

# CMOS VLSI DESIGN

<b>VLSI Methodologies</b>	<b>2</b>
<b>MOSFET</b>	<b>20</b>
<b>Unit process in VLSI and IC fabrication</b>	<b>57</b>
<b>CMOS Logic Circuits</b>	<b>88</b>
<b>Physical Design Automation</b>	<b>115</b>

## NOTE:

MAKAUT course structure and syllabus of 6<sup>th</sup> semester has been changed from 2021. CMOS VLSI DESIGN has been introduced as a new subject in present curriculum. The syllabus of this subject is almost same as Microelectronics & VLSI Designs [EC 702]. Taking special care of this matter we are providing the relevant MAKAUT university solutions of Microelectronics & VLSI Designs [EC 702] and some model questions & answers for newly introduced topics, so that students can get an idea about university questions patterns.

POPULAR PUBLICATIONS

## VLSI METHODOLOGIES

### Multiple Choice Type Questions

1. The data refresh operation is needed in  
a) SRAM      b) DRAM      c) EEPROM      d) FLASH  
Answer: (b) [WBUT 2014]
2. Which DRAM i/p buffer has lowest power consumption and fastest in operation?  
a) inverter type buffer      b) latch-type buffer      c) differential amplifier type buffer      d) both (b) and (c)  
Answer: (c) [WBUT 2014]
3. The disadvantages of flash memory is  
a) high electrical voltage required to erase data  
b) total block of memory is erased at a time  
c) very slow writing speed  
Answer: (b) [WBUT 2014]
4. Among the following which one has the greatest gate integration capacity?  
a) FPGA      b) CPLD      c) PLD      d) ASIC  
Answer: (a) [WBUT 2014, 2017]
5. FPGA is a  
a) full-custom ASIC  
b) semi-custom ASIC  
c) programmable ASIC  
d) none of these  
Answer: (c) [WBUT 2014, 2017]
6. In a PAL  
a) only AND array is programmable  
b) only OR array is programmable  
c) both (a) and (b)  
d) macro cell is the building block  
Answer: (a) [WBUT 2015, 2018]
7. Which one is not used in FPGA?  
a) static RAM technology  
b) dynamic RAM technology  
c) anti-fuse technology  
d) look up tables  
Answer: (b) [WBUT 2015]
8. Which design is more efficient?  
a) Pull up & Pull down design  
b) TG design  
c) Pre-charge & evaluate logic  
Answer: (c) [WBUT 2016]

9. Memory configuration of CPLD is

- a) volatile
- b) non-volatile
- c) both volatile and non-volatile

Answer: (b)

[WBUT 2017]

- b) non-volatile
- d) does not have memory

10. LUT is used in

- a) CPLD
- b) ASIC

Answer: (c)

[WBUT 2018]

- c) FPGA
- d) SPLD

11. The fastest logic family is

- a) TTL
- b) CMOS

Answer: (c)

[WBUT 2018]

- c) ECL
- d) DTL

12. DRAM is widely used because

- a) refreshing operating is not needed
- c) of low power consumption

Answer: (b)

[WBUT 2018]

- b) of low cost and high density
- d) of high speed

13. To store the configuration bits of CPLD

- a) external PROM is required
- c) FPGA is required

Answer: (b)

[WBUT 2018]

- b) no external PROM is required
- d) RAM is required

### **Short Answer Type Questions**

1. What is modularity and locality of VLSI design?

[WBUT 2016]

Answer:

#### **Modularity:**

Sub modules must have well defined functions and interfaces. Where modules are well formed, the interactions with other modules are easy to characterize. In the case of layout, the interface is defined by the ports of the sub modules which must be at specified locations and using specified conductors. If modules are “well-formed”, the interaction with other modules may be well characterized. In an IC this correspond a well defined behavioural, structural and physical interface that indicates the position, name, layer type, size, and signal type of external interconnections, along with the logic function and electrical characteristics. For instance, connection points may indicate the power and ground, inputs and outputs to a module. The function must also be defined in an unambiguous manner. Modularity helps the designer to clarify and document an approach to a problem, and also allows a design system to more easily check the attributes of a module as it is constructed. The ability to divide a task into a set of well defined modules also aids in a team of designers where each member has a portion of a complete chip to design.

The concept of modularity enables the parallelization of the design process. It also allows the use of generic modules in various designs. The well defined functionality and signal interface allow plug and play design. By defining well-characterized interfaces for each

## POPULAR PUBLICATIONS

module in the system, we effectively ensure that the internals of each module become unimportant to the exterior modules. Internal details remain at the local level. At a system level, the correct decision regarding modularity allows one to break-up a system with the confidence that when the parts are combined, the whole system will function as specified.

### **Locality**

By defining well characterized interfaces for a module, we are effectively stating that the other internals of the module are unimportant to any exterior interface. In this way we are performing a form of "information hiding" that reduces the apparent complexity of the module. In software, this is paralleled by the reduction of global variables to a minimum. A side effect of this complexity hiding is that a sub-module may be changed at any time without disturbing the overall design provided that the changed sub-module continues to support the same interface.

The concept of locality ensures that connections are mostly between neighbouring modules, avoiding long distance connections as much as possible to reduce the delay and complexity of routing. Time critical operations should be performed locally, without the need to access distant modules or signals.

For locality to work, we must impose restrictions on the use of a sub-module. In the case of layout, we must avoid making unwanted connections to elements in the sub-module and we must avoid design rule violations caused by proximity of external elements to internal elements.

To support these we come up with two general rules for good hierarchical design using locality:

- Connections to Sub-Cells

When making connections to any sub-module connections may only be made at defined ports

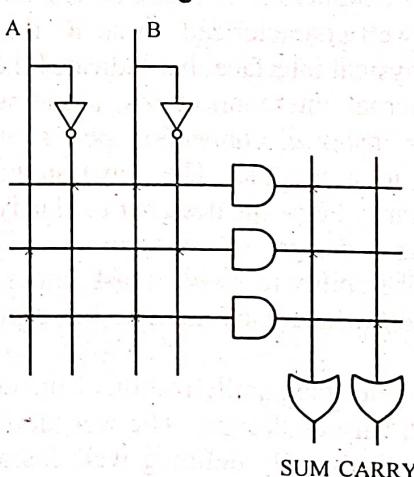
- Keep Out Areas

In addition, external wiring may not encroach on certain explicit or implicit keep out areas, over or around the cell.

## **2. Design a half subtractor circuit using PLA.**

[WBUT 2016]

**Answer:**



**CMOS-4**

### CMOS VLSI DESIGN

**3. Design a PROM which takes 3 binary bits as input and generates the output which is the square of the input.** [WBUT 2016]

**Answer:**

**Truth Table:**

Inputs			Outputs			
B <sub>2</sub>	B <sub>1</sub>	B <sub>0</sub>	E <sub>3</sub>	E <sub>2</sub>	E <sub>1</sub>	E <sub>0</sub>
0	0	0	0	0	1	1
0	0	1	0	1	0	0
0	1	0	0	1	0	1
0	1	1	0	1	1	0
1	0	0	0	1	1	1
1	0	1	1	0	0	0
1	1	0	1	0	0	1
1	1	1	1	0	1	0

$$E_0 = \Sigma m(0, 2, 4, 6)$$

$$E_1 = \Sigma m(0, 3, 4, 7)$$

$$E_2 = \Sigma m(1, 2, 3, 4)$$

$$E_3 = \Sigma m(5, 6, 7)$$

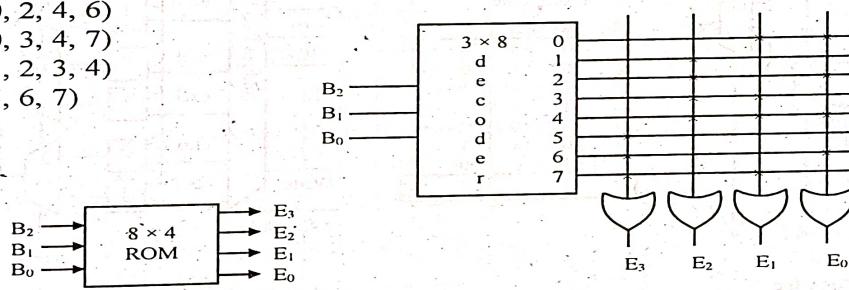


Fig: Combinational circuit using a PROM

[WBUT 2018]

**4. Write down the difference between CPLD and FPGA.**

**Answer:**

The primary differences between CPLDs and FPGAs are architectural. A CPLD has a somewhat restrictive structure consisting of one or more programmable sum-of-products logic arrays feeding a relatively small number of clocked registers. The result of this is less flexibility, with the advantage of more predictable timing delays and a higher logic-to-interconnect ratio. The FPGA architectures, on the other hand, are dominated by interconnect. This makes them far more flexible (in terms of the range of designs that are practical for implementation within them) but also far more complex to design for. Another notable difference between CPLDs and FPGAs is the presence in most FPGAs of higher-level embedded functions (such as adders and multipliers) and embedded memories.

