{DD} through Q_1. And clock ϕ_1 which comes after ϕ_2 performs the evaluation. If V_{in} is high, Q_2 is turned ON and Q_3 is ON, the capacitor discharges to the GND level and the output V_0 attains low logical level.
6. If V_{in} is low the Q_2 is OFF and there is no path for the capacitor to discharge. Therefore, the output V_0 remains at high logic level.
7. It may be noted that the pull-up and pull-down transistors are never simultaneously ON. The circuit has no DC circuit path regardless of the state of the clocks or the information stored on the parasitic capacitors.
8. Moreover, the output is not ratioed, *i.e.*, the low-level output is independent of the relative value of the aspect ratio of the transistors.
9. That is why the circuits based on two-phase clocking are often termed as ratioless and powerless.

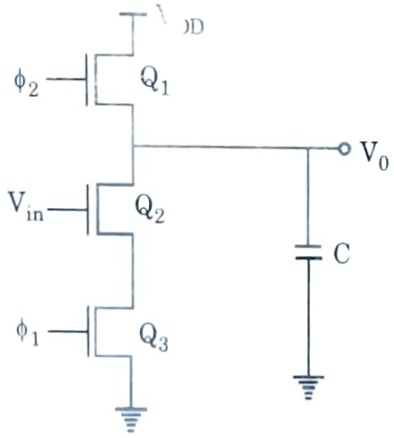


Fig. 3.14.2. Two-phase n-type MOS (nMOS) inverter.

Que 3.15. **Describe the working of three stage pseudo nMOS dynamic shift register driven with two-phase clocking giving its circuit.**

AKTU 2019-20, Marks 07

Answer

1. In depletion load dynamic shift register circuits, the input data are inverted once and transferred or shifted into the next stage during each clock pulse as shown in Fig. 3.15.1.

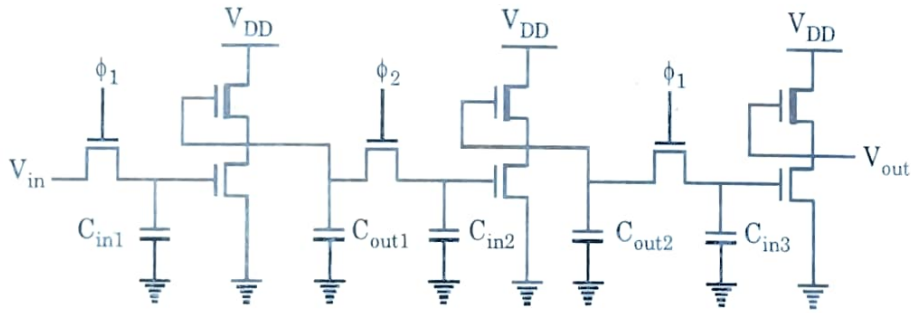


Fig. 3.15.1.

2. The operation of three stage dynamic shift register circuit is as follows.
3. When ϕ_1 is active, input V_{in} is transferred into capacitance C_{in1} and the output appears at C_{out1}.
4. When ϕ_2 becomes active during next phase, the C_{out1} is transferred as input to second stage C_{in2} and output of second stage is determined.

5. During the first stage C_{in1} continues to retain its previous level via charge storage.
6. When ϕ_1 becomes active again, the original data written into register *i.e.*, into third stage and the first stage can now accept the next data bit.
7. The maximum clock frequency is determined by the signal propagation delay through one inverter stage.
8. One-half period of clock signal is long enough to allow C_{in} to charge up or down and to propagate the logic to C_{out}. The logic high of each inverter stage is one threshold voltage lower than the power supply voltage.

PART-7

Sequential CMOS Logic Circuits, Layout Design.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.16. Explain sequential CMOS circuit.

Answer

1. Fig. 3.16.1 shows a sequential circuit consisting of a combinational circuit and a memory block in the feedback loop.

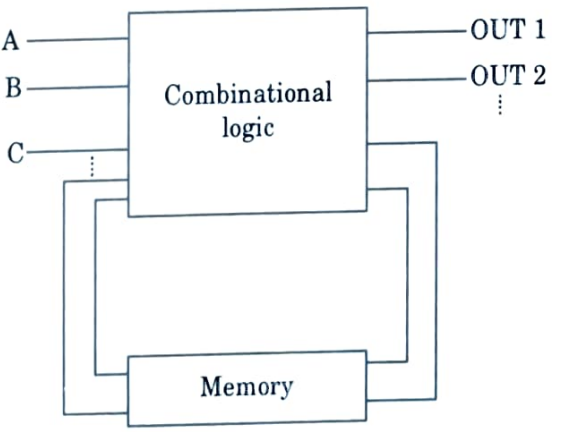


Fig. 3.16.1.

2. In most cases, the regenerative behavior of sequential circuits is due to either a direct or an indirect feedback connection between the output and the input.

3. Regenerative operation can, under certain conditions, also be interpreted as a simple memory function.
4. The critical components of sequential systems are the basic regenerative circuits, which can be classified into three main groups :

i. **Bistable :** Bistable circuits have, as their name implies, two table states or operation modes, each of which can be attained under certain input and output conditions.

ii. **Monostable :** Monostable circuits, on the other hand, have only one stable operating point. Even if the circuit experiences an external perturbation, the output eventually returns to the single stable state after a certain time period.

iii. **Astable :** In astable circuits, there is no stable operating point or state which the circuit can preserve for a certain time period. Consequently, the output of an astable circuit must oscillate without setting into a stable operating mode.

Que 3.17. Explain SR latch based on NOR gate.

Answer

1. The gate-level schematic of the SR latch consisting of two NOR gates are shown in Fig. 3.17.1.

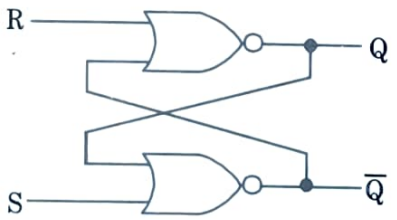


Fig. 3.17.1.

2. If the set input (S) is equal to logic 1 and the reset input is equal to logic 0 then the output Q will be forced to logic 1. While $\bar{Q}$ is forced to logic 0. This means the SR latch will be set, irrespective of its previous state.
3. Similarly, if S is equal to 0 and R is equal to 1 then the output Q will be forced to 0 while $\bar{Q}$ is forced to 1. This means the latch is reset, regardless of its previously held state.
4. Finally, if both of the inputs S and R are equal to logic 1 then both output will be forced to logic 0 which conflicts with the complementarity of Q and $\bar{Q}$.

5. Therefore, this input combination is not allowed during normal operation.
6. Truth table of NOR based SR Latch is given in Table 3.17.1.

Table 3.17.1.

S	R	Q	$\bar{Q}$	Operation
0	0	Q	$\bar{Q}$	Hold
1	0	1	0	Set
0	1	0	1	Reset
1	1	0	0	Not allowed

7. CMOS SR latch based on NOR gate is shown in Fig. 3.17.2.

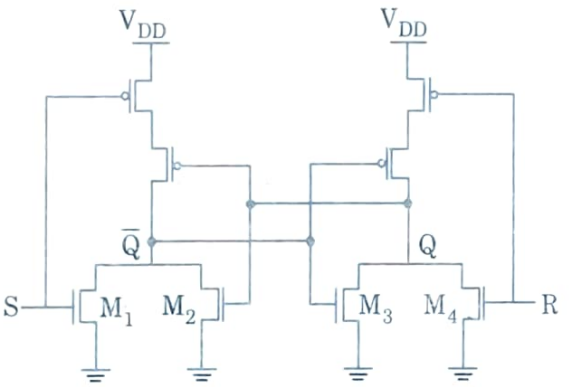


Fig. 3.17.2. CMOS SR latch circuit based on NOR2 gates.

8. If the S is equal to V_{OH} and the R is equal to V_{OL}, both of the parallel-connected transistors M_1 and M_2 will be ON. The voltage on node $\bar{Q}$ will assume a logic-low level of $V_{OL} = 0$.
9. At the same time, both M_3 and M_4 are turned OFF, which results in a logic-high voltage V_{OH} at node Q. If the R is equal to V_{OH} and the S is equal to V_{OL}, M_1 and M_2 turned OFF and M_3 and M_4 turned ON.

Que 3.18. **Discuss NOR-based SR latch with a clock.**

Answer

1. Fig. 3.18.1 shows a NOR-based SR latch with a clock added. The latch is responsive to inputs S and R only when CLK is high.

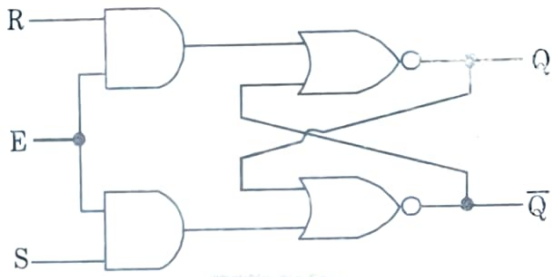


Fig. 3.18.1.

2. When CLK is low, the latch retains its current state. Observe that Q changes state :

i. When S goes high during positive CLK.

ii. On leading CLK edge after changes in S and R during CLK low time.

iii. A positive glitch in S while CLK is high.

iv. When R goes high during positive CLK.

3. CMOS AOI implementation of clocked NOR based SR latch is shown in Fig. 3.18.2.

4. When clock is high, the circuit becomes simply a NOR based CMOS latch which will respond to input S and R.

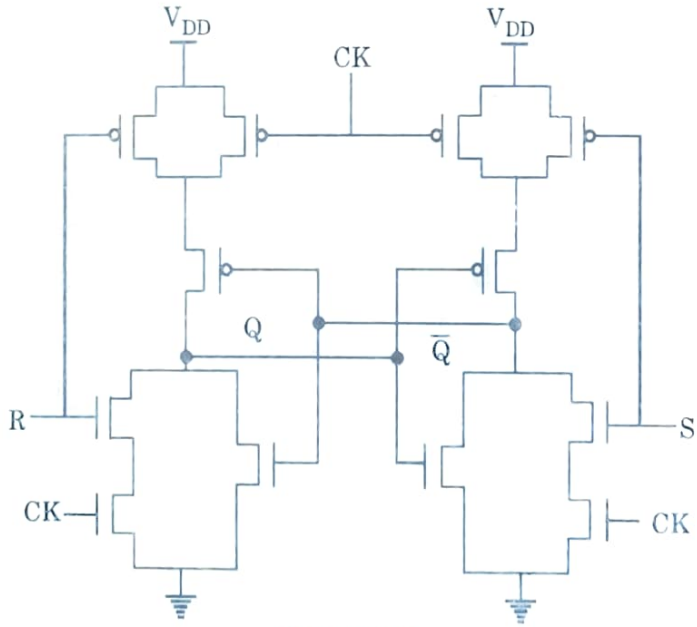


Fig. 3.18.2.

Que 3.19. **Write a short note on clocked *JK* flip-flop.**

Answer

1. Fig. 3.19.1 shows a clocked *JK* latch, based on NAND gates.

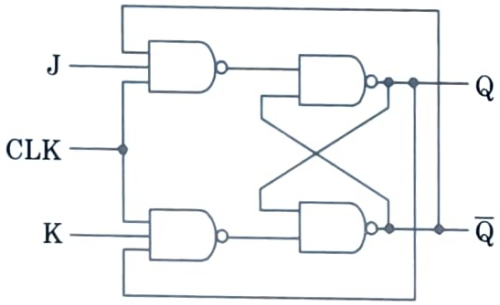


Fig. 3.19.1.

2. The disadvantage of an SR latch is that when both S and R are high, its output state becomes indeterminant.
3. The *JK* latch eliminates this problem by using feedback from output to input, such that all input states of the truth table are allowable. If $J = K = 0$, the latch will hold its present state.
4. If $J = 1$ and $K = 0$, the latch will set on the next positive-going clock edge, *i.e.*, $Q = 1$, $\bar{Q} = 0$.
5. If $J = 0$ and $K = 1$, the latch will reset on the next positive-going clock edge, *i.e.*, $Q = 1$ and $\bar{Q} = 0$.
6. If $J = K = 1$, the latch will toggle on the next positive-going clock edge.
7. The operation of the clocked *JK* latch is given in the Table 3.19.1.

Table 3.19.1.

J	K	Q	$\bar{Q}$	S	R	Q	$\bar{Q}$	Operation
0	0	0	1	1	1	0	1	Hold
		1	0	1	1	1	0	
0	1	0	1	1	1	0	1	Reset
		1	0	1	0	0	0	
1	0	0	1	0	1	1	0	Set
		1	0	1	1	1	0	
1	1	0	1	0	1	1	0	Toggle
		1	0	1	0	0	1	

4
UNIT

Semiconductor Memories

CONTENTS

PART-1

Semiconductor Memories : Dynamic Random Access Memories (DRAM).

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.1. What is semiconductor memory ? Also explain its types.

Answer

Semiconductor memory : Semiconductor memory arrays capable of storing large quantities of digital information are essential to all digital systems.

B. Types of semiconductor memory :

The Fig. 4.1.1 shows an overview of semiconductor memory types.

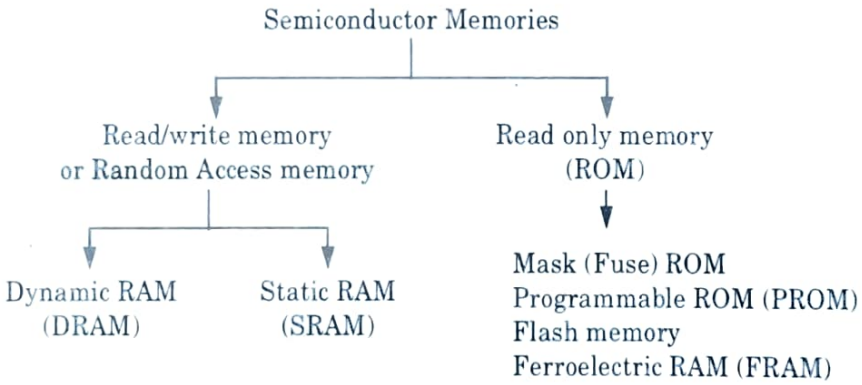


Fig. 4.1.1. Overview and characteristic summary of different memory types and their classification.

i. RAM :

1. The Read/write memory must permit the modification of data bits stored in the memory array as well as their retrieval on demand.
2. The stored data is volatile *i.e.*, the stored data is lost when the power supply voltage is turned OFF.
3. Based on the operation RAM's are classified into two main categories.

a. Dynamic RAM : The DRAM cell consists of a capacitor to store binary information '1' (high voltage) or '0' (low voltage) and a transistor to access the capacitor. Due to low cost and high density DRAM is widely used for main memory.

b. Static RAM : The SRAM cell consists of a latch therefore, the cell data is kept as long as the power is turned ON and refresh operation is not require as in the case of DRAM cells. SRAM is mainly used for the cache memory in various applications.

ii. Read only memory (ROM) :

1. It allows only retrieval of previously stored data and does not permit modifications of the stored information during normal operation.
2. ROM are non-volatile memories *i.e.*, the stored data is not lost even when the power supply is OFF and refresh operation is not required.
3. Mask ROM are the one in which data are written during chip fabrication by using a photo mask and in programmable ROM (PROM), data are written electrically after the chip is fabricated.
4. Depending on data erasing characteristics PROMs are further classified into Fuse ROM, erasable PROM (EPROM) and electrically erasable PROM (EEPROM). The data written by blowing the fuse electrically cannot be erased or modified in Fuse ROM.
5. Data in EPROMs and EEPROMs can be rewritten, but the number of subsequent re-write operation is limited.
6. Flash memory is similar to EEPROM, where data in the block can be erased by using a high electrical voltage.
7. Ferroelectric RAM (FRAM) utilizes the hysteresis characteristics of a ferroelectric capacitor to overcome the slow write operation of other EEPROMs.