Some FPGAs have the capability of partial re-configuration that lets one portion of the device be re-programmed while other portions continue running.

**CMOS-5**

## POPULAR PUBLICATIONS

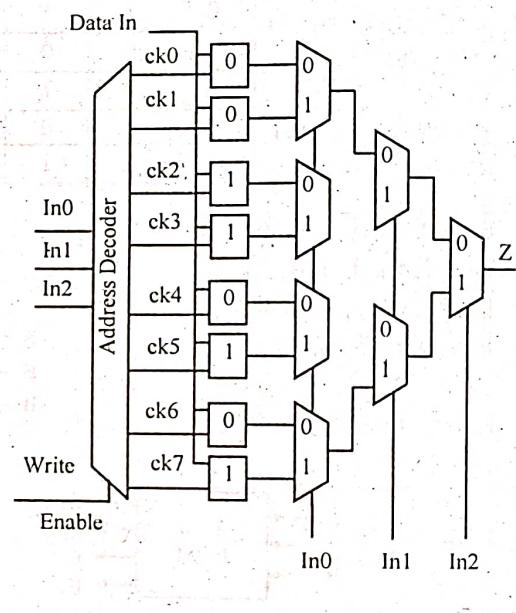
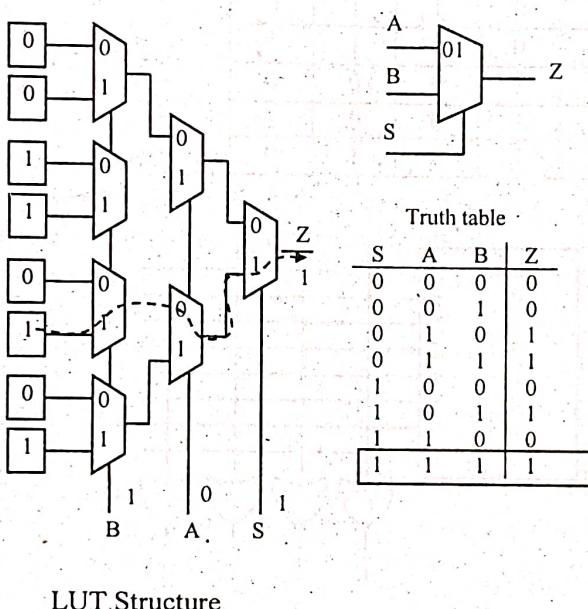
### 5. Explain LUT architecture of FPGA.

[WBUT 2018]

**Answer:**

The fundamental building blocks inside of an FPGA are the flip-flop and the lookup table (LUT). A LUT, which stands for **Look-Up Table**, in general terms is basically a table that determines what the output is for any given input(s). In the context of combinational logic, it is the **truth table**.

In other words, whatever behavior you get by interconnecting any number of gates (like AND, NOR, etc.), without feedback paths (to ensure it is state-less), can be implemented by a LUT.



### 6. How to download a program to FPGA?

[WBUT 2018]

**Answer:**

Step 1: Download and Install ISE WebPACK Design Software.

Step 2: Setup Xilinx Platform Cable. A **download** cable is needed to program the configuration file to the target **FPGA** device.

Step 3: Create the **FPGA** Project.

Step 4: **Program** the **FPGA** chip.

Step 5: Download the Program to SPI Flash.

## Long Answer Type Questions

1. a) Classify the different types of ASIC design.

[WBUT 2013, 2015]

b) Design the following circuit using PAL, PLA and PROM:

$$Y_1 = AB + A'C + ABC', Y_2 = AB'C, Y_3 = BC + ABC'.$$

**Answer:**

a) ASIC, short form of *Application Specific Integrated Circuit*, a chip designed for a particular application. ASICs are built by connecting existing circuit building blocks according to the need.

ASICs are broadly classified as:

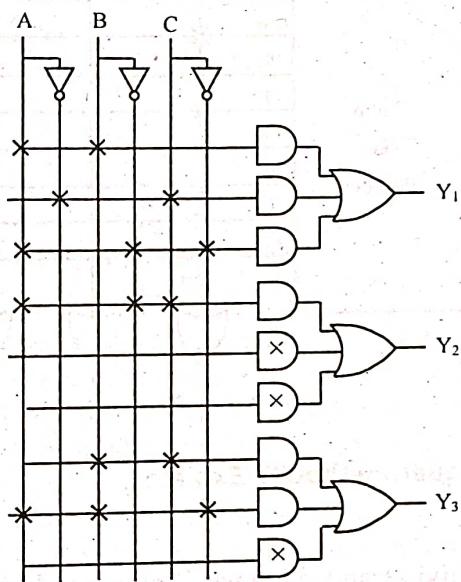
- Full custom ASIC and
- Semi custom ASIC

The semi custom ASICs are again classified as:

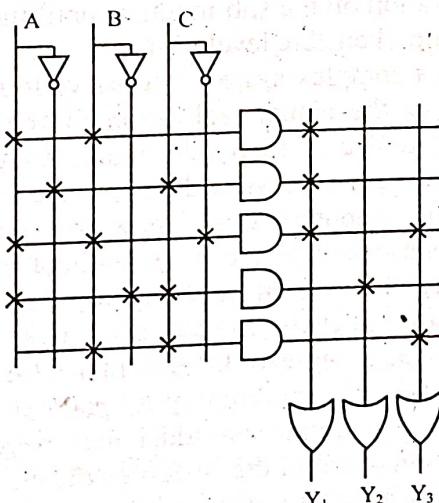
- Standard cell based ASIC
- Gate array based ASIC

b)

**PAL**



**PLA**

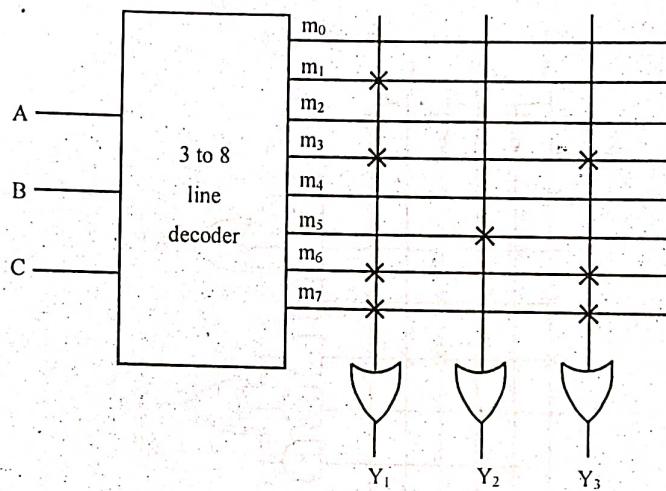


**CMOS-7**

## POPULAR PUBLICATIONS

$$\begin{aligned}
 Y_1 &= AB + A'C + ABC' \\
 &= ABC + ABC' + A'BC + A'B'C + ABC' \\
 &= m_7 + m_6 + m_3 + m_1 + m_6 \\
 Y_2 &= AB'C = m_5 \\
 Y_3 &= BC + ABC' \\
 &= ABC + A'BC + ABC' \\
 &= m_7 + m_3 + m_6
 \end{aligned}$$

## PROM



2. What is "divide and conquer method"? Explain.

[WBUT 2016]

Answer:

### Hierarchy

The use of hierarchy, or "divide and conquer", involves dividing a module into sub-modules and then repeating this operation on the sub-modules until the complexity of the sub-modules is at an appreciably comprehensible level of detail.

In the physical domain, partitioning a complex system into its various functional blocks will provide a valuable guidance for the actual realization of these blocks on chip. Obviously, the approximate shape and size of each sub-module should be estimated in order to provide a useful floorplan. This physical view describes the external geometry of the block, the locations of input and output pins, and how pin locations allow some signals to be transferred from one sub-block to the other without external routing. At lower levels of the physical hierarchy, the internal mask layout of each sub-block which defines the locations and the connections of each transistor and wires.

As described, a design may be expressed in terms of three domains (architecture, Register Transfer Level (RTL), logic & circuit). We can employ a "parallel hierarchy" in each domain to document the design. As an example, an adder may have a subroutine that models the behaviour, a gate connection diagram that specifies the circuit structure, and a piece of layout that specifies the physical nature of the adder. Composing the adder into

other structures can proceed in parallel for all three domains, with domain-to-domain comparison ensuring that the representations are consistent.

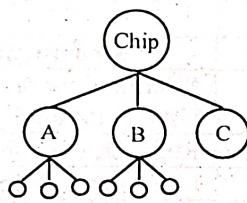
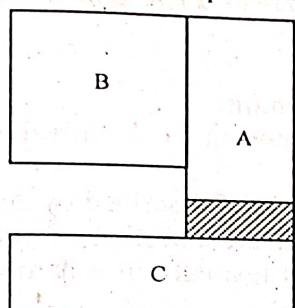


Fig. 1

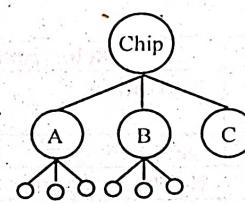
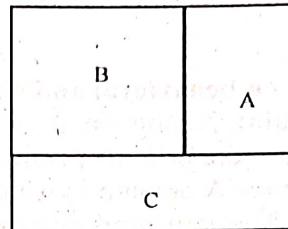


Fig. 2

Fig. Repartitioning the structural hierarchy to suit  
the physical hierarchy

At a system level, the use of hierarchy allows one to specify single designer projects, at which level the schedule is proportional to the number of available personnel. Significant benefits accrue where modules may be re-used within a system design. In this case we create a single design for this module but we use several instances of this design in different parts of the system.

The hierarchy is stopped at the level where modules are defined in terms of simulation models and physical layouts i.e., where the modules are standard cells. There may be two "disjoint" hierarchies describing the same structure. Generally it is a good practice to maintain identical hierarchies between the function, structure and physical aspects of a design because this allows consistent checks between description domains from the lowest level of the hierarchy to the very top levels. Frequently if the physical hierarchy is designed first without a structural or functional hierarchy, it will be found that the resulting hierarchy is cumbersome. On the other hand structural hierarchies may be defined that do not map well to the physical constraints. As an example, consider the floor plan shown in figure below where module A has to fit within a certain area constraint. Module B has space for some of the contents of module A but, due to the structural hierarchy the floor plan in fig. results.

Usually after a few iterations the physical and structural hierarchies may be reconciled. Many times the issue is moot because an automatic layout system is able to take the structural hierarchy and create a layout that meets both timing and area requirements.

**3. Describe Y-chart Model. Difference between Structural and Behavioral domain.  
Define PLA and PAL architectures.** [WBUT 2018]

## POPULAR PUBLICATIONS

**Answer:**

**1<sup>st</sup> Part:** Refer to Question No. 4(d) of Long Answer Type Questions.

**2<sup>nd</sup> Part:**

**Difference between behavioral and structural domain:**

**Structural domain:** A domain in which a component is described in terms of an interconnection of more primitive components.

**Behavioral domain:** A domain in which a component is described by defining its input/output response. The structural description is an interconnection of gate and flip-flop primitives. The behavioral description is expressed textually in a hardware description language (HDL).

Level of details	Behavioral Domain	Structural Domain
Chip	Algorithm	Microprocessor, RAM, ROM, UART, Parallel port
Register	Data flow	Register, Counter, MUX, ROM
Gate	Boolean equation	Logic gates, Flip-Flop
Circuit	Differential equations	Transistors, R, L, C
Layout	Equations of electron & hole motion	Geometric shape

**3<sup>rd</sup> Part:**

**PLA and PAL Architecture**

A Programmable Logic Array (PLA) consists of two levels of logic gates: a programmable “wired” AND-plane followed by a programmable “wired” OR-plane. A PLA is structured so that any of its inputs (or their complements) can be AND’ed together in the AND-plane; each AND-plane output can thus correspond to any product term of the inputs. Similarly, each OR plane output can be configured to produce the logical sum of any of the AND-plane outputs. With this structure, PLAs are well-suited for implementing logic functions in sum-of-products form. They are also quite versatile, since both the AND terms and OR terms can have many inputs.

Programmable Array Logic (PAL) is a relatively small FPD that has a programmable AND-plane followed by a fixed OR-plane. PALs feature only a single level of programmability, consisting of a programmable “wired” AND plane that feeds fixed OR-gates. To compensate for lack of generality incurred because the fixed OR output plane, AND Inputs & Flip-flop feedbacks are provided. Several variants of PALs are produced, with different numbers of inputs and outputs, and various sizes of OR-gates. PALs usually contain flip-flops connected to the OR-gate outputs so that sequential circuits can be realized.

**4. Write short notes on the following:**

- a) CPLD [WBUT 2013]
- b) FPGA [WBUT 2016, 2017, 2018]
- c) Programmable Logic Array [WBUT 2016]
- d) Y-chart of VLSI design flow [WBUT 2016]

**CMOS-10**

e) ASIC

[WBUT 2017]

f) PLD

[WBUT 2018]

g) VLSI Design

[WBUT 2018]

h) Standard Cell Based Design

[WBUT 2018]

**Answer:**

a) CPLD:

The term "complex programmable logic device" (CPLD) was introduced to differentiate these devices from their PAL and GAL predecessors, which were then sometimes referred to as "simple programmable logic devices" or SPLDs.

One type of large CPLD is often referred to as a "field-programmable gate array" or FPGA. A complex programmable logic device (CPLD) is a programmable logic device with complexity between that of PALs and FPGAs, and architectural features of both. The building block of a CPLD is the macro cell, which contains logic implementing disjunctive normal form expressions and more specialized logic operations.

The characteristic of non-volatility makes the CPLD the device of choice in modern digital designs to perform 'boot loader' functions before handing over control to other devices *not having this capability*. A good example is where a CPLD is used to load configuration data for an FPGA from non-volatile memory. CPLDs were an evolutionary step from even smaller devices that preceded them, PLAs and PALs.

The architectural advantage of CPLDs combining low cost, non-volatile configuration, and macro cells with predictable timing characteristics will likely be sufficient to maintain a product differentiation for the foreseeable future.

b) FPGA:

The cost of designing traditional standard cell ASICs is increasing every year. In addition to the non-recurring engineering (NRE) and mask costs, development costs are increasing due to design complexity. Issues such as power, signal integrity, clock tree synthesis, and manufacturing defects can add significant risk and time-to-market delays. FPGAs offer a viable and competitive option to traditional standard cell ASIC development by reducing the risk of re-spins, high NRE costs, and time-to-market delays.

Programmable logic has progressed from being used as glue logic to today's FPGAs, where complete system designs can be implemented on a single device. The number of gates and features has increased dramatically to compete with capabilities that have traditionally only been offered through ASIC devices. In today's FPGA, we have higher density devices, intellectual property (IP) integration, and high-speed I/O interconnects. All of these elements have allowed FPGAs to play a central role in digital technology. All of these elements have allowed FPGAs to play a central role in digital systems implementations.

i) **Goals and Techniques**

In addition to the logical functions to be performed, many other attributes must be satisfied to obtain a proper FPGA based system design. They are:

- **Performance:** The logic must run at a required speed. Performance can be measured in several ways, such as throughput and latency. Clock rate is often used as a measure of performance.

## POPULAR PUBLICATIONS

- **Power/energy:** The chip must often run with an energy or power budget. Every consumption is critical for battery powered systems. Even if the system is to run off the power grid, heat dissipation costs money and must be controlled.
- **Design time:** As a standard part, FPGAs have several advantages in design time. They can be used as prototypes, they can be programmed quickly, and they can be used as parts in the final design.
- **Design cost:** Design time is one important component of design cost, but other factors, such as support tools, must be taken into consideration. FPGA tools are often less expensive than custom VLSI tools.
- **Manufacturing cost:** it is the cost for replicating the system many times. FPGAs are generally more expensive than ASICs due to overhead cost for programming. However, bulk production reduces the cost.

### *ii) Design Challenges*

Design of an FPGA is particularly hard because several problems are to be solved. They are:

- **Multiple levels of abstraction:** FPGA design requires refining an idea through many levels of details. Starting from a specification, designer must create an architecture which performs the required functions, and then transform it into a logic design.
- **Multiple and Conflicting Costs:** Sometimes particular expensive software is needed to design a portion of it. Costs may also be in performance or power consumption of the final FPGA.
- **Short Design Time:** To hit the market first, and also at a competitive price is a real challenge.

### *iii) Requirements and Specifications*

A design project may start with varying amount of information. Some projects are revisions of earlier designs, some are implementation of proposed structures. In these cases, the function is well specified and some aspects of implementations are clear. But starting with some basic notions we use the term 'requirements' for a written description of what the system is to do, and we use the term 'specification' for a more formal description of the function. Attributes like performance, power consumption, and material cost are known as 'non-functional requirements'.