Que 4.2. Explain the working of three-transistor DRAM cell with concept of leakage currents and refresh operation. What are the features required to select a proper RAM ?

OR

Write short note on DRAM cell. Explain leakage and refresh operation in DRAM cells. **AKTU 2019-20, Marks 07**

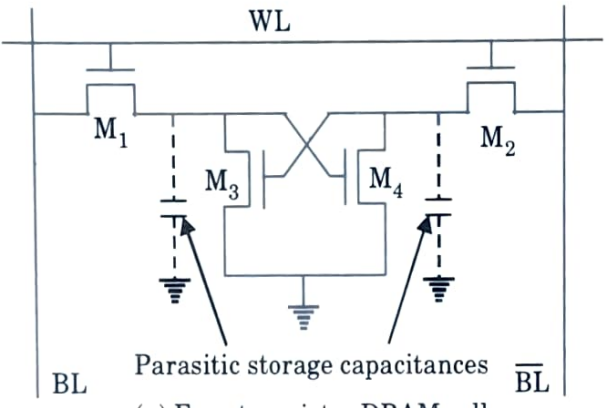
Answer

A. DRAM : Refer Q. 4.1, Page 4–2F, Unit-4.

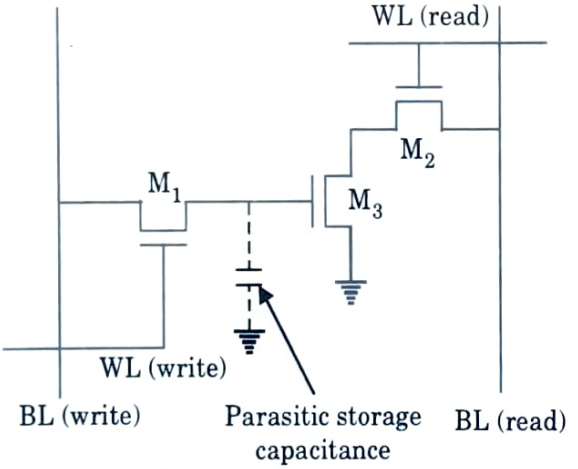
B. Working of three-transistor DRAM cell :

1. The various configuration and step wise historical evolution of the DRAM cells is shown in Fig. 4.2.1.
2. In Fig. 4.2.1(*a*), during write operation, a word line is enabled and complementary data are written from a pair of bit lines. Charge is stored at parasitic and gate capacitance of a node.
3. In read operation, the voltage of a bit line is discharged to ground through transistor where the gate is charged with the high voltage.

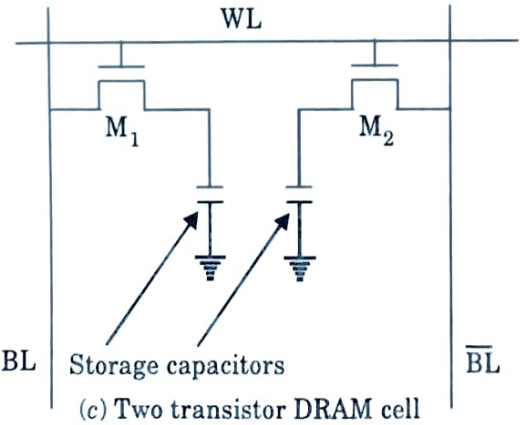
4. In Fig. 4.2.1(*b*), M_3 acts as st· age device and one transistor each for read and write operation.



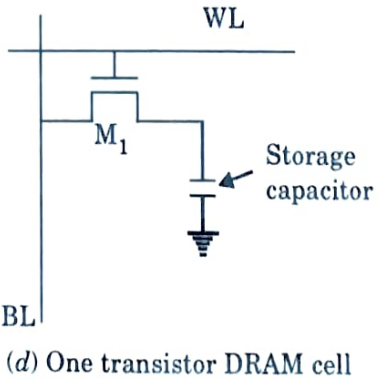
(*a*) Four transistor DRAM cell



(*b*) Three transistor DRAM cell



(*c*) Two transistor DRAM cell



(*d*) One transistor DRAM cell

Fig. 4.2.1. Various configurations of the dynamic RAM cell.

5. During write operation, the write word line is enabled and the voltage of the write bit line is passed onto the gate of storage device through the M_1 transistor.
6. During read operation, the voltage of read bit line is discharged to ground through M_2 and M_3 transistor when the gate voltage of the storage device is high.
7. In Fig. 4.2.1(*c*) and Fig. 4.2.1(*d*) separate capacitor is manufactured for each storage cell instead of relying on the gate and diffusion capacitances of the transistors.
8. The read and write operations of these two cells are almost the same.
9. In write operation, after the word line is enabled, the data are written into the cell through the M_1 or M_2 transistor and stored at the storage capacitor.
10. The read operation is destructive. When the charge stored in the storage is shared with the bit line, its charge can be changed significantly (destructed).

B. Leakage currents in DRAM cells and refresh operation :

1. Different leakage mechanisms contribute to the decay of charge at the cell and total leakage current is expressed as

$$I_{\text{leakage}} = I_{\text{sub}} + I_{\text{tunneling}} + I_j + I_{\text{cell-to-cell}}$$

where I_{sub} is the leakage current through the cell access transistor, $I_{\text{tunneling}}$ is the tunneling current through thin dielectric material, I_j is the junction leakage current at the storage node,

$I_{\text{cell-to-cell}}$ is the leakage current across the field oxide.

2. As the minimum feature size is scaled–down I_j tends to take a smaller portion of the total leakage current and $I_{\text{cell-to-cell}}$ is usually negligible due to the thick field oxide as in the trench isolation technique.
3. Since I_{sub} is a strong function of the threshold voltage, increasing V_{SB} of the access transistor is an effective and practical method to reduce this leakage component.
4. The refresh operation to recharge the cell capacitor is accomplished by reading and restoring operations for the cell using different control signals than those used for the normal read/write operations.
5. A feature of all refresh operations is that the DRAM chip does not send the cell data out to the external devices.
6. Fig. 4.2.2 shows typical timing diagrams for refresh operations called ROR($\overline{\text{RAS}}$ – only refresh), CBR($\overline{\text{CAS}}$ – before – $\overline{\text{RAS}}$) refresh and self refresh.

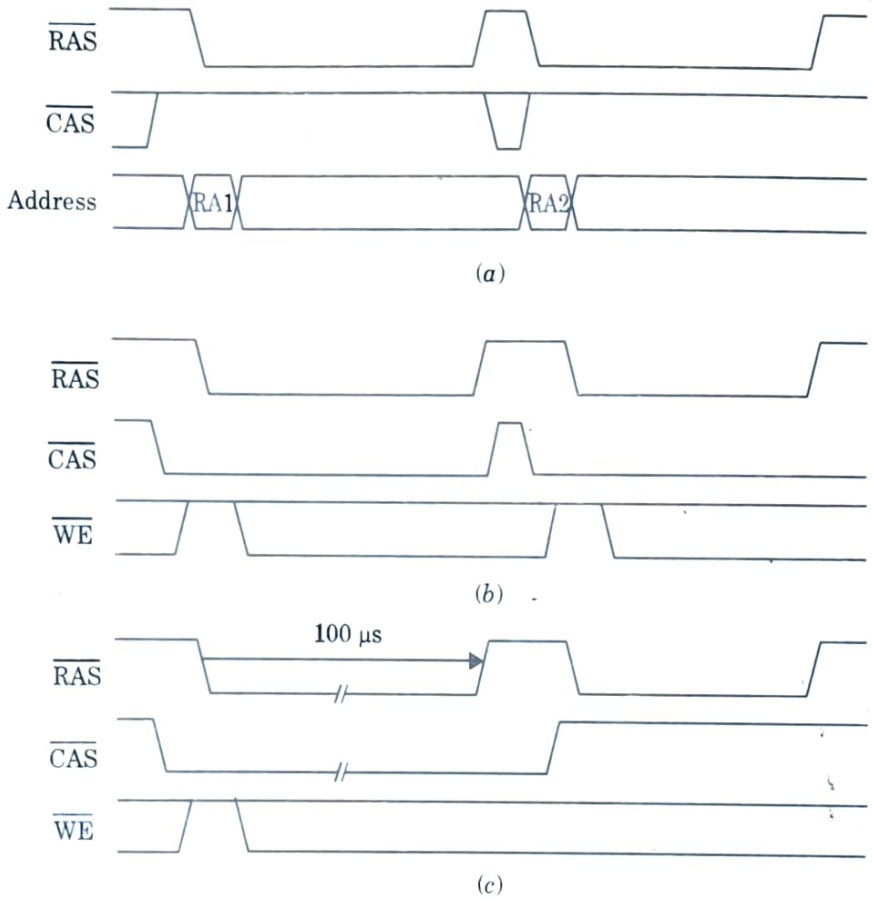


Fig. 4.2.2. Timing diagram for various DRAM refresh operation models, (*a*) ROR refresh, (*b*) CBR refresh, (*c*) Self refresh.

C. Features required to select a proper RAM are :

1. Long life.
2. No need to refresh.
3. Faster.
4. Used as cache memory.
5. Large size.

PART-2

Static RAM.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

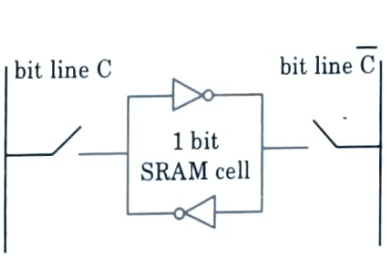
Que 4.3. **What is SRAM ? Explain CMOS SRAM cell design strategy.** **AKTU 2020-21, Marks 07**

Answer

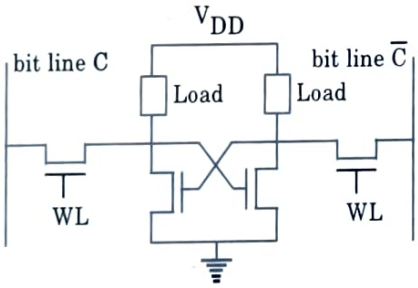
A. SRAM : Refer Q. 4.1, Page 4–2F, Unit-4.

B. CMOS SRAM cell design strategy :

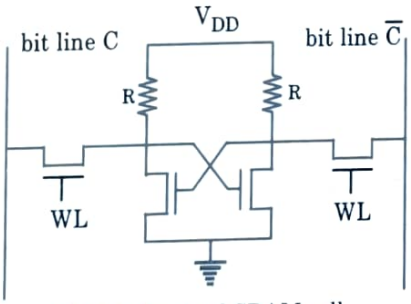
1. The various configuration of the static **RAM** cell is shown in Fig. 4.3.1.
2. Fig. 4.3.1(*a*) : Two complementary access switches consisting of nMOS pass transistor are implemented to connect the 1-bit SRAM cell to the complementary bit lines.



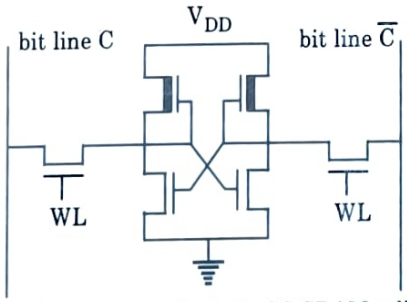
(*a*) Symbolic representation of the two inverter latch circuit



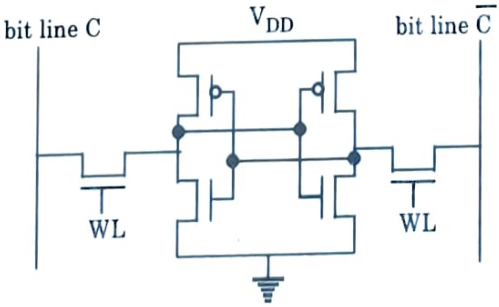
(*b*) Generic circuit topology



(*c*) Resistive load SRAM cell



(*d*) Depletion load nMOS SRAM cell



(*e*) Full CMOS SRAM cell

Fig. 4.3.1. Various configuration of the static RAM cell.

3. Fig. 4.3.1(*b*) : The generic structure of MOS static RAM cell, consisting of two cross coupled inverters and two access transistors.

4. The load devices may be polysilicon resistors, depletion type nMOS transistor, or pMOS transistor, depending on the type of the memory cell.
5. In Fig. 4.3.1(*c*) : We use a resistive load inverters with undoped polysilicon resistors.
6. The resistors can be stacked on top of the cell, thereby reducing the cell size to four transistors, as opposed to the six transistor cell topologies.
7. If multiple polysilicon layers are available, one layer can be used for the gates of the enhancement type nMOS transistor, while another can be used for the load resistors and interconnects.
8. Fig. 4.3.1(*d*) : The six transistor depletion load nMOS SRAM cell can be easily implemented with one polysilicon and one metal layer, and the cell size tends to be relatively small, especially with the use of buried metal diffusion contacts.
9. The static power consumption of the depletion load SRAM cell, however, makes it an unsuitable candidate for high density SRAM arrays.
10. Fig. 4.3.1(*e*) : The full CMOS SRAM cell are currently most popular due to the lowest static power dissipation among the various circuit configurations and compatibility with current logic process.

Que 4.4. Draw the circuit diagram of SRAM and explain read and write operation.

AKTU 2017-18, Marks 10

AKTU 2019-20, Marks 07

Answer

A. Circuit diagram of SRAM :

The circuit diagram of SRAM cell is shown in Fig. 4.4.1.