### *iv) Hierarchical Design*

Hierarchical design is a standard method for dealing with complex digital designs. It is commonly used in programming: each procedure breaks down the task into smaller operations until each step is refined into a procedure simple enough to be written directly. This technique is commonly known as 'divide-and-conquer'. Components are interconnected in physical design with a defined connectivity description, usually called a 'net list file'. Each component is used as a black box-to understand how the system works, we only have to know each component's input-output behaviour.

### *v) Design Abstraction*

Design abstraction is critical to hardware system design. Hardware designers use multiple levels of design abstraction to manage the design process and ensure that they meet major

design goals, such as speed and power consumption. The simplest design abstraction is the logic gates. Some design tasks, such as accurate delay calculation is almost impossible to calculate at this level. Other design tasks, such as logic optimization, are too cumbersome to be done on the circuit. So a design abstraction is chosen that is best suited to the design task.

We may use higher abstractions to make first cut decisions that are later refined using more detailed models.

### *FPGA Abstraction:*

- **Behaviour:** A detailed, executable description of what the chip should do, but not how it should do it. It allows us to describe in detail what needs to be done, error and boundary conditions, etc.
  - **Register Transfer**
  - **Logic**
  - **Configuration:** The logic must be placed into logic elements around the FPGA and the proper connections must be made between those logic elements. Placement and routing perform these important steps.

### c) Programmable Logic Array:

c) Programmable Logic Array:  
A programmable Logic Array (PLA) is a combinational module that provides a NOT-AND-OR or NOT-OR-AND implementation of switching modules. The modules are easily tri-stated and the enable input E is used to select the module for designing purpose. For  $E=1$ , a high or a low binary value can be obtained and for  $E=0$ , the outputs are in a high impedance state. The block diagrams are showing the PLA structures of SOP and POS forms.

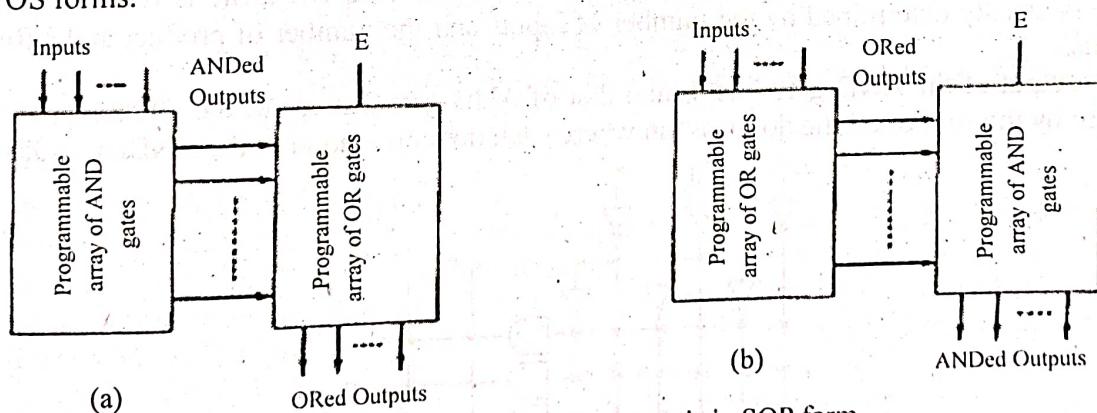


Fig: 1 a) PLA block where output is in SOP form,  
 b) PLA block where output is in POS form

b) PLA block where output is  $m_1$

The inputs applied to the input block are in both complemented and un-complemented form by using NOT gates at the inputs. If implementation is done using diodes or bipolar junction transistors in the bit position then AND-OR array is used and if MOS implementation is done then OR-AND array is used. The AND gates are arranged in a

## POPULAR PUBLICATIONS

regular structure known as AND array and the OR gates are also arranged in a regular structure as known as OR array.

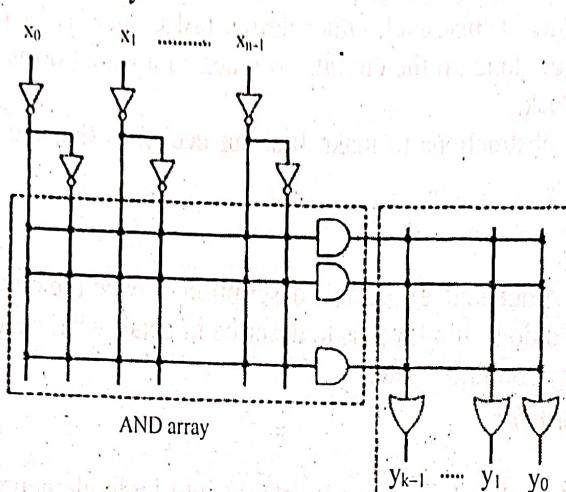


Fig 2 Regular arrangement of AND array and OR array

The cross points shown in the fig 2 are the points that are to be included in the product term in the AND array. Each cross point represents the input (in complemented or un-complemented form). If a certain input is used in the product term then a dot is shown. Same is the case with OR array. If the dot is placed then the resultant added terms are ORed. As an example, the following function can be represented as shown in fig 3.

$$Y_1 = x_0$$

$$Y_2 = x_0x_1 + x_0x_2 + x_1x_2$$

From the function it is clear that 5X6 AND matrix and 5X2 OR array is required. The size is usually determined by the number of inputs and the number of product and ORed terms.

The output of the AND gate 1 is  $x_0$  and that of AND gate 2 is  $x_1$ . In the OR array,  $y_1$  is given by the sum of all the dot position where each dot correspond to the ANDed results.

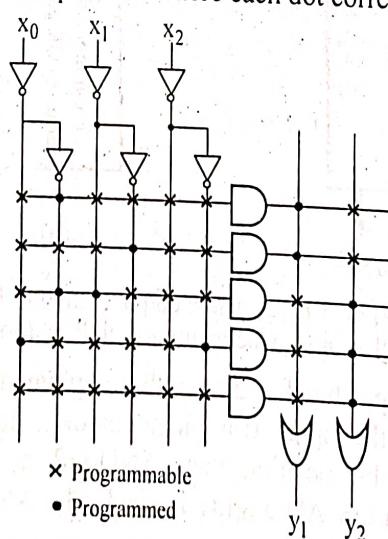


Fig: 3 AND and OR matrix to implement  $y_1$  and  $y_2$

When the device is implemented actually in PLA, the cross points are either diodes or transistors and it depends on the technology. In an NMOS implementation, the transistors are so connected in the AND and OR array that they act as NOT gates. So we have NOR-NOR implementation. The AND-OR array is equivalent to NOR-NOR array.

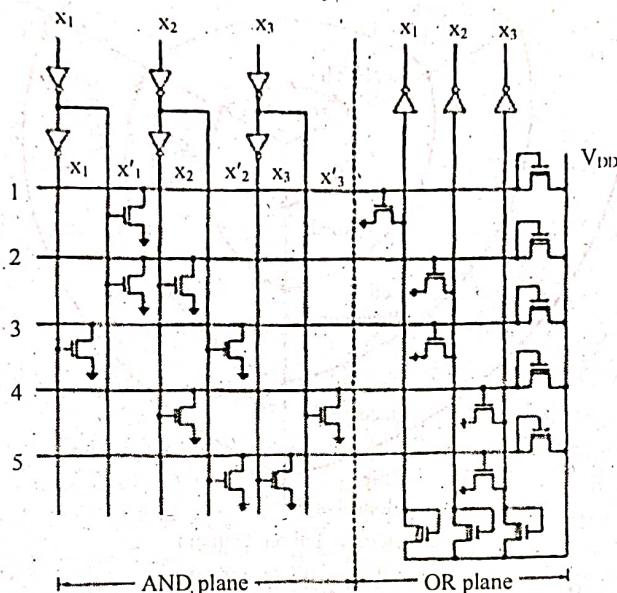


Fig 4 An NMOS NOR-BOR implementation of PLA

A NMOS NOR-NOR implementation of PLA is shown in fig 4. From the figure it can be noticed that PLA utilizes less space as compared to ROM. A more better way of utilization of empty space in PLA can be done by allowing two or more input lines to share the same column, or to let two product lines share the same row or by both. This sharing is known as folding. When a single column or a row is shared by more than two lines it is known as multiple folding and when both columns and rows are folded it is known as composite folding.