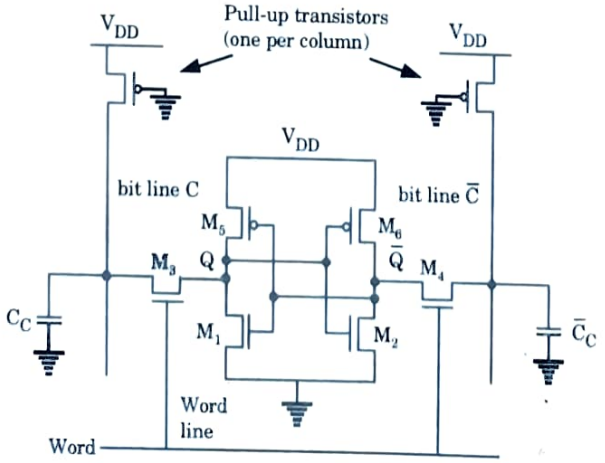


Fig. 4.4.1. Circuit topology of the CMOS SRAM cell.

B. Read operation of SRAM :

1. Assume Q is initially 0 and thus $\bar{Q}$ is initially 1. $\bar{Q}$ and bit line $\bar{C}$ both should remain 1.
2. When the word line is raised, bit line C should be pull down through driver and access transistors M_1 and M_3.
3. At the same time C is being pull down, node Q tends to rise. Q is held low by M_1, but raised by current flowing in from M_3. Hence, the driver M_1 must be stronger than the access transistor M_3.
4. Specifically, the transistor must be ratioed such that node Q remains below the switching threshold of the M_6/M_2 inverter. This constraint is called read stability.
5. Waveforms for the read operation are shown in Fig. 4.4.2 as 0 is read onto C.

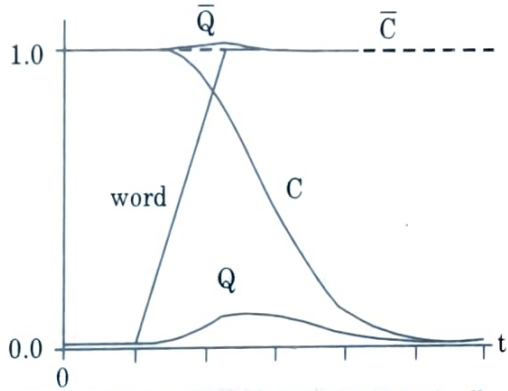


Fig. 4.4.2. Read operation of SRAM cell.

C. Write operation :

1. Assume Q is initially 0 and that we wish to write a 1 into the cell. C is precharged high and left floating. $\bar{C}$ is pulled low by a write driver.
2. We know on account of read stability constraint that C will be unable to force Q high through M_3, hence, the cell must be written by $\bar{Q}$ low through M_4.

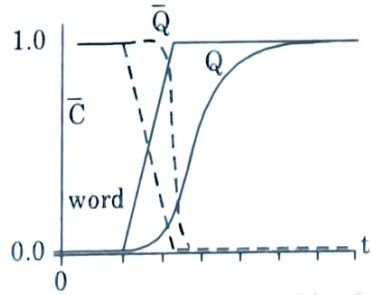


Fig. 4.4.3. Write operation of SRAM cell.

3. M_6 opposes this operation, thus, M_6 must be weaker than M_4, so that $\bar{Q}$ can be pulled low enough. This constraint is called writability.

4. Once $\bar{Q}$ falls low, M_1 turns off and M_5 turns on, pulling Q high as desired.

Que 4.5. Explain read/write operation of SRAM memory cell. How 1-bit cell is used in bigger memory system.

AKTU 2018-19, Marks 10

Answer

A. Read/write operation of SRAM : Refer Q. 4.4, Page 4–8F, Unit-4.

B.

1. A SRAM chip is made from an array of cells, where each cell can store one bit information. A cell is made from flip-flop.
2. The flip-flops consist of six (or four) transistors. The basic block of a SRAM cell is shown in Fig. 4.5.1.
3. Since these devices require only power supply to retain the data, they are also referred as static memory.
4. Address line is used to select the cell while the data bit is written to (or read from) the cell using a data bit line.
5. A large number of such cells are arranged in matrix form to get biggest storage capacity.
6. The 61xx family of memory chips is SRAM, where xx in the chip number represent capacity in kbits, for example, 6116 has 16 kbits (organized as 2k × 8) bits.

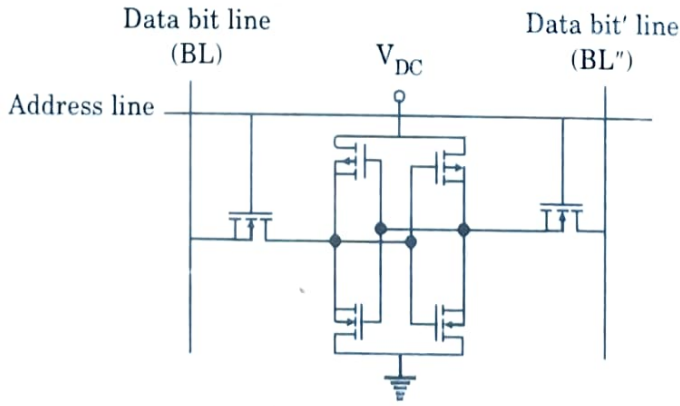


Fig. 4.5.1. SRAM memory cells.

PART-3

Non-volatile Memories.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.6. **What do you mean by ROM ? Also explain NAND based ROM array design style.**

Answer

A. ROM : Refer Q. 4.1, Page 4–2F, Unit-4.

B. NAND based ROM array :

1. Fig. 4.6.1 shows a NAND-based ROM array.

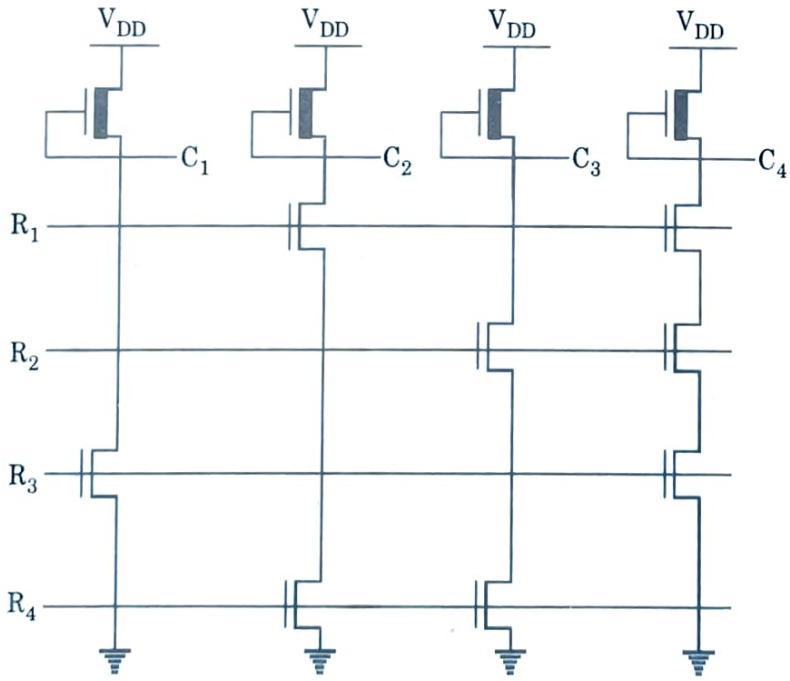


Fig. 4.6.1. 4 bit × 4 bit NAND-based ROM array.

R_1	R_2	R_3	R_4	C_1	C_2	C_3	C_4
0	1	1	1	0	1	0	1
1	0	1	1	0	0	1	1
1	1	0	1	1	0	0	1
1	1	1	0	0	1	1	0

2. Here each bit line consists of a depletion load NAND gate, driven by some of the row signals.
3. In normal operation, all word lines are held at the logic high voltage except for the selected line, which is pulled down to logic low level.
4. If a transistor exists at the crosspoint of a column and the selected row, that transistor is turned OFF and the column voltage is pulled high by the load device.
5. If no transistor exists at that particular crosspoint, the column voltage is pulled low by the other nMOS transistors in the multi-input NAND structure.
6. A logic '1' bit in stored by the presence of a transistor that can be deactivated, while a logic '0' bit is stored by a shorted or normally on transistor at the crosspoint.
7. The NAND based ROM array can be fabricated initially with a transistor connection present at every row-column intersection.
8. A '0' bit is then stored by lowering the threshold voltage of the corresponding nMOS transistor through a channel implant, so that the transistor remains on regardless of the gate voltage.
9. Fig. 4.6.2 shows a sample 4 bit × 4 bit layout of the implant mask NAND ROM array.

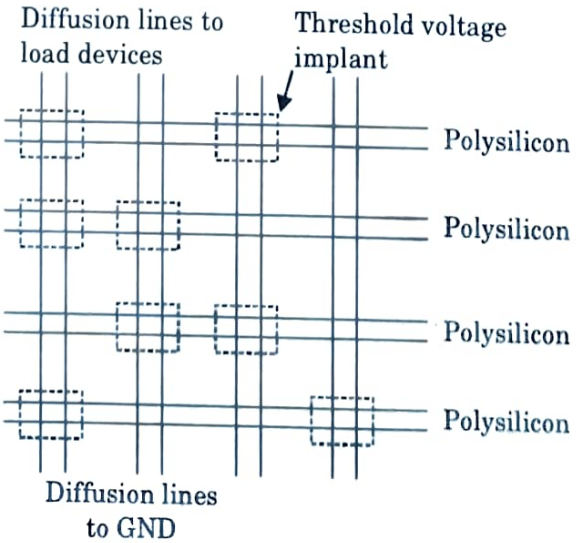


Fig. 4.6.2. Implant mask layout of the NAND ROM array.

10. Vertical columns of n-type diffusion intersect at regular intervals with horizontal rows of polysilicon, which results in an nMOS transistor at each intersection point.
11. The threshold voltage of '0' bit transistors are lowered below 0 V through implant.

PART-4

Flash Memories.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.7. Explain flash memory.

Answer

1. The flash memory cell consists of one transistor with a floating gate, whose threshold voltage can be changed repeatedly by applying an electrical field to its gate.
2. The memory cell can have two threshold voltages corresponding to the existence of charges at the floating gate.
3. When electrons are accumulated at the floating gate, the threshold voltage of the memory cell becomes higher and the memory cell is considered to be in a '1' state as a convention.
4. This is because the memory cell is not turned on with the read signal voltage applied to the control gate and the bit line precharge level is maintained.
5. The threshold voltage of the memory cell can be lowered by removing electrons from the floating gate and the memory cell is regarded to be in a '0' state.
6. In such case the cell transistor is turned on with the applied voltage and the bit line is discharged to ground.
7. Therefore, the cell data of flash memory is programmed by either storing or ejecting electrons in the floating gate of a MOS transistor through channel hot electron injection or Fowler Nordheim tunneling mechanism.

8. The schematic cross-section views of the flash memory cell for two programming mechanisms are shown in Fig. 4.7.1(*a*) and Fig. 4.7.1(*b*) respectively.

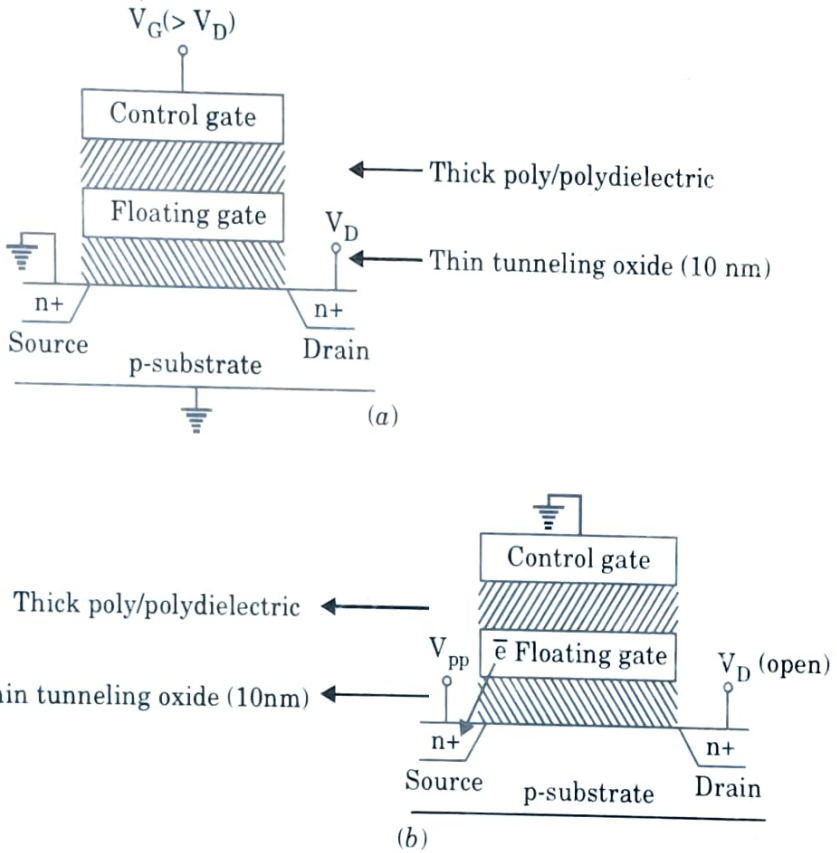


Fig. 4.7.1. Data programming and erasing methods in the flash memory (*a*) Hot electron injection mechanism (*b*) Fowler nordheim tunneling mechanism.

9. When the high voltage is applied to the control gate and across the drain to source, electrons are heated by the high lateral electric field.
10. Avalanche break down occurs at the near of the drain and electron hole pairs are generated by the impact ionization.
11. The high voltage on the control gates attracts and injects electrons into the floating gate through the oxide and the holes flow to the substrate as the substrate current.
12. When the 0 V and a high voltage of 12 V are applied to the control gate and the source, electrons at the floating gate are ejected to the source by the tunneling effect.

Que 4.8. Discuss NAND flash memory cell.

Answer

1. By placing eight or sixteen cells in series, the memory cell area can be reduced by the elimination of contacts in cells. The equivalent circuit of the eight bit NAND cell structure is shown in Fig. 4.8.1.

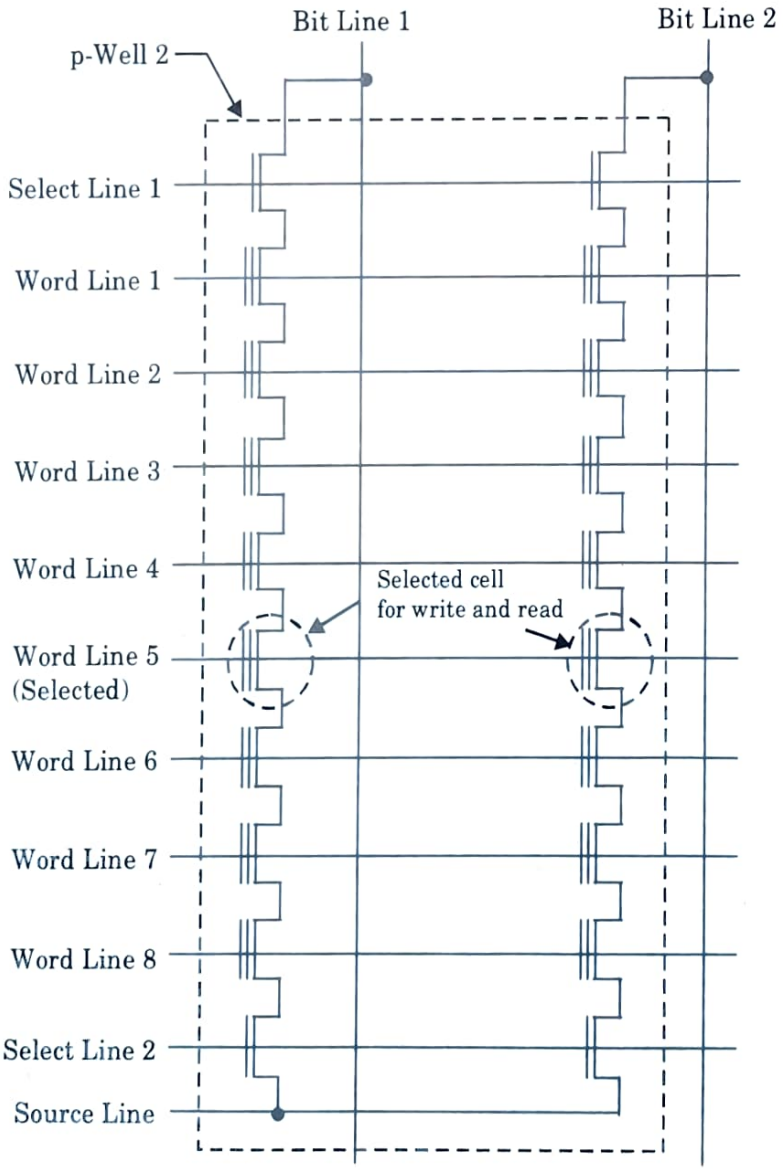


Fig. 4.8.1.

2. Bias conditions for erase, programming and read operations are shown in Table 4.8.1.

Table 4.8.1. Bias conditions of the NAND cells for erase, programming, and read operations.

Signal	Operation		
	Erase	Programming	Read
Bit line 1	Open	0 V	1
Bit line 2	Open	0 V	1 V
Select line 1	Open	5 V	5 V
Word line 1	0 V	10 V	5 V
Word line 2	0 V	10 V	5 V
Word line 3	0 V	10 V	5 V
Word line 4	0 V	10 V	5 V
Word line 5	0 V	20 V	0 V
Word line 6	0 V	10 V	5 V
Word line 7	0 V	10 V	5 V
Word line 8	0 V	10 V	5 V
Select line 2	Open	0 V	5 V
Source line	Open	0 V	0 V
p-well 2	20 V	0 V	0 V
n-sub	20 V	0 V	0 V

3. The NAND cell uses *F-N* tunneling mechanism for the erase operation. A high voltage (*e.g.*, 20 V) is applied to source line; *p*-well 2 and *n*-substrate while 0 V to all the word lines to eject electrons from the floating gate to the *p*-well 2.

4. In the programming operation, a high voltage (*e.g.*, 20 V) is applied only to the select word line (word line 5) and a moderate voltage (*e.g.*, 10 V) is applied to all unselected word lines. The NAND cell uses the *pn* mechanism to program the cell.

5. To read the cell data, 0 V is applied to the selected word line while 5 V is applied to the select lines and all the other unselected word lines.

Que 4.9. **Compare the characteristic of NOR and NAND cell.**

Answer

S. No.	Characteristic	NOR cell	NAND cell
1.	Erase method	Fowler-Nordheim tunneling	Fowler-Nordheim tunneling
2.	Programming method	Hot electron injection	Fowler-Nordheim tunneling
3.	Erase speed	Slow	Fast
4.	Program speed	Fast	Slow
5.	Read speed	Fast	Slow
6.	Cell size	Large	Small
7.	Scalability	Difficult	Easy
8.	Application	Embedded system	Mass storage

PART-5

Pipeline Architecture.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.10. Examine the process to convert single stage structure into a pipeline structure. Is pipeline approach significant in reducing the power consumption ?

Answer

1. In many applications VTCMOS and MTCMOS are infeasible due to technological limitations. Pipelining and hardware replication techniques offer feasible alternatives for maintaining the system performance.
2. Consider the single functional block as shown in Fig. 4.10.1(*a*) with logic function F (INPUT) of input vector, INPUT.
3. Both input and output vectors are sampled through register arrays, driven by a clock signal CLK.
4. Maximum sampling frequency is f_{CLK} at power supply voltage of V_{DD} *i.e.*, maximum input-to-output propagation delay $\tau_{p\,max}$ of this logic block is equal to or less than $T_{CLK} = 1 / f_{CLK}$.

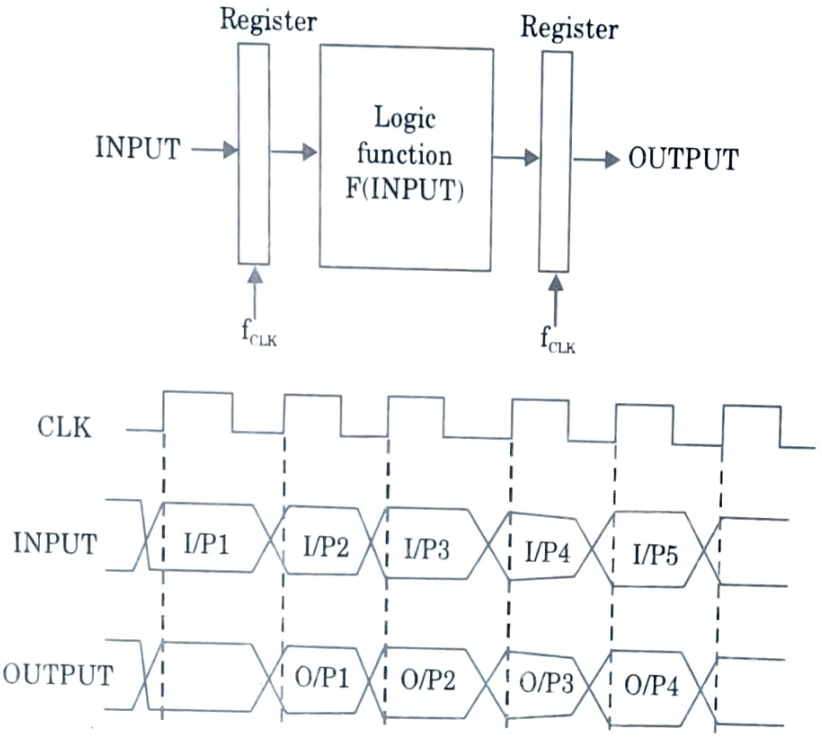


Fig. 4.10.1.

5. A new input vector is latched into the input register array at each clock cycle, and the output data become valid with a latency of one cycle.
6. Let C_{total} be the total capacitance switched every clock cycle. C_{total} be the total capacitance switched in input register and output register array and the capacitance switched to implement the logic function.
7. The dynamic power consumption of this structure can be found as

$$P_{reference} = C_{total} \cdot V^2_{DD} \cdot f_{CLK} \cdot \quad ...(4.10.1)$$

8. Now consider an N-stage pipelined structure for implementing the same logic function. The F (INPUT) has been partitioned into N successive stages, and a total of $(N-1)$ register arrays have been introduced.
10. If all stages of partitioned function have approximately equal delays of

$$\tau_p(\text{pipeline-stage}) = \frac{\tau_{p.max}(\text{input} - \text{to} - \text{output})}{N} = T_{CLK} \quad ...(4.10.2)$$

Then the logic blocks between two successive registers can operate N-times slower while maintaining the same functional throughput as before *i.e.*, V_{DD} can be reduced to value of $V_{DD\ new}$ to effectively slow down the circuit by factor of N.

11. The dynamic power consumption of N-stage pipelined structure with low V_{DD} and with same functional throughput as the single stage structure can be found as

$$P_{\text{pipeline}} = [C_{\text{total}} + (N-1)\,C_{\text{reg.}}].V^2_{DD\,\text{new}} f_{\text{CLK}} \quad ...(4.10.3)$$

Where, C_{reg} = Capacitance switched by each pipeline register.

12. Then, the power reduction factor achieved in a N-stage pipeline structure is

$$\frac{P_{\text{pipeline}}}{P_{\text{reference}}} = \frac{[C_{\text{total}} + (N-1)C_{\text{reg}}] \cdot V^2_{DD,\text{new}} \cdot f_{\text{CLK}}}{C_{\text{total}} \cdot V^2_{DD} \cdot f_{\text{CLK}}} = \left[1 + \frac{C_{\text{reg}}}{C_{\text{total}}}(N-1)\right] \frac{V^2_{DD\,\text{new}}}{V^2_{DD}} \quad ...(4.10.4)$$

13. A total of $(N-1)$ register arrays have to be added to convert original single-stage structure into a pipeline.
14. With little area overhead, this approach also increases the latency from one to N-clock cycles, but latency is not a very significant concern.

PART-6

Low-power CMOS Logic Circuits : Introduction, Overview of Power Consumption.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.11. What are the sources of power dissipation in CMOS circuits ? Explain dynamic and static power consumption.

AKTU 2017-18, Marks 10

OR

Describe leakage (static) power dissipation and dynamic power dissipation.

AKTU 2018-19, Marks 10

OR

Explain various types of power dissipation in CMOS circuits.

AKTU 2019-20, Marks 07

Answer

A. Sources of power dissipation :

1. Power dissipation in CMOS circuits comes from two components :

A. Dynamic dissipation due to :

a. Charging and discharging load capacitances as gates switch.

b. "Short-circuit" current while both pMOS and nMOS stacks are partially ON.

B. Static dissipation due to :

a. Subthreshold leakage through OFF transistors.

b. Gate leakage through gate dielectric.

c. Junction leakage from source/drain diffusions.

d. Contention current in ratioed circuits.

Putting this together gives the total power of a circuit

$$P_{\text{dynamic}} = P_{\text{switching}} + P_{\text{short circuit}}$$

$$P_{\text{static}} = (I_{\text{sub}} + I_{\text{gate}} + I_{\text{Junct}} + I_{\text{contention}})V_{DD}$$

$$P_{\text{total}} = P_{\text{dynamic}} + P_{\text{static}}$$

2. Power can also be considered in active, standby and sleep modes.

3. Active power is the power consumed while the chip is doing useful work. It is usually dominated by $P_{\text{switching}}$.

4. Standby power is the power consumed while the chip is idle. If clocks are stopped and ratioed circuits are disabled, the standby power is set by leakage.

5. In sleep mode, the supplies to unneeded circuits are turned off to eliminate leakage. This drastically reduces the sleep power required, but the chip requires time and energy to wake up so sleeping is only viable if the chip will idle for long enough.

B. Static power consumption :

1. Considering the static CMOS inverter shown in Fig. 4.11.1, if the input = '0', the associated nMOS transistor is OFF and the pMOS transistor is ON. The output voltage is V_{DD} of logic '1'.

2. When the input = '1', the associated nMOS transistor is ON and the pMOS transistor is OFF. The output voltage is 0 volts (GND).

3. Note that one of the transistors is always OFF when the gate is in either of these logic states. Ideally, no current flows through the OFF transistor so the power dissipation is zero when the circuit is quiescent, *i.e.*, when no transistors are switching.

4. Zero quiescent power dissipation is principle advantage of CMOS over competing transistor technologies. However, secondary effects including subthreshold conduction, tunneling and leakage lead to small amounts of static current flowing through the OFF transistor.

5. Assuming the leakage current is constant so instantaneous and average power are the same; the static power dissipation is the product of total leakage current and the supply voltage.

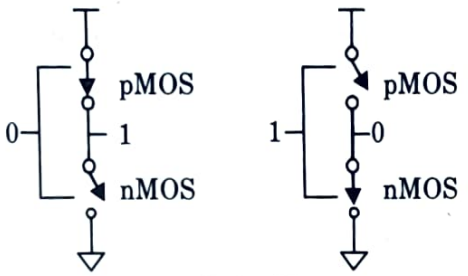


Fig. 4.11.1.

$$P_{static} = I_{static} V_{DD} \qquad ...(4.11.1)$$

B. Dynamic power consumption :

1. The primary dynamic dissipation component is charging the load capacitance. Suppose a load C is switched between GND and V_{DD} at an average frequency of f_{sw}.
2. Over any given interval of time T, the load will be charged and discharged Tf_{sw} times. Current flows from V_{DD} to the load to charge it. Current then flows from the load to GND during discharge.
3. In one complete charge/discharge cycle, a total charge of $Q = CV_{DD}$ is thus transferred from V_{DD} to GND.
4. The average dynamic power dissipation is

$$P_{dynamic} = \frac{1}{T}\int_0^T i_{DD}(t) V_{DD} dt = \frac{V_{DD}}{T}\int_0^T i_{DD}(t) dt \qquad ...(4.11.2)$$

Taking the integral of the current over some interval T as the total charge delivered during that time, we simplify to

$$P_{dynamic} = \frac{V_{DD}}{T}[Tf_{sw} CV_{DD}] = CV_{DD}^2 f_{sw} \qquad ...(4.11.3)$$

5. Because most gates do not switch every clock cycle, it is often more convenient to express switching frequency f_{sw} as an activity factor α times the clock frequency f. Now, the dynamic power dissipation may be rewritten as:

$$P_{dynamic} = \alpha\, CV_{DD}^2 f \qquad ...(4.11.4)$$

Que 4.12. Calculate the average dynamic switching power consumption in CMOS logic circuits.

Answer

1. Switching power dissipation represents the power dissipated during a switching event.
2. During charge-up phase, the output node charges upto V_{DD} and one half of energy drawn from power supply is dissipated as heat in the conducting pMOS transistors.

3. No energy is drawn from power supply during charge down-phase; still energy stored in output capacitance is dissipated as heat in conducting nMOS transistors.
4. For the dynamic power dissipation during switching consider the circuit is shown in Fig. 4.12.1.
5. In Fig. 4.12.1 the two-input NOR gate driving two NAND gates through interconnection lines.

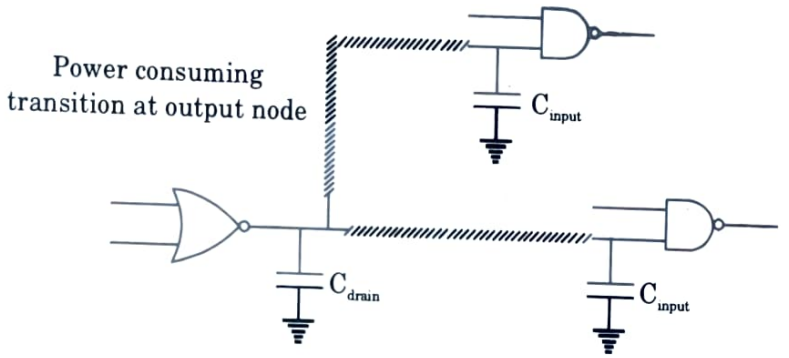


Fig. 4.12.1.