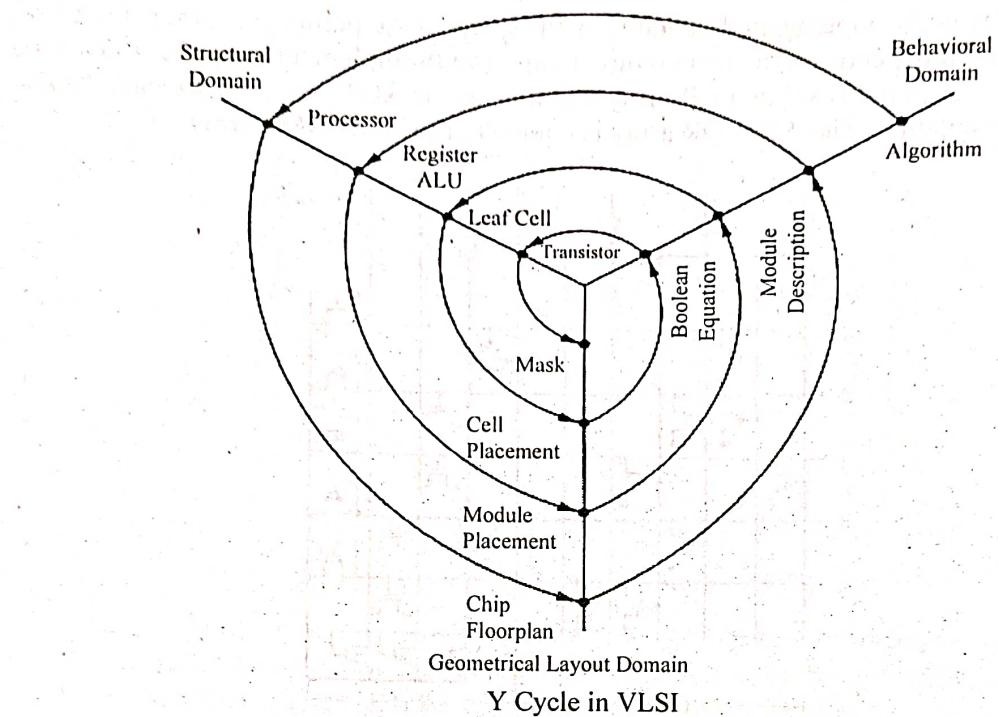
#### d) Y-chart of VLSI design flow:

The design process starts with a given set of requirements. Initial design is developed and tested against the requirements. When requirements are not met, the design has to be improved. If such improvement is either not possible or too costly, then the revision of requirements and its impact analysis must be considered. The Y-chart (first introduced by D. Gajski) shown in Fig. illustrates a design flow for most logic chips, using design activities on three different axes (domains) which resemble the letter Y.

The Y-chart consists of three major domains, namely:

- behavioral domain,
- structural domain,
- geometrical layout domain.

## POPULAR PUBLICATIONS



The design flow starts from the algorithm that describes the behavior of the target chip. The corresponding architecture of the processor is first defined. It is mapped onto the chip surface by floor-planning. The next design evolution in the behavioral domain defines finite state machines (FSMs) which are structurally implemented with functional modules such as registers and arithmetic logic units (ALUs). These modules are then geometrically placed onto the chip surface using CAD tools for automatic module placement followed by routing, with a goal of minimizing the interconnect areas and signal delays. The third evolution starts with a behavioral module description. Individual modules are then implemented with leaf cells. At this stage the chip is described in terms of logic gates (leaf cells), which can be placed and interconnected by using a cell placement & routing program. The last evolution involves a detailed Boolean description of leaf cells followed by a transistor level implementation of leaf cells and mask generation.

### e) Application Specific Integrated Circuits (ASIC):

ASIC, short form of *Application Specific Integrated Circuit*, a chip designed for a particular application. ASICs are built by connecting existing circuit building blocks according to the need. Since the building blocks already exist in a library, it is much easier to produce a new ASIC than to design a new chip from scratch.

*Examples of ICs that are ASICs include:* a chip for a toy bear that talks; a chip for a satellite; a chip designed to handle the interface between memory and a microprocessor for a workstation CPU; and a chip containing a microprocessor as a cell together with other logic. *Examples of ICs that are not ASICs include* standard parts such as: memory

chips sold as a commodity item—ROMs, DRAM, and SRAM; microprocessors; TTL or TTL-equivalent ICs at SSI, MSI, and LSI levels.

As a general rule, if it is mentioned in a data book, then it is probably not an ASIC, but there are some exceptions. For example, two ICs that might or might not be considered ASICs are a controller chip for a PC and a chip for a modem. Both of these examples are specific to an application (shades of an ASIC) but are sold to many different system vendors (shades of a standard part).

ASICs are broadly classified as:

- Full custom ASIC and
- Semi custom ASIC

In a **full-custom ASIC** an engineer designs some or all of the logic cells, circuits, or layout specifically for one ASIC. This means the designer abandons the approach of using pretested and pre-characterized cells for all or part of that design. It makes sense to take this approach only if there are no suitable existing cell libraries available that can be used for the entire design. This might be because existing cell libraries are not fast enough, or the logic cells are not small enough or consume too much power. One may need to use full-custom design if the ASIC technology is new or so specialized that there are no existing cell libraries or because the ASIC is so specialized that some circuits must be custom designed. Full-custom ICs are the most expensive to manufacture and to design. The **manufacturing lead time** (the time it takes just to make an IC—not including design time) is typically eight weeks for a full-custom IC.

The members of the IC family that we are more interested in are **semicustom ASICs**, for which all of the logic cells are predesigned and some (possibly all) of the mask layers are customized. Using predesigned cells from a **cell library** makes our lives as designers much, much easier. There are two types of semicustom ASICs: standard-cell-based ASICs and gate-array-based ASICs.

There are many reasons for choosing an ASIC-based solution over discrete components. Some examples:

- An ASIC solution normally provides better total functionality and performance than the corresponding discrete solution. The reason is that the design can more easily be optimized in an ASIC. This is especially true for systems requiring low power consumption.
- In most cases, using an ASIC solution means that the weight and volume of the product can be reduced considerably. For portable products, that is an important and sometimes vital advantage.
- ASIC solutions offer a remarkable improvement of overall product economy, provided the product is manufactured in a certain minimum volume.
- If you choose ASIC technology, your product will automatically be very well protected against copying.

## POPULAR PUBLICATIONS

### f) Programmable Logic Devices (PLD):

Programmable Logic Devices may be configured or programmed to create a part customized to a specific application, and so they also belong to the family of ASICs. PLDs are different technologies to allow programming of the device. Common features of PLDs are:

- No customized mask layers or logic cells
- Fast design turnaround
- A single large block of programmable interconnect
- A matrix of logic macro cells consisting programmable array logic followed by flip-flops

The design process using PLDs require the following steps:

- Function specification
- Generation of Boolean equations
- Minimization of Boolean equations
- Generation of fuse maps
- Logic simulation
- Programming the select device
- Testing

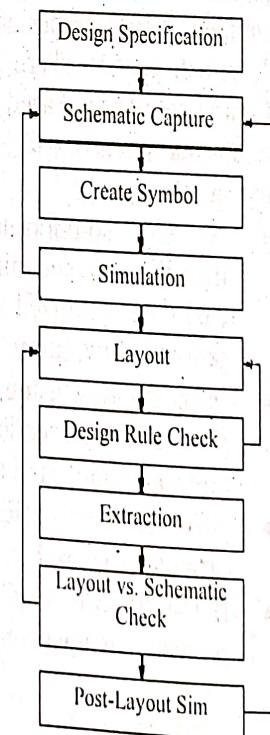
The design strategies for PLD are:

- Rom-based design
- Programmable Logic Array (PLA) based design
- Programmable Array Logic (PAL) based design

### g) VLSI Design:

The VLSI IC circuits design flow is shown in the figure below. The various levels of design are numbered and the blocks show processes in the design flow.

Specifications comes first, they describe abstractly, the functionality, interface, and the architecture of the digital IC circuit to be designed.



Behavioral description is then created to analyze the design in terms of functionality, performance, compliance to given standards, and other specifications.

RTL description is done using HDLs. This RTL description is simulated to test functionality. From here onwards we need the help of EDA tools.

RTL description is then converted to a gate-level netlist using logic synthesis tools. A gate-level netlist is a description of the circuit in terms of gates and connections between them, which are made in such a way that they meet the timing, power and area specifications.

Finally, a physical layout is made, which will be verified and then sent to fabrication.

#### **h) Standard Cell Based Design:**

A cell-based ASIC uses pre-designed logic cells (AND gates, OR gates, multiplexers, and flip-flops, for example) known as **standard cells**.