6. The total capacitive load at the output of NOR gate consists of :

i. Output node capacitance of the gate itself.

ii. The total interconnect capacitance.

iii. The input capacitance of driven gates.

7. The average dynamic (switching) power consumption in CMOS logic circuits is,

$$P_{avg.} = \frac{1}{T} C_{load} V^2_{DD} \qquad ...(4.12.1)$$

or,

$$P_{avg.} = C_{load} V^2_{DD} f_{CLK} \qquad ...(4.12.2)$$

8. $P_{avg.}$ is independent of all transistor characteristics and sizes as long as a full voltage swing is between 0 and V_{DD}.
9. We will introduce α_T (node transition factor), which is the effective number of power consuming voltage transitions experienced per clock cycle.

$$P_{avg.} = \alpha_T \cdot C_{load} V^2_{DD} f_{CLK} \qquad ...(4.12.3)$$

10. In complex CMOS logic, nodes also makes full or partial voltage transitions. Since there is a parasitic capacitance associated with each internal node, these internal transitions contribute to the overall power dissipation of the circuit.
11. An internal node may undergo several transitions while the output node voltage of the circuit remains unchanged as shown in Fig. 4.12.2.

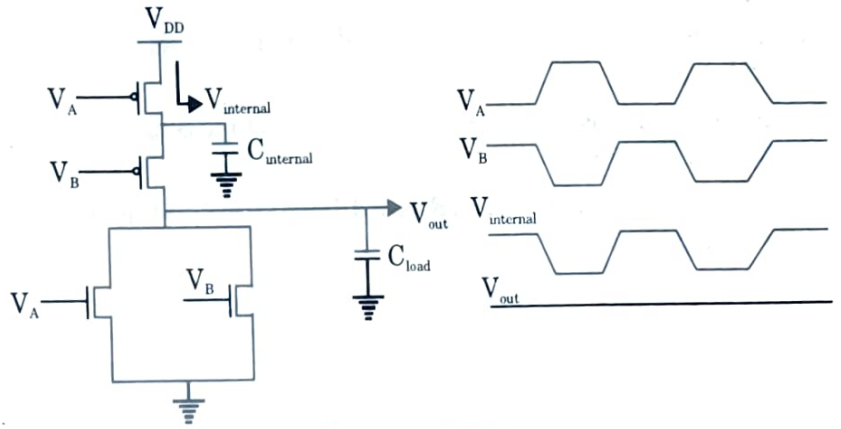


Fig. 4.12.2.

12. Now, $$P_{avg.} = \left(\sum_{i=1}^{\text{No. of Nodes}} \alpha_{Ti} C_i V_i \right) \cdot V_{DD} \cdot f_{CLK}$$

where, V_i = Internal node voltage swing < full voltage swing of V_{DD}
C_i = Parasitic capacitance associated with each node

Que 4.13. Analyze the CMOS inverter circuit for short-circuit power dissipation during the switching events across the nMOS and pMOS transistors.

Answer

1. If a CMOS inverter is driven with input voltage waveforms with finite rise and fall times, both nMOS and pMOS may conduct simultaneously for a short amount of time, forming direct current path between power supply and the ground as shown in Fig. 4.13.1.

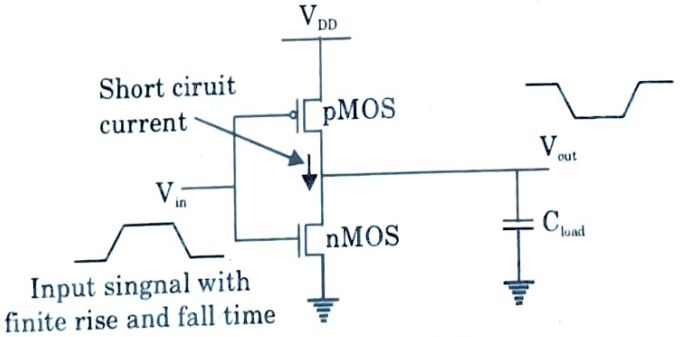


Fig. 4.13.1.

2. This direct current does not contribute to the charging of capacitances and is called the short-circuit current component.
3. The magnitude of short-circuit current component will be approximately the same during both the rising input transition and falling input transition.

4. The pMOS also conducts current needed to charge output load capacitance, but only during the falling-input transition.
5. Let a CMOS inverter with $k_n = k_p = k$, $V_{T,n} = |V_{T,p}| = V_T$ and ($\tau_{rise} = \tau_{fall} = \tau$), the time averaged short circuit current drawn from the power supply is

$$I_{avg.}\text{ (short circuit)} = \frac{1}{12} \cdot \frac{k\tau f_{CLK}}{V_{DD}} (V_{DD} - 2V_T)^3 \qquad ...(4.13.1)$$

Hence, the short-circuit power dissipation becomes

$$P_{avg.}\text{ (short-circuit)} = \frac{1}{12} \cdot k \cdot \tau \cdot f_{CLK} (V_{DD} - 2V_T)^3 \qquad ...(4.13.2)$$

6. The short-circuit power dissipation is linearly proportional to the input signal rise and fall times, and also to the transconductance of the transistors. Hence reducing the input transition times will decrease the short-circuit current component.
7. Consider the CMOS inverter with large output capacitance and small input transition times.
8. During the rising input transition the output voltage is V_{DD} until the input voltage completes its swing *i.e.*, input reaches to its final value.
9. Both nMOS and pMOS are ON during transition, the pMOS transistor cannot conduct a significant amount of current since the voltage drop between its source and drain is almost zero.
10. Similarly the output voltage will remain at approximately 0 V during the falling input transition and it will start to rise only after the input voltage completes its swing.
11. The peak value of supply current to charge up the output load capacitance is larger in this case.
12. This is because pMOS transistor remains in saturation during entire input transition as opposed to previous case where transistor leaves the saturation region before the input transition is completed.

Que 4.14. Discuss the reverse leakage current paths and sub-threshold leakage current path in CMOS inverter with high input voltage condition.

Answer

1. The nMOS and pMOS used in CMOS generally have non-zero reverse leakage and sub-threshold currents.
2. These currents contribute to overall power dissipation even when transistors are not undergoing any switching event.
3. Consider a CMOS inverter with high input voltage, nMOS is turned ON and output discharged to zero.

4. Although pMOS is OFF, there will be a reverse potential difference of V_{DD} between its drain and the n-well causing diode leakage through the drain junction.
5. The n-well region of pMOS is also reverse-biased with V_{DD} with respect to p-type substrate. Therefore another significant leakage current component exists because of the n-well junction as shown in Fig. 4.14.1.
6. A similar situation can be observed when the input voltage is zero, and the output voltage is charged up to V_{DD} through pMOS.
7. Then, the reverse potential difference between nMOS drain and p-type substrate causes reverse leakage current which is also drawn from power supply (through pMOS).

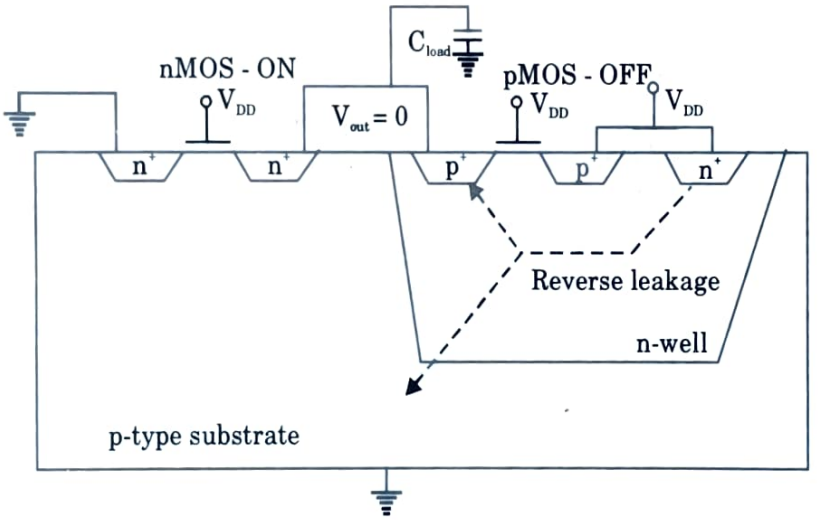


Fig. 4.14.1.

8. The reverse leakage current of p-n junction is expressed by,

$$I_{\text{reverse}} = A \cdot J_s \left(e^{\frac{qV_{\text{bias}}}{kT}} - 1 \right) \qquad \ldots(4.14.1)$$

where, V_{bias} = Reverse bias voltage across the junction

J_S = Reverse saturation current density

= 1-5 pA/μm^2

A = Junction area

9. J_S increases quite significantly with temperature. The reverse leakage occurs even during the standby operation when no switching takes place.
10. Another component of leakage is sub-threshold current due to carrier diffusion in weak inversion. This current exhibits an exponential dependence on the gate voltage.

11. Sub-threshold current become significant when gate-to-source voltage is less than or equal to threshold voltage of device.
12. Sub-threshold current can occur even when there is no switching activity in the circuit as shown in Fig. 4.14.2.

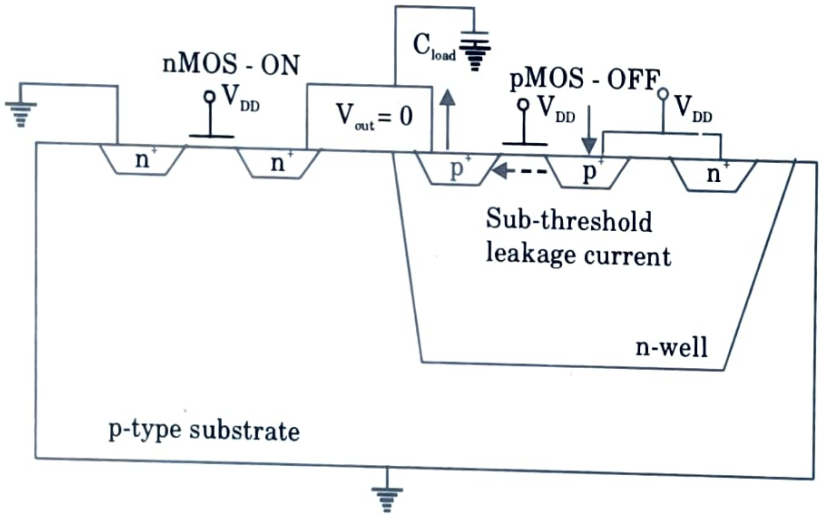


Fig. 4.14.2.

$$I_D(\text{sub-threshold}) \cong \frac{qD_n W x_c n_o}{L_B} \cdot e^{\frac{q\phi_t}{kT}} \cdot e^{\frac{q}{kT}(A V_{GS} + B \cdot V_{DS})} \quad \ldots(4.14.2)$$

13. To limit sub-threshold current avoid very low threshold voltages, although with some speed penalty. That's why the V_{GS} of nMOS remain safely below $V_{T,n}$ when the input is logic zero and $|V_{GS}|$ of pMOS safely below $|V_{T,p}|$ when input is logic one.
14. The total power dissipation in CMOS digital circuit is

$$P_{\text{total}} = \alpha_T \cdot C_{\text{load}} \cdot V^2_{DD} \cdot f_{CLK} + V_{DD} (I_{\text{short circuit}} + I_{\text{leakage}} + I_{\text{static}}) \quad \ldots(4.14.3)$$

I_{static} denotes the DC current component drawn from the power supply.

Que 4.15. Explain the variable threshold CMOS circuits.

AKTU 2017-18, Marks 10

OR

Briefly explain variable threshold CMOS (VTCMOS) circuit.

AKTU 2019-20, Marks 07

Answer

1. In VTCMOS (Variable-Threshold CMOS) circuit, the transistors are designed inherently with low V_T and substrate bias voltages of nMOS and pMOS are generated by a variable substrate bias control circuit as shown in Fig. 4.15.1.

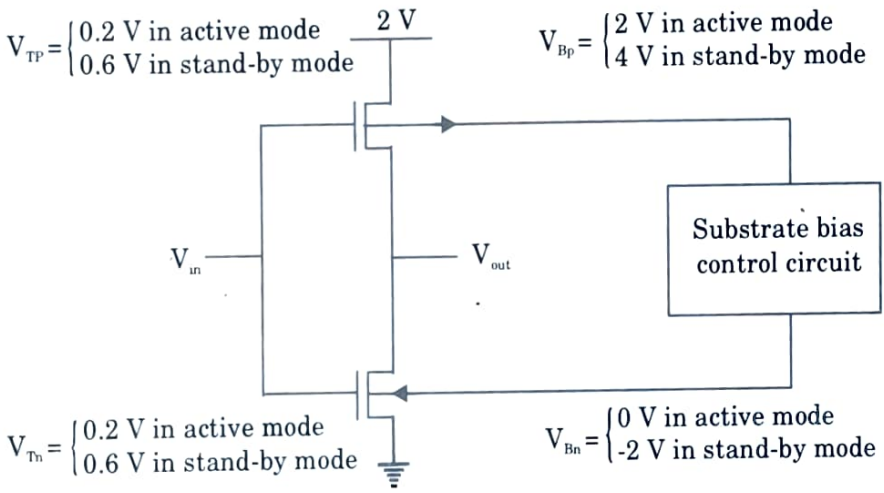


Fig. 4.15.1.

2. In active mode substrate bias voltage of nMOS is $V_{Bn} = 0$ and substrate bias voltage of pMOS is $V_{Bp} = V_{DD}$. Thus, inverter circuit do not experience any back gate bias effect.
3. The circuit with low V_{DD} and V_T benefits both low power dissipation (due to low V_{DD}) and high switching speed (due to low V_T).
4. When inverter circuit in stand-by mode, the substrate bias control circuit generates a lower substrate bias voltage for nMOS and higher for pMOS. As a result the magnitude of V_{Tn} and V_{Tp} increases due to back gate bias effect.
5. Since, the sub-threshold leakage current drops exponentially with increasing V_T the leakage power in stand-by mode can be reduced.
6. VTCMOS is also used for automatic control of V_T which is called self-adjusting threshold voltage scheme (SATS).
7. VTCMOS requires twin/triple well CMOS technology to provide different substrate bias voltage to different parts of the chip.
8. Also, separate power pins may be required if substrate bias voltage levels are not generated on chip. The additional area is still negligible compared to overall chip area.

Que 4.16. Discuss the low power MTCMOS VLSI design techniques.

OR

Explain the following circuits :

i. Variable threshold CMOS circuits

ii. Multiple threshold CMOS circuits.

AKTU 2018-19, Marks 10

Answer

A. VTCMOS : Refer Q. 4.15, Page 4–26F, Unit-4.

B. MTCMOS :

1. MTCMOS VLSI design technique which can be applied for reducing leakage currents in low voltage circuits in stand-by mode is based on different V_T in the circuit.
2. The low V_T transistors are typically used to design the logic gates where switching speed is essential, whereas high V_T transistors are used to effectively isolate the logic gates in stand-by and to prevent leakage dissipation.

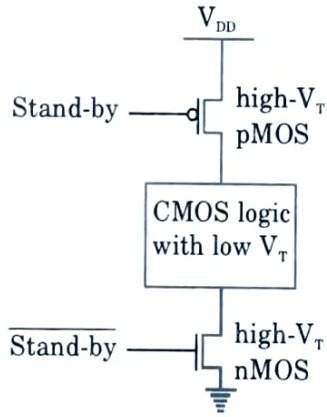


Fig. 4.16.1.

3. The circuit structure of Multiple Threshold CMOS (MTCMOS) is shown in Fig. 4.16.1.
4. In active mode, the high V_T transistors are turned ON and the logic gates consisting of low V_T transistors can operate with low switching power dissipation and small propagation delay.
5. In stand-by-mode, the high V_T transistors are turned OFF and the conduction paths for any subthreshold leakage currents that may originate from the internal low V_T circuitry are effectively cut-off.
6. A *D*-latch circuit designed with the MTCMOS is shown in Fig. 4.16.2.
7. The path from input to output consists of low V_T transistors, while cross-coupled inverter pair consisting of high V_T transistors is used to preserve the data in stand-by mode.

8. MTCMOS is conceptually easier than VTCMOS which usually requires substrate bias control mechanism.

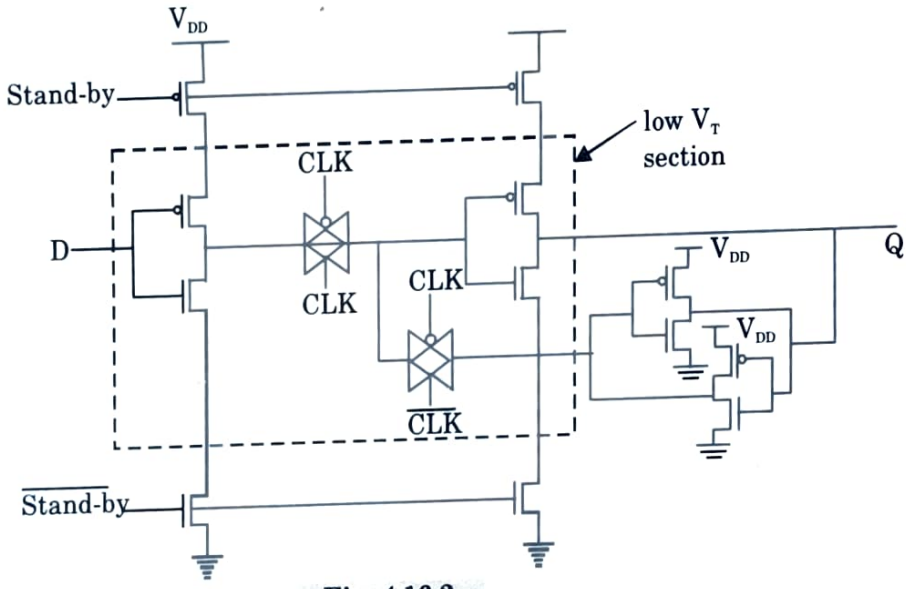


Fig. 4.16.2.

9. The disadvantage of MTCMOS is the presence of series-connected stand-by transistors, which increase the overall circuit area and also add extra parasitic capacitance and delay.

Que 4.17. Explain parallel processing method.

OR

Discuss the various design techniques involved in low power CMOS VLSI circuits. **AKTU 2020-21, Marks 07**

Answer

A. Design techniques : Refer Q. 4.15, Page 4–26F, Q. 4.16, Page 4–27F; Unit-4.

B. Parallel processing approach :

1. Another method for trading off area for low power dissipation is to use parallelism or hardware replication. Consider N identical processing element with function F (INPUT) each in parallel as shown in Fig. 4.17.1.
2. The consecutive input vectors arrive at same rate as in single stage case and are routed to all registers of the N processing blocks.
3. Gated clock signals, with period NT_{CLK} are used to load each register every N clock cycles.

4. Each input register is clocked at lower frequency of f_{CLK}/N, the time allowed to compute the function for each input vector is increased by factor of N.

5. This implies that the V_{DD} can be reduced until the critical path delay equals the new clock period of NT_{CLK}.

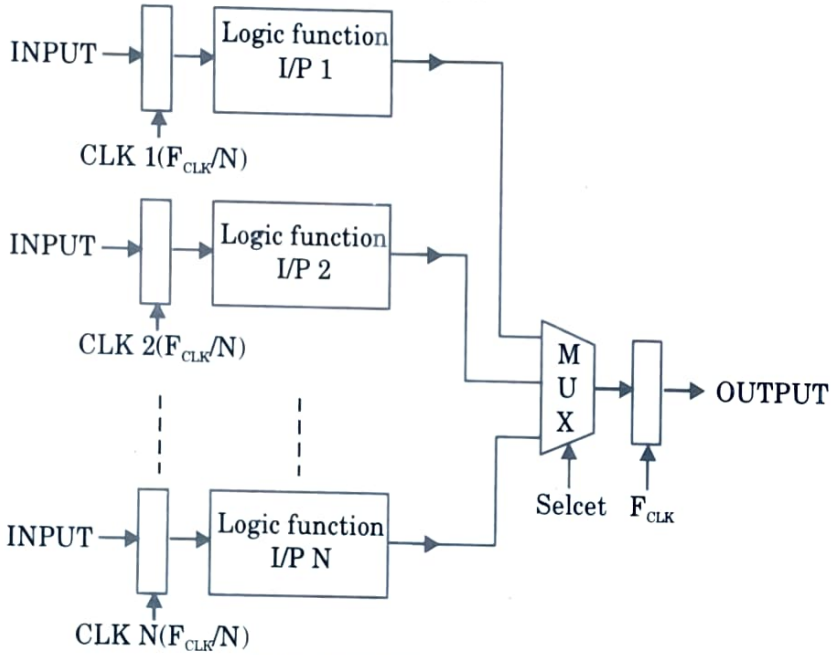


Fig. 4.17.1.

6. The output of the N processing blocks are multiplexed and send to output register ensuring the same data throughput rate as before.

7. The timing diagram of this parallel arrangement is shown in Fig. 4.17.2.

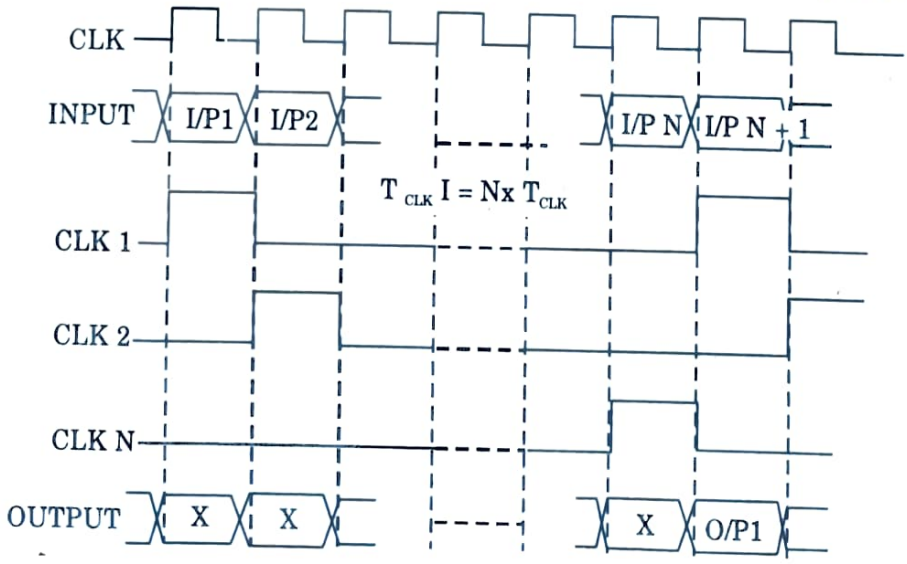


Fig. 4.17.2.

8. The total dynamic power dissipation of the parallel structure is found as the sum of the power dissipated by the input registers and logic blocks at clock frequency of f_{CLK}/N, and output register operating at clock frequency of f_{CLK}.

$$P_{parallel} = N \cdot C_{total} \cdot V^2_{DD,new} \cdot \frac{f_{CLK}}{N} + C_{reg.} V^2_{DD,new} f_{CLK}$$

$$= \left(1 + \frac{C_{reg}}{C_{total}}\right) C_{total} \cdot V^2_{DD,new} f_{CLK}$$

9. The amount of power reduction achievable in N-block parallel implementation is

$$\frac{P_{parallel}}{P_{reference}} = \frac{V^2_{DD\,new}}{V^2_{DD}} \cdot \left(1 + \frac{C_{reg}}{C_{total}}\right)$$

10. Two consequences of this approach are increased area and the increased latency.
11. A total of N identical processing blocks must be used to slow down the operating speed by a factor of N.
12. Actually the silicon area will grow even faster than the number of processors because of signal routing and overhead circuitry.

PART-7

Low-power Design Through Voltage Scaling.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.18. Examine the effects of reducing V_{DD} upon switching power consumption and dynamic performance of the gate. Discuss influence of voltage scaling on power and delay in low-power design.

Answer

1. The reduction of V_{DD} reduces the dynamic power dissipation. This can be seen by following expressions for CMOS inverter circuit.

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

$$\tau_{PLH} = \frac{C_{load}}{k_n(V_{DD} - |V_{T,p}|)}\left[\frac{2|V_{T,p}|}{V_{DD} - |V_{T,p}|} + \ln\left(\frac{4(V_{DD} - |V_{T,p}|)}{V_{DD}} - 1\right)\right]$$

2. If power supply voltage is scaled down while other parameters kept constant, the propagation delay time would increase.
3. The dependence of circuit speed on V_{DD}, may also influence the relationship between dynamic power dissipation and supply voltage.
4. If the circuit is always operated at f_{max} allowed by its propagation delay, the number of switching events per unit time will drop as the propagation delay becomes larger with the reduction of V_{DD}.
5. The above expressions for delay shows that the negative effect of reducing V_{DD} upon the delay can be compensated if the threshold voltage is scaled down accordingly.
6. As shown in Fig. 4.18.1 reducing the threshold voltage from 0.8 V to 0.2 V can improve the delay at $V_{DD} = 2$ V by a factor of 2.

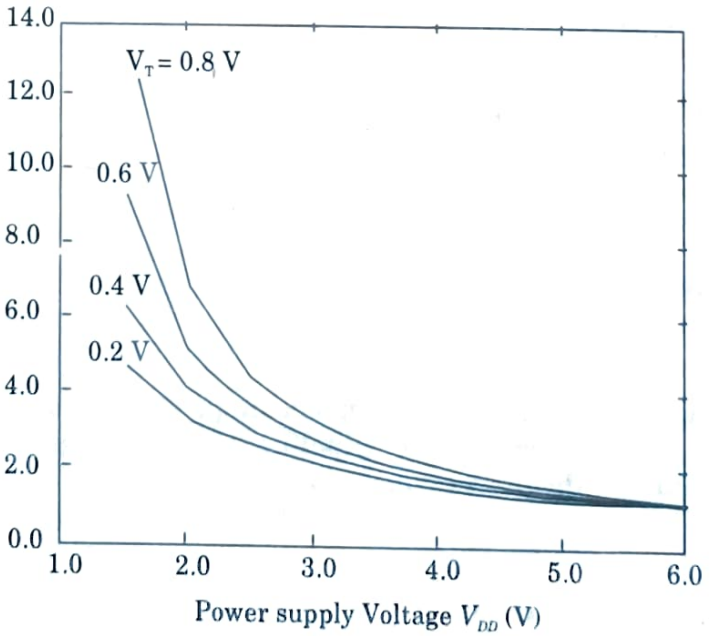


Fig. 4.18.1. Variation of the normalized propagation delay of a CMOS inverter, as a function of the power supply voltage V_{DD} and the threshold voltage V_T.

7. Low V_T transistors raise significant concerns about noise margins and sub-threshold conduction.
8. Smaller V_T lead to smaller noise margins for CMOS logic gates.
9. For $V_T < 0.2$ V leakage due to subthreshold conduction in stand-by mode become a very significant component of overall power consumption.

5 UNIT

Introduction to Testing

CONTENTS

PART-1

Faults in Digital Circuits, Modeling of Faults.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.1. Enlist the different kinds of physical defect (faults) that can occur on a CMOS circuits.

Answer

1. The physical defects include :

i. Defects in silicon substrate.

ii. Photolithographic defects.

iii. Mask contamination and scratches.

iv. Process variations and abnormalities.

v. Oxide defects.

2. The physical defects can cause electrical faults and logical faults. The electrical faults include :

i. Shorts (bridging faults).

ii. Opens.

iii. Transistor stuck-on, stuck-open.

iv. Resistive shorts and opens.

v. Excessive change in threshold voltage.

vi. Excessive steady-state currents.

3. The electrical faults in turn can be translated into logical faults. The logical faults include :

i. Logical stuck-at-0 or stuck-at-1.

ii. Slower transition (delay fault).

iii. AND-bridging, OR-bridging.

Que 5.2. Give a logic circuit example in which stuck-at-0 fault and stuck-at-1 fault are indistinguishable.

Answer

1. For the circuit shown below, the output function is

$$F = ab + \bar{a}c$$

2. Thus the AND gate with b, c as input is redundant. So stuck at 1 fault and stuck at 0 fault are indistinguishable at node X.

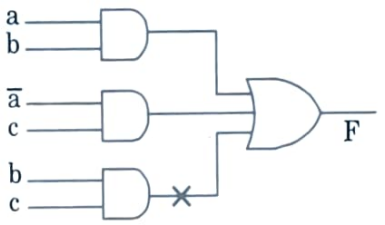


Fig. 5.2.1.

Que 5.3. Differentiate between single stuck at fault and multiple stuck at fault.

Answer

S. No.	Single stuck at fault	Multiple stuck at fault
1.	In single stuck at fault there is only two faults per line, stuck-at-1 and stuck-at-0.	In Multiple stuck at fault there are more than two faults per line.
2.	Complexity in single stuck simulation is less.	Complexity of multiple fault simulation increases with increase in number of faults being handled together.

PART-2

Functional Modeling at the Logic Level, Functional Modeling at the Register.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.4. **Explain functional modeling.**

Answer

1. A functional model of a digital system is a representation of its logic function.
2. Any digital system can be represented by a black box which processes the input to produce its output.