The standard-cell areas (also called flexible blocks) in these types of ASICs are built of rows of standard cells — like a wall built of bricks. The standard-cell areas may be used in combination with larger predesigned cells, perhaps microcontrollers or even microprocessors, known as **mega cells**. Mega cells are also called mega functions, full-custom blocks, system-level macros (SLMs), fixed blocks, cores, or Functional Standard Blocks (FSBs).

The standard cells can be placed anywhere on the silicon; this means that all the mask layers of a CBIC are customized and are unique to a particular customer. The advantage of CBICs is that designers save time, money, and reduce risk by using a predesigned, pretested, and pre-characterized **standard-cell library**. In addition each standard cell can be optimized individually. The disadvantages are the time or expense of designing or buying the standard-cell library and the time needed to fabricate all layers of the ASIC for each new design.

The important features of this type of ASIC are as follows:

- All mask layers are customized — transistors and interconnect.
- Custom blocks can be embedded.
- Manufacturing lead time is about eight weeks

**POPULAR PUBLICATIONS**

## **MOSFET**

### **Multiple Choice Type Questions**

1. A MOS device can be used as a resistor in [WBUT 2014]

- a) linear region
- b) saturation region
- c) sub-threshold condition
- d) none of these

Answer: (a)

2. The equivalent resistor in a switched capacitor circuit is [WBUT 2015]

- a) directly proportional to square of the time period of clock
- b) inversely proportional to square of the time period of clock
- c) inversely proportional to the time period of clock
- d) directly proportional to the time period of clock

Answer: (d)

3. In a saturated MOSFET, with increasing V<sub>DS</sub>, channel length [WBUT 2015]

- a) increases
- b) decreases
- c) remain same
- d) may increase or decrease

Answer: (b)

4. A constant current source gives a current of 200mA for a load resistance of 500Ω. When short circuited, the current is [WBUT 2015]

- a) 60mA
- b) 80mA
- c) 150mA
- d) 200mA

Answer: (d)

5. In which device at zero gate voltage the channel already exist? [WBUT 2016]

- a) Depletion type MOSFET
- b) CMOS device
- c) Enhance type MOSFET

Answer: (a)

6. The condition, where the majority carrier concentration is greater near the Si-SiO<sub>2</sub> interface compared to the bulk in the MOSFET is called [WBUT 2016]

- a) Accumulation
- b) Depletion
- c) Inversion

Answer: (a)

7. The potential at which the inversion layer dominates the substrate behavior is

[WBUT 2016]

- a) Pinch-off voltage
- b) Cut-off voltage
- c) Threshold voltage

Answer: (c)

8. The overlap capacitance is

- a) voltage dependent      b) voltage independent

- c) none of these

Answer: (b)

[WBUT 2016]

9. Overlapping capacitance of MOS denotes

- a) capacitance between drain and oxide layer  
b) capacitance between source and oxide layer  
c) both (a) and (b)

Answer: (c)

[WBUT 2016]

10. Which is correct for the accumulation region?

- a)  $V_{GB} < V_{FB}$       b)  $V_{GB} > V_{FB}$       c)  $V_{GB} = V_{FB}$

Answer: (a)

[WBUT 2016]

11. Pinch off region of MOS transistor the current become

[WBUT 2017]

- a) saturated      b) non-saturated  
c) decrease      d) increase exponentially

Answer: (a)

### Short Answer Type Questions

1. What is the problem of realizing a large value resistor by a MOSFET structure?

How can a switched capacitor be used to overcome this problem? [WBUT 2013]

Answer:

Problems for realizing a large value of resistor using MOS structure are:

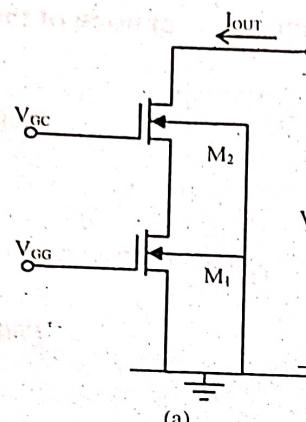
- i) space requirement is large, ii) accuracy is not good.

Switched capacitor circuits are used to replace the conventional MOS resistances where the value of resistance is dependent on clock frequency and the value of a MOS capacitance. So dependence on physical parameters are eliminated, which increase accuracy and the reduction of size is also possible.

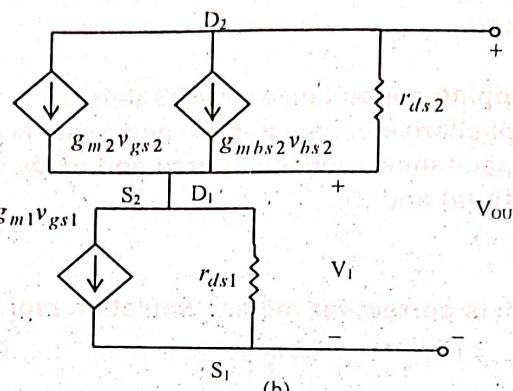
2. How can resistance of a current source/sink be improved? [WBUT 2013]

**POPULAR PUBLICATIONS**

**Answer:**



(a)



(b)

a) Circuit for increasing  $r_{out}$  of a current sink

b) Small signal model of circuit (a)

The common gate configuration is used to multiply the source resistance by the approximate voltage gain of the common gate configuration.

Here the output resistance  $r_{ds1}$  of the current sink is increased by the common gate voltage gain of  $M_2$ . Therefore small signal output resistance is increased by the factor  $g_m 2 r_{ds2}$ .

3. What is current source and current sink is VLSI circuit? Design a current sink using  $V_{DD} = -V_{SS} = 2.5V$  to sink a current of  $10\mu A$ . Estimate the minimum voltage across the current source and the output resistance. Assume  $K_p = 50\mu A/V^2$ ,  $L = 5\mu m$ ,  $V_{THN} = 0.83V$ ,  $\lambda = 0.06$ . [WBUT 2014]

**Answer:**

Minimum voltage across the current source can be calculated as:

$$V_{min} = \left[ 2I_D / K'(W/L) \right]^{1/2}$$

and resistance =  $1/\lambda I_D$

$$V_{out, min} = 2\sqrt{\frac{2I_{ref}}{\beta}} + V_{th,n}, \text{ where } \beta = \frac{K'W}{L}$$

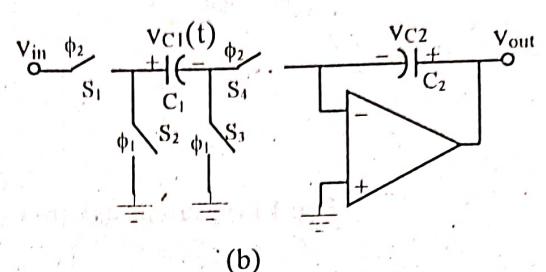
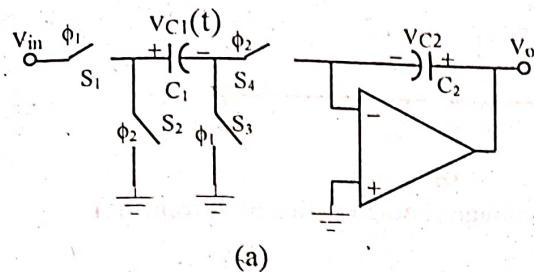
$$r_{out} \approx \frac{1}{\lambda I_d}$$

Here,  $\lambda = 0.06$ ,  $L = 5\mu m$ ,  $V_{th,n} = 0.83V$ ,  $K_p = K' = 50\mu A/V^2$ ,  $W = ?$

$$I_d = 10\mu A = I_{ref}$$

4. Draw and explain the operation of MOS Switched Capacitor Integrator and also find the expression for output voltage. [WBUT 2014, 2017].

Answer:



Integrator Circuit using switched capacitor (a) Non-inverting and (b) Inverting.

If we replace the resistor with the negative trans-resistance equivalent circuit, we obtain non-inverting switched capacitor integrator circuit. If we replace the resistor with positive trans-resistance, we obtain inverting switched capacitor integrator circuit.

Beginning with the phase  $\phi_1$ , during the time from  $t=(n-1)T$  to  $t=(n-1/2)T$ , the voltage across capacitor is

$$V_{c1}^o(n-1)T = v_{in}^o(n-1)T$$

$$\text{and } V_{c2}^o(n-1)T = v_{out}^o(n-1)T$$

During the  $\phi_2$  phase, we can write

$$V_{out}^e \left(1 - \frac{1}{2}\right)T = \frac{C_1}{C_2} v_{in}^o(n-1)T + v_{out}^o(n-1)T$$

If we advance one more phase period, the voltage output remain unchanged.

$$\text{Thus, } v_{out}^o(n)T = \frac{C_1}{C_2} v_{in}^o(n-1)T + v_{out}^o(n-1)T$$

[WBUT 2015]

5. Describe the current sink and current source.

Answer:

- Current sinks and sources are two terminal components whose current at any instant of time is independent of the voltage across their terminals.
- Through sink and source, current flows from positive node to negative node.
- Current sink typically has a negative node at  $V_{SS}$  and a current source has a positive node at  $V_{DD}$ .
- The gate potential of the MOS should be taken to such level to create desired value of current. The non-saturation region of MOS device is not a good current source.