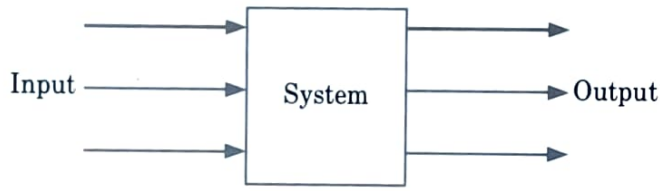


Fig. 5.4.1.

3. The value of the outputs varies with inputs and a certain delay exists.
4. The logic function is the input-output mapping that deals with the value transformation.
5. It makes the relationship between input and output value of a circuit.
6. It does not talk about any timing aspect.
7. Simulation implies calculation of the output (response) of a circuit.
8. Functional models provide the information for simulation.
9. It can be done at two levels :

a. Logic level.

b. RTL (program) level.

Que 5.5. **What are the different functional model at logic level ?**

Answer

Different functional model at logic levels are :

i. Truth table :

1. It is representation of the function in terms of rows and columns.
2. It is the simplest way to represent a circuit.
3. In boolean arithmetic, the possible values are 0 and 1. So for an n-input gate, number of possible outcomes = 2^n.
4. The inputs are given in an increment of 1 from 0 to $2^n - 1$ in boolean form :

ii. Primitive cubes : Primitive cubes are compressed form of truth table.

iii. Binary Decision Diagram (BDD) :

1. BDD is a way to reduce the number of entries for describing the circuit.
2. It can determine the output by simple graph traversal procedure.

iv. Programming model :

1. All the combinational elements in circuit can be represented in a programming model.
2. Any logic function is expressed in terms of basic logic operations namely :

i. NOT

ii. AND, NAND

iii. OR, NOR

iv. EXOR, EXNOR.

3. Assembly coding is an example of programming model.

Que 5.6. Write a short note on RTL model.

Answer

1. RTL is a type of Hardware Description Language (HDL).
2. It is used in describing the registers of a computer or digital electronic system, and the way in which data is transferred between them.
3. In this we make HDL models of registered circuits and how signals interact between them such as memories, flip-flops, latches and shift register.
4. RTL codes are fully synthesizable (can be realized in hardware), because they are written using basic HDL structures.
5. RTL models are characterized as functional, because they emphasize functional description while providing only summary about structural circuits.

PART-3

Structural Model and Level and Modeling.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.7. **Explain structural modeling.**

Answer

1. In this modeling, an entity is described as a set of interconnected components. A component instantiation statement is a concurrent statement. Therefore, the order of these statements is not important.
2. The structural style of modeling describes only an interconnection of components (viewed as black boxes), without implying any behavior of the components themselves nor of the entity that they collectively represent.
3. In structural modeling, architecture body is composed of two parts : the declarative part (before the keyword begin) and the statement part (after the keyword begin).

Que 5.8. **Differentiate between behavioral and structural modeling.**

Answer

No.	Behavioral model	Structural model
1.	A way of describing the function of a design as a set of concurrent algorithms.	A way of describing functions defined using basic components such as inverters, multiplexers, adders, decoders and basic logic gates.
2.	It is called black box modeling.	It is called glass box modeling.
3.	It focuses on showing the relationships between input and outputs.	It focuses on constructing the design using logic gates and predefined modules.

PART-4

Design for Testability.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.9. Write a short note on VLSI testing.

AKTU 2018-19, Marks 05

Answer

1. Following three factors conspire to create considerable difficulties for the test engineer and indeed, for the designer testing his own prototypes:

a. The complexity of VLSI systems.

b. The fact that the entire surface of the chip, other than over the pads, is sealed by an over glass layer and thus, circuit nodes cannot be probed for monitoring.

c. With minor exceptions, there is no way that the circuit can be modified during tests to make it work.

2. It is also essential for faults to be detected as early as possible in the manufacturing of a system. A relationship, known as the "rule of ten" tends to apply as far as test costs are concerned.

3. If the chip test cost = x, then once that chip is soldered into a PCB with other components, test cost = $10x$.

4. Further once that board is integrated into a system, then the test cost increases further by factor of ten to test cost = $100x$.

5. Finally a factor which is often overlooked is that test costs may increase further by factor of ten when the equipment/system is in service in the field.

6. Thus, chip design and fabrication mistakes can be very costly both in terms of time and money. So testability is very important.

Que 5.10. What is need of VLSI testing ? Discuss about functional and manufacturing tests.

AKTU 2017-18, Marks 10

Answer

A. Need of VLSI testing :

1. To ensure high yield and proper detection of faulty chips after manufacturing.

2. To achieve high quality manufacturing with acceptable cost.

B. Functional test :

1. The first set of tests verifies that the chip performs its intended function. These tests, called functionality tests or logic verification, are run before tape out to verify the functionality of the circuit.

2. Verification tests are usually the first ones a designer construct as part of the design process.
3. Verification tests were required to prove that a synthesized gate description was functionally equivalent to the source RTL.
4. Functional equivalence involves running a simulator on the two descriptions of the chip (*e.g.*, one at the gate level and one at a functional level) and ensuring that the outputs are equivalent at some convenient check points in time for all inputs applied.
5. One can check functional equivalence through simulation at various levels of the design hierarchy. If the description is at the RTL level, the behavior at a system level may be able to be fully verified.

C. Manufacturing tests :

1. The second set of tests verifies that every transistor, gate, and storage element in the chip functions correctly. These tests are conducted on each manufactured chip before shipping to the customer to verify that the silicon is completely intact. These are called manufacturing tests.
2. The functionality tests seek to confirm the function of a chip as a whole, whereas, manufacturing tests are used to verify that every gate operates as expected.
3. The need to do this arises from a number of manufacturing defects that might occur during either chip fabrication or accelerated life testing.
4. Tests are required to verify that each gate and register is operational and has not been compromised by a manufacturing defect.
5. Tests can be carried out at the wafer level to eliminate bad dies, or can be left until the parts are packaged. This decision would normally be determined by the yield and package cost.

PART-5

Ad-Hoc Design for Testability Techniques.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.11. Explain the Ad-Hoc testable design techniques.

OR

Explain the following :

i. Ad Hoc testable design techniques.

ii. Fault types and models. **AKTU 2017-18, Marks 10**

Answer

A. Ad-Hoc testable design techniques :

1. One way to increase the testability is to make nodes more accessible at some cost by physically inserting more access circuits to the original design.
2. Ad-Hoc testable design techniques are given as :

a. Partition–and–MUX technique :

i. Since the sequence of many serial gates, functional blocks, or large circuits are difficult to test, such circuits can be partitioned and multiplexers can be inserted such that some of the primary inputs can be fed to partitioned parts through multiplexers with accessible control signals.

ii. With this design technique, the number of accessible nodes can be increased and the number of test patterns can be reduced.

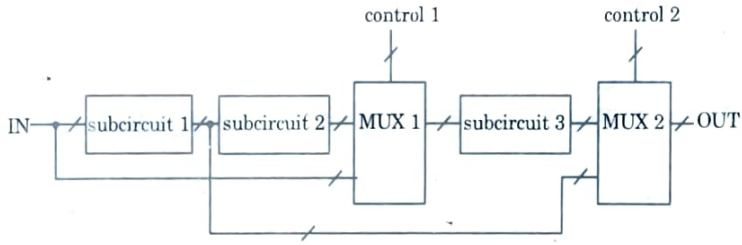


Fig. 5.11.1. Partition-and-MUX method for large circuits.

b. Initialize sequential circuit :

i. When the sequential circuit is powered up, its initial state can be a random, unknown state.

ii. In this case, it is not possible to start the test sequence correctly.

iii. The state of a sequential circuit can be brought to a known state through initialization. Initialization can be easily done by connecting asynchronous preset or clear-input signals from primary or controllable input to flip-flops or latches.

c. Disable internal oscillators or clocks :

i. To avoid synchronization problems during testing, internal oscillators and clocks should be disabled.

ii. For example, rather than connecting the circuit directly to the on-chip oscillator, the clock signal can be ORed with a disabling signal followed by an insertion of a testing signal as shown in Fig. 5.11.2.

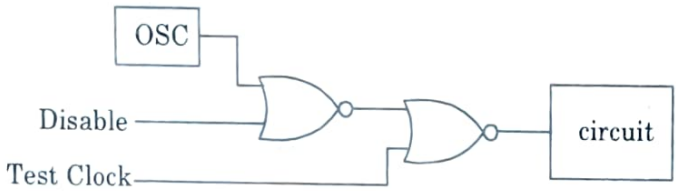


Fig. 5.11.2. Avoid synchronization problems via disabling of the oscillator.

d. **Avoid delay-depending logic :** Chains of inverters can be used to design in delay times and use AND operation of their outputs along with input to generate pulses.

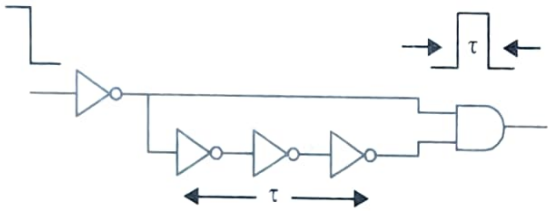


Fig. 5.11.3. A pulse-generation circuit using a delay chain of three inverters.

B. **Fault types and models :** Refer Q. 5.1, Page 5–2F, Unit-2.

PART-6

Controllability and Observability.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.12. **Explain the concept of observability, controllability and predictability.** **AKTU 2017-18, Marks 10**

OR

Explain the term controllability and observability in detail. **AKTU 2020-21, Marks 07**

Answer

A. Observability :

1. Observability is a degree to which it can be observed that node at output of an integrated circuit operates correctly.
2. 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.
3. Higher observability indicates less number of cycles is required to measure output node value.
4. Circuit having poor observability includes sequential circuit with long feedback of 100 ps.

B. Controllability :

1. Controllability of an internal circuit node within a chip is a measure of the ease of setting the node to logic 1 or 0 state.
2. This is important when assessing the degree of difficulty of testing a particular signal within a circuit.
3. An easily controllable node would be directly settable via an input pad. A node with little controllability may take hundreds of cycle to get it to the right state.
4. Masking of all flip-flops resettable via a global reset signal is one step toward good controllability.

C. Predictability :

1. Predictability is the ability to obtain known output values in response to given input stimuli.
2. This is the case with which the network being tested can be placed in a known state.
3. There are some factors *e.g.*, initial state of a circuit, race around condition etc. which may affect the predictability.

Que 5.13. Explain the scan based techniques.

AKTU 2018-19, Marks 05

OR

Explain the following :

i. Scan Based Technique.

ii. Fault types and models. **AKTU 2019-20, Marks 07**

OR

With the help of suitable diagram explain scan based technique.

AKTU 2020-21, Marks 07

Answer

A. Scan based technique :

1. The scan design techniques are structured approach to designing sequential circuits so that testability is 'designed in' from the outset.
2. The major difficulty in sequential circuit testing is in determining the internal state of the circuit.
3. Scan design technique are directed at improving the controllability and observability of the internal states.

The scan path :

1. The sequential circuit for scan path is shown in Fig. 5.13.1.
2. Scan path design techniques configure the logic so that the inputs and outputs of the combinational part can be accessed and the storage element reconfigured to form a shift register known as the scan path.

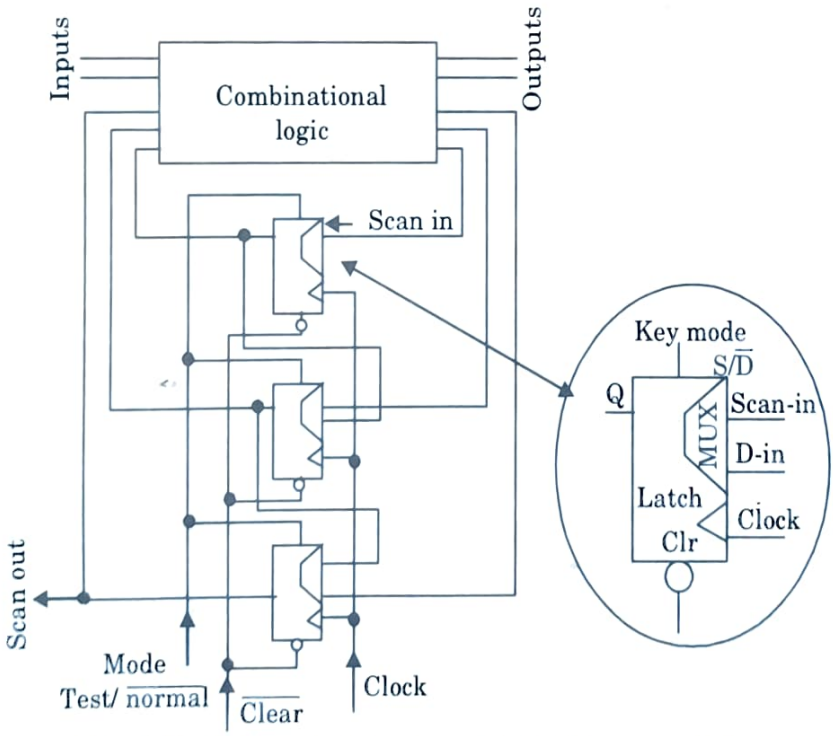


Fig. 5.13.1. Sequential circuit configured for scan path testing.

3. The storage elements are usually *D*, *JK*, or *RS* flip-flop elements with the classical structure being modified by the addition of a two way multiplexer on the data input.

4. In figure a basic *D* flip-flop has been shown with the added input multiplexer. This configuration is commonly known as 'MD' (multiplexed *D*) flip-flop.
5. The sequential circuit containing the scan path has two modes of operation a normal and a test mode.
6. The configuration associated with each basic mode is set out in Fig. 5.13.2(a) and Fig. 5.13.2(b) normal and test mode respectively.

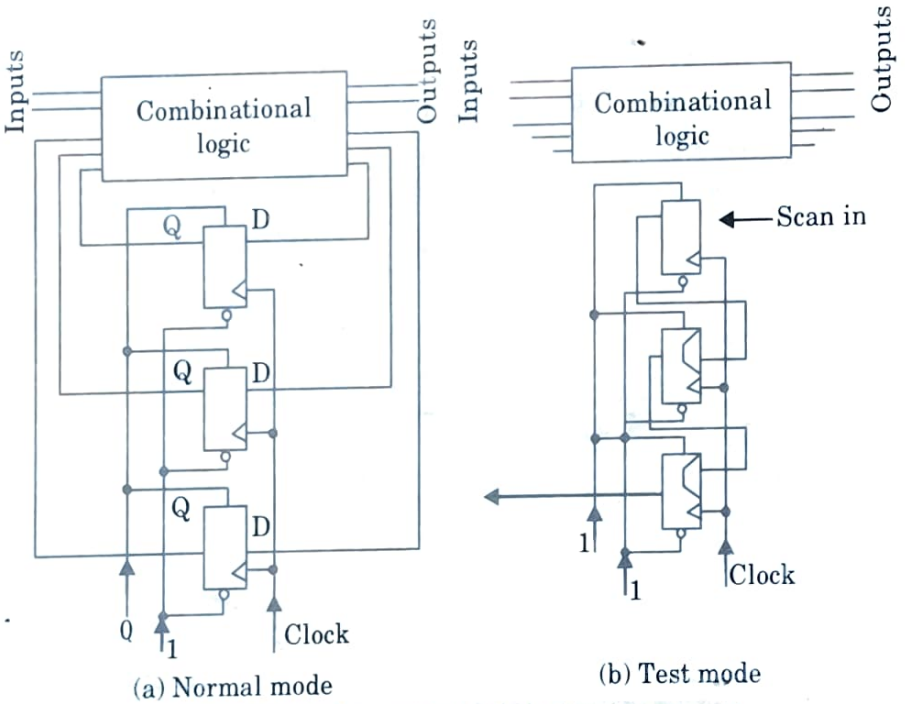


Fig. 5.13.2. Sequential circuit showing normal and test mode configurations.

7. Before applying test patterns, the scan path shift register is verified by shifting in all ones then all zeros.
8. A general method for testing with the scan path approach is as follows :

i. Set the mode to test so that the scan path is configured.

ii. Verify the scan path by shifting test data in and out.

iii. Set the shift register to a known initial state.

iv. Apply a test pattern to the primary inputs of the overall circuit.

v. Set the mode to normal. The circuit then settles and the primary outputs are monitored.

vi. Activate the circuit with one clock pulse.

vii. Return to the test mode.

viii. Scan out the contents of the scan path registers and simultaneously scan in next pattern.

ix. Repeat from step (iv) etc.

B. Fault types and models : Refer Q. 5.1, Page 5–2F, Unit-5.

Que 5.14. Explain the following :

i. Level-sensitive scan design

ii. Boundary scan test

OR

Write the advantages of level sensitive design and BST.

Answer

i. Level-Sensitive Scan Design (LSSD) :

1. This technique incorporates two aspects level sensitivity and a scan path approach. The general arrangement is indicated in Fig. 5.14.1.

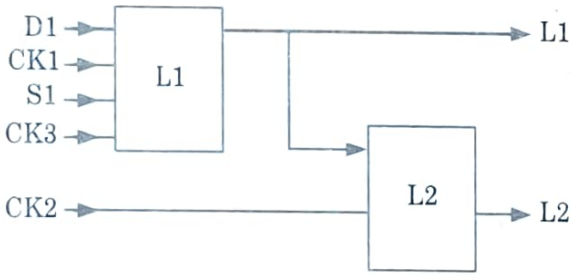


Fig. 5.14.1. Level sensitive scan design (LSSD) configuration.

2. The level sensitive means that the sequential network is designed so that when an input change occurs, the response is independent of the component and wiring delays within the network.

3. The scan path aspect is due to the use of shift register latches (SRL) employed as storage elements.

4. In the test mode they are connected as a long serial shift register.

5. Each SRL has a specific design similar to a master slave flip-flop. It is driven by two non-overlapping clocks which can be controlled readily from the primary inputs to the circuit.

6. Input $D1$ is the normal data input to the SRL, clocks CK1 and CK2 control the normal operation of the SRL while clocks CK3 and CK2 control scan path movement through the SRL.

7. The SRL output is derived at $L2$ in both modes of operation.

Advantages of LSSD approach :

a. The circuit operation is independent of dynamic characteristics of the logic elements rise and fall times and propagation delays.

b. ATP generation is simplified since tests need only be generated for a combinational circuit.

c. LSSD methods, when adopted in design, eliminate hazards and races.

ii. Boundary Scan Test (BST) :

1. This technique involving scan path and self testing to resolve the problems associated with the testing of boards carrying VLSI circuits.

2. The boundary scan path is provided with serial input and output pads and appropriate clock pads which make it possible to :

a. Test the interconnections between the various chips on the board.

b. Deliver test data to the chips on board for self testing.

c. Test the chip themselves with internal self test facilities.

The advantages of BST are :

a. No need for complex testers.

b. The test engineer's work is simplified.

c. Time spent on test pattern generation and application is reduced.

d. Fault coverage is increased.

PART-7

Introduction to Built-in-Self-Test(BIST) Concept.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.15. **What are the different scan based techniques. Explain built in self-test technique.** **AKTU 2017-18, Marks 10**

OR

Explain the issues involved in Built-In Self Test (BIST) techniques in detail. **AKTU 2018-19, Marks 10**

OR

Write a short note on Built-in-self test (BIST) techniques.

AKTU 2019-20, Marks 07

Answer

A. Scan based techniques : Refer Q. 5.13, Page 5–11F, Unit-5.

B. Built in self-test techniques :

1. As the complexity of individual VLSI circuits and as overall system complexity increases, test generation and application becomes an expensive and not always very effective means of testing.
2. Further, there are also very difficult problems associated with the high speed at which many VLSI systems are designed to operate.
3. Such problems required the use of very sophisticated, but not always affordable, test equipment's.
4. The objectives of this test are :

i. To reduce test pattern generation cost.

ii. To reduce the volume of test data.

iii. To reduce test time.

Compact Test :

1. Data compression techniques are currently used in BIST systems and consist of making comparisons on compacted system test responses instead of on the entire test data, which can be huge in some cases.
2. The test compacting scheme used most is called signature analysis. Here the data is represented as a polynomial $P(x)$ which is divided by a characteristic polynomial $C(x)$ to give the signature $R(x)$, so that

$$R(x) = P(x)/\ C(x)$$

This analysis is represented in Fig. 5.15.1.

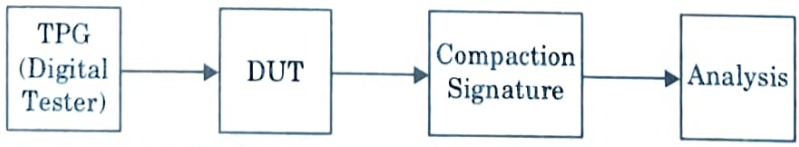


Fig. 5.15.1. Built-in-self-test-Signature analysis.

3. The signature from the DUT (Device Under Test) is compared with the expected signature to find out if the DUT is fault free.
4. Difference between the faulty signature and a good signature may also be used to indicate the nature of the fault.
5. Another technique of data compression transition counting has been in use for some considerable time.

6. It consists of counting transitions of a specified direction (0 to 1 or 1 to 0) and then comparing this count with the count obtained from the simulation model.

Que 5.16. Write a short note on the following :

i. Linear feedback shift register

ii. Built-in logic block observer

Answer

i. Linear Feedback Shift Register (LFSR) :

1. The LFSR model is that of a finite state machine comprising storage elements and modulo two adders connected in feedback loops as shown in Fig. 5.16.1.

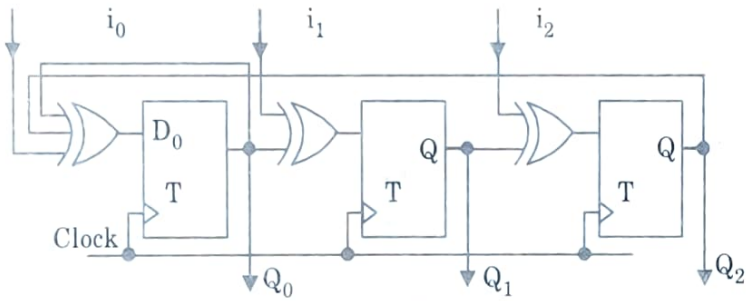


Fig. 5.16.1. Built-in-test-linear feedback shift register.

2. LFSR technique can be applied in number of ways, including random number generation, polynomial division for signature analysis and n-bit counting.

3. LFSR can be either series or parallel, the difference being in the operating speed and in the area of silicon occupied, parallel LFSR being faster but larger than serial LFSR.

ii. Built-In Logic Block Observer (BILBO) :

1. BILBO is built- in test generation scheme which uses signature analysis in conjunction with a scan path. It is aimed at integrated modular and bus-oriented systems, such as microprocessor and similar circuits.

2. Major component of a BILBO is an LFSR with a few gates. In Fig. 5.16.2 the BILBO is controlled by two signals, B_1 and B_2 which define the modes.

3. In the normal mode, $B_1 = B_2 = 1$ and the storage elements are used independently by the circuit as shown in Fig. 5.16.3.

4. In the test 1 mode, $B_1 = B_2 = 0$ and the storage elements are configured as a scan path, all storage elements being connected as a serial shift register as shown in Fig. 5.16.4.

5. Test vectors are then applied to the scan-in input and responses shifted out at the scan path output. The analysis of data is then similar to that for a simple scan-path test.
6. In the test 2 mode as shown in Fig. 5.16.5, $B_1 = 1, B_2 = 0$ and the circuit is then configured in a LFSR mode and be used either as a polynomial divider to compact data or as a random test pattern generator.
7. In the final mode $B_1 = 0, B_2 = 1$ which resets the BILBO.

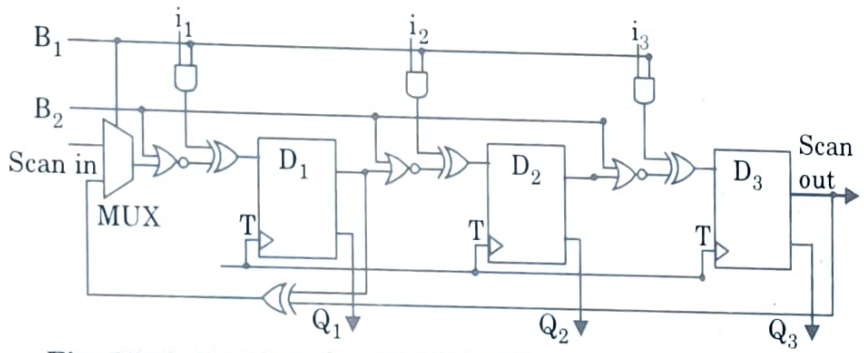


Fig. 5.16.2. Built-in-test-Built-in logic block observer (BILBO).

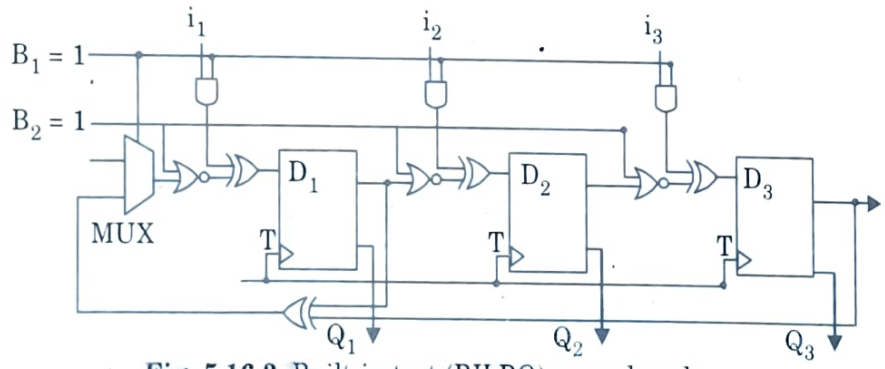


Fig. 5.16.3. Built-in-test-(BILBO) normal mode.

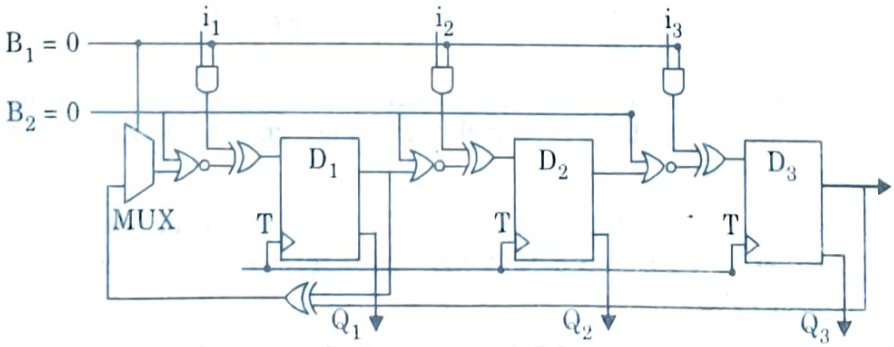


Fig. 5.16.4. Built-in-test-(BILBO) scan path mode.

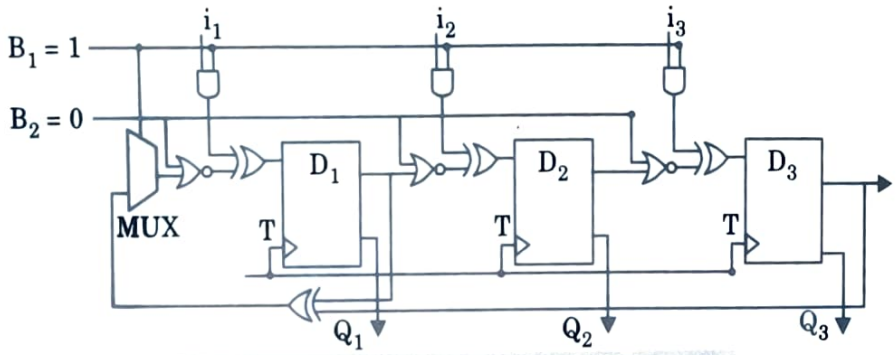


Fig. 5.16.5. Built-in-test-(BILBO) LFSR mode.

Que 5.17. Explain self checking techniques.

Answer

1. Data transmission in computer systems commonly makes use of coding to allow for the ready detection of errors. Such error detection techniques have been adapted and extended for built in test purposes.
2. Self checking techniques consist of supplying coded input data to the logic block under test and comparing the output in a checker designed to delete any errors. This is shown in Fig. 5.17.1.

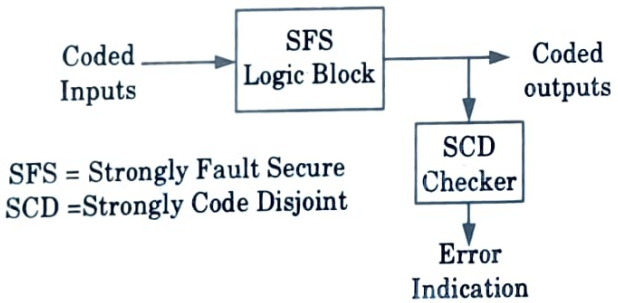


Fig. 5.17.1. Self-checking logic.

3. The design of logic blocks and checkers should then obey a set of rules in which the logic block is 'strongly fault secure' and the checker 'strongly code disjoint'.
4. The code used in data encoding depends on the types of errors that may occur at the logic block output. In general, three types are possible :

i. Simple error
ii. Unidirectional error
iii. Multiple error

5. For each type of error there is a set of appropriate coding techniques. Self checking techniques are applied to circuits in which security is important so that fault tolerance is of major interest.