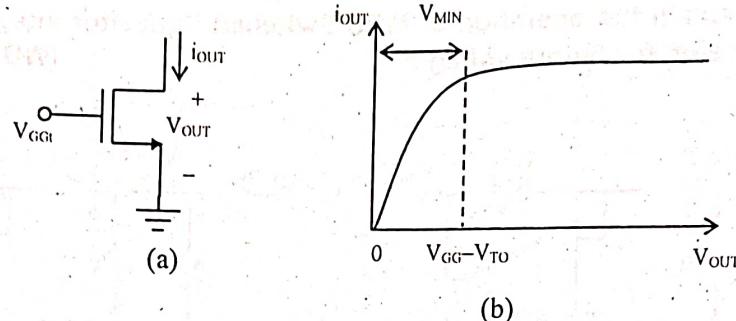


Fig: 1 (a) Current sink (b) Current – voltage characteristics of current sink

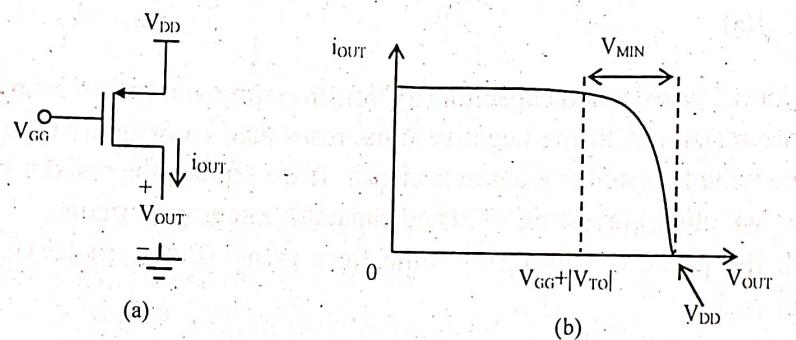


Fig: 2 (a) current source (b) current-voltage characteristics of current source

To perform current sink properly –

$$V_{OUT} \geq V_{GG} - V_{TO} \quad \dots \text{(i)}$$

If both source and bulk are connected to ground, then small signal output is given by:

$$r_{out} = \frac{1 + \lambda V_{DS}}{\lambda I_D} \approx \frac{1}{\lambda I_D} \quad \dots \text{(ii)}$$

If the source and bulk are not connected to the same potential, the characteristics will not change as long as  $V_{BS}$  is constant.

Fig. 2(a) shows an implementation of a current source using a p-channel transistor. Here also the gate is taken to a constant potential as in the sink. The small signal output resistance of the source is also given by equation (ii). The drain - source voltage must be larger than  $V_{MIN}$  for this current source to work properly. This current source only works for values of  $V_{OUT}$  given by:

$$V_{OUT} \leq V_{GS} + |V_{TO}| \quad \dots \text{(iii)}$$

The current sink and source of Fig 1 and 2 are very simple circuit. Their performance may be improved further by:

- a) Increasing the small signal output resistance, resulting in more constant current over the range of  $V_{OUT}$  values and b) Reducing in  $V_{MIN}$  value, allowing a larger range of  $V_{OUT}$  over which current sink/source works properly.

**6. Describe the operation of MOSFET differential Amplifier.**  
**Answer:**

[WBUT 2015]

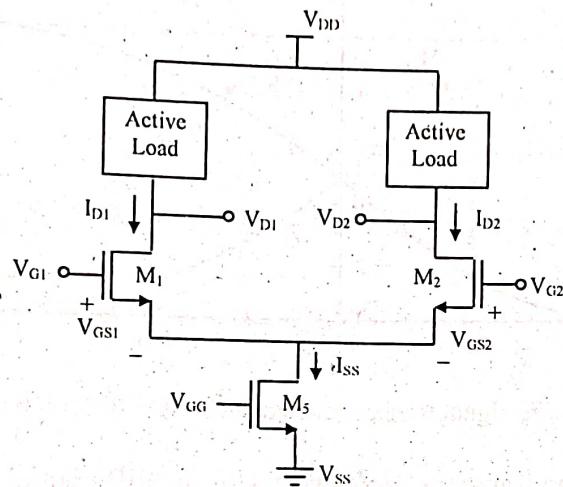


Fig: (a) A general MOS differential amplifier configuration

The figure (a) shows the circuit of a general MOS differential amplifier. The active loads consist of MOS transistors at different configurations. The key aspect of the differential amplifier is the input source coupled pair, M<sub>1</sub> and M<sub>2</sub>.

For large signal considerations, let us assume that M<sub>1</sub> and M<sub>2</sub> are in saturation and their threshold voltages are equal. Neglecting channel modulation,

$$V_{ID} = V_{G1} - V_{G2} = V_{GS1} - V_{GS2} = \left| \frac{2I_{D1}}{\beta_1} \right|^{\frac{1}{2}} - \left| \frac{2I_{D2}}{\beta_2} \right|^{\frac{1}{2}} \quad \dots (i)$$

and  $I_{SS} = I_{D1} + I_{D2}$  .... (ii)

where,  $\beta = K'(W/L)$ . Assuming  $\beta_1 = \beta_2 = \beta$ , from equation (i) and (ii), we get

$$I_{D1} = \frac{I_{SS}}{2} + \frac{I_{SS}}{2} \left[ \frac{\beta V_{ID}^2}{I_{SS}} - \frac{\beta^2 V_{ID}^4}{4I_{SS}^2} \right]^{\frac{1}{2}} \quad \dots (iii)$$

$$\text{and } I_{D2} = \frac{I_{SS}}{2} - \frac{I_{SS}}{2} \left[ \frac{\beta V_{ID}^2}{I_{SS}} - \frac{\beta^2 V_{ID}^4}{4I_{SS}^2} \right]^{\frac{1}{2}} \quad \dots (iv)$$

These relationships are valid only for

$$\left| V_{ID} \right| \leq \left| \frac{2I_{SS}}{\beta} \right|^{\frac{1}{2}} \quad \dots (v)$$

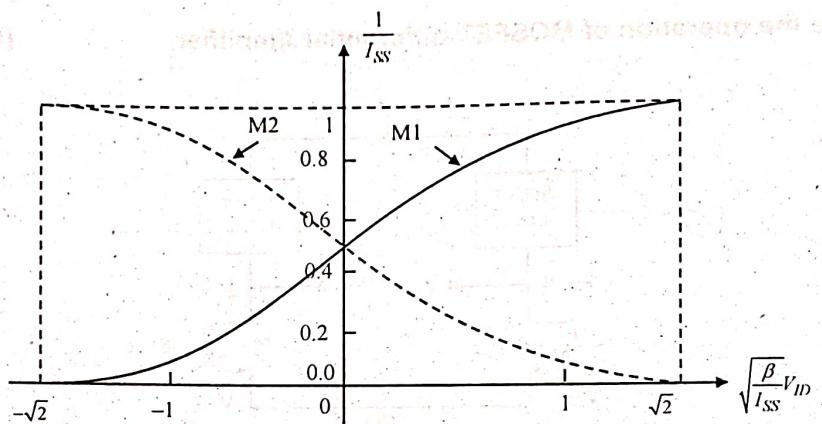


Fig: (b) Large signal transconductance of MOS differential amplifier

The large signal voltage transfer characteristics of the differential amplifier can be found by using the equations (iii) and (iv) along with voltage current characteristics of the active load devices.

### 7. Describe the operation of a current Mirror circuit using MOSFET. [WBUT 2015]

**Answer:**

The current mirrors are just an extension of the current source/sink. The current mirror uses the principle that if the gate source potentials of two identical MOS transistors are equal, the channel currents should be equal.

Fig. shows the implementation of a simple n-channel current mirror. Here the current  $I_i$  is defined by a current source or some other means and  $I_o$  is the output mirrored current.

Here,  $M_1$  is always in saturation as  $V_{DS1}$  is equal to  $V_{GS1}$  and assuming that  $V_{DS2} \geq V_{GS2} - V_{T2}$ , we can use the equations in saturation region of the transistor to get the ratio of  $I_o$  to  $I_i$  as:

$$\frac{I_o}{I_i} = \left( \frac{L_1 W_2}{W_1 L_2} \right) \left( \frac{V_{GS} - V_{T2}}{V_{GS} - V_{T1}} \right)^2 \left[ \frac{1 + \lambda V_{DS2}}{1 + \lambda V_{DS1}} \right] \left( \frac{K'_2}{K'_1} \right) \quad \dots (i)$$

As in normal cases, the components of a current mirror are processed on the same integrated circuit, all of the physical parameters such as  $V_T$  and  $K'$  are identical for both devices. As a result the above equation reduces to

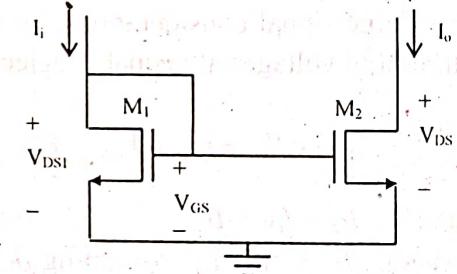


Fig: n-channel current mirror

$$\frac{I_o}{I_i} = \left( \frac{L_1 W_2}{W_1 L_2} \right) \left( \frac{1 + \lambda V_{DS2}}{1 + \lambda V_{DS1}} \right) \quad \dots \text{(ii)}$$

If  $V_{DS2}$  is maintained as equal to  $V_{DS1}$ , then the ratio  $I_o/I_i$  becomes,

$$\frac{I_o}{I_i} = \frac{L_1 W_2}{W_1 L_2} \quad \dots \text{(iii)}$$

Consequently, the ratio  $I_o/I_i$  becomes a function of the aspect ratios that are under control of the designer.

There are **three effects** that cause the **current mirror** to be **drifted** from its ideal situation considered in equation (i).

They are

- i) Channel length modulation effect
- ii) Threshold offset between the two transistors and
- iii) Imperfect geometrical matching.

#### 8. Explain the following phenomena in a MOS structure:

[WBUT 2016]

a) Channel length modulation

b) Scaling of MOSFET

**Answer:**

a) **Channel Length Modulation:**

One of several short channel effects in MOSFET scaling, **channel length modulation (CLM)** is a shortening of the length of the inverted channel region with increase in drain bias for large drain biases. The result of CLM is an increase in current with drain bias and a reduction of output resistance.

To understand the effect, first the notion of **pinch-off** of the channel is introduced. The channel is formed by attraction of carriers to the gate, and the current drawn through the channel is nearly a constant independent of drain voltage in active mode. However, near the drain, the gate *and drain jointly* determine the electric field pattern. Instead of flowing in a channel, beyond the pinch-off point the carriers flow in a subsurface pattern made possible because the drain and the gate both control the current. In the figure, the channel is indicated by a dashed line and becomes weaker as the drain is approached, leaving a gap of un-inverted silicon between the end of the formed inversion layer and the drain (the *pinch-off region*).

As the drain voltage increases, its control over the current extends further toward the source, so the un-inverted region expands toward the source, shortening the length of the channel region, the effect called *channel-length modulation*. Because resistance is proportional to length, shortening the channel decreases its resistance, causing an increase in current with increase in drain bias for a MOSFET operating in saturation. The effect is more pronounced the shorter the source-to-drain separation, the deeper the drain junction, and the thicker the oxide insulator.

## POPULAR PUBLICATIONS

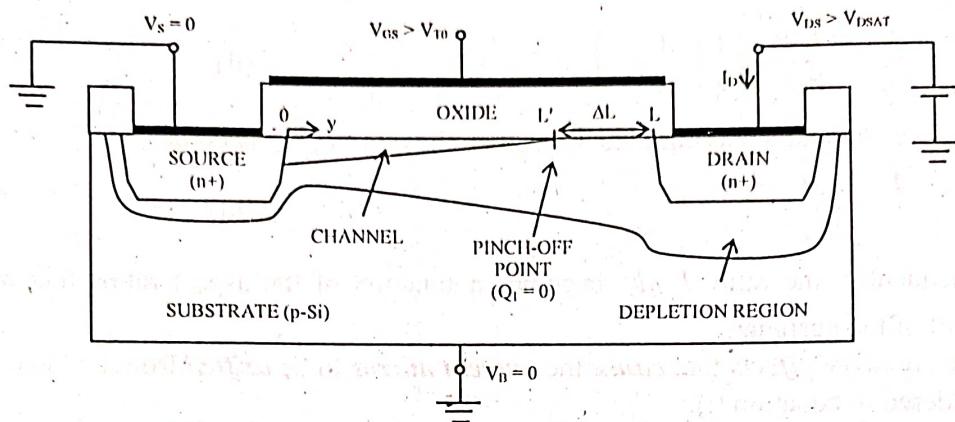


Fig: 1 Cross-section of a MOS operating in the saturation region

In the weak inversion region, the influence of the drain analogous to channel-length modulation leads to poorer device turn off behavior known as *drain-induced barrier lowering*, a drain induced lowering of threshold voltage.

In bipolar devices a similar increase in current is seen with increased collector voltage due to base-narrowing, known as the **Early effect**. The similarity in effect upon the current has led to use of the term "Early effect" for MOSFETs as well, as an alternative name for "channel-length modulation".

Channel length modulation in active mode usually is described using the **Shichman-Hodes model**, accurate only for old technology:

$$I_D = K_n' \frac{W}{L} (V_{GS} - V_{th})^2 (1 + \lambda V_{DS}) \quad \dots (1)$$

where  $I_D$  = drain current,  $K_n'$  = technology parameter sometimes called the transconductance coefficient,  $W$ ,  $L$  = MOSFET width and length,  $V_{GS}$  = gate-to-source voltage,  $V_{th}$  = threshold voltage,  $V_{DS}$  = drain-to-source voltage and  $\lambda$  = channel-length modulation parameter. The threshold voltage  $V_{th}$  in this formula is subject to drain-induced barrier lowering.

Channel-length modulation is important because it decides the **MOSFET output resistance**, an important parameter in circuit design of current mirrors and amplifiers.

### b) MOSFET Scaling:

High density chips in MOS VLSI require high packing density of MOSFETs in circuits. Consequently, transistor sizes should be as small as possible.

- Reduction in dimensions of MOSFET is called Scaling. Some operational characteristics of MOS transistors will change with reduction in dimensions. So, some physical limitations restrict the extent of scaling.

Two size reduction strategies are:

- Full Scaling or Constant Field Scaling
- Constant Voltage Scaling or Generalized Scaling.

Scaling concerned with systematic reduction of overall dimensions of the device, preserving the geometrical ratios found in larger devices. Proportional scaling of all devices in a circuit results in a reduction of total silicon area occupied by the circuit, increasing overall functional density of the chip.

All horizontal and vertical dimensions of large size transistors are then reduced by a scaling factor K ( $<1$ ) to obtain a scaled device. Extent of scaling is dominated by fabrication technology and by minimum feature size.

### 9. Explain how MOSFET can be used as resistor.

[WBUT 2016]

**Answer:**

#### MOS Resistor

MOS provides a good switch. The ON resistance of the device consists of series combination of  $r_D$ ,  $r_S$  and whatever channel resistance exists. By design,  $r_D$  and  $r_S$  contribution is less; hence channel resistance is the primary consideration.

In the ON state of the switch, voltage across the switch should be small and  $V_{GS}$  should be large. In this state, MOS device is assumed to be in the linear region.

$$\therefore I_D = \frac{K'W}{L} \left[ (V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \quad \dots (1)$$

So, the small signal channel resistance is

$$r_{ON} = \frac{1}{\partial I_D / \partial V_{DS}} \Big|_{Q=\text{Quiescent point}} = \frac{L}{K'W(V_{GS} - V_T - V_{DS})} \quad \dots (2)$$

A plot of  $r_{ON}$  as a function of  $V_{GS}$  for  $V_{DS} = 0.1V$  and for  $W/L$  equal to 1, 2, 5 & 10 is shown below.

From the above illustration and expressions it is clear that

- A lower value of  $r_{ON}$  achieved for large  $W/L$ .
- When  $V_{GS} \approx V_T$ ,  $r_{ON} \approx \infty$  as switch is turning off.
- When  $V_{GS} \leq V_T$ ,  $r_{OFF} = \infty$ . At this moment, performance of the device is dominated by drain - bulk and source - bulk leakage current from drain to source.

The non-linearity in the graph would become negligible where drain and source voltages vary over a small range so that the transistor ON resistance can be approximated as small signal resistance.

## POPULAR PUBLICATIONS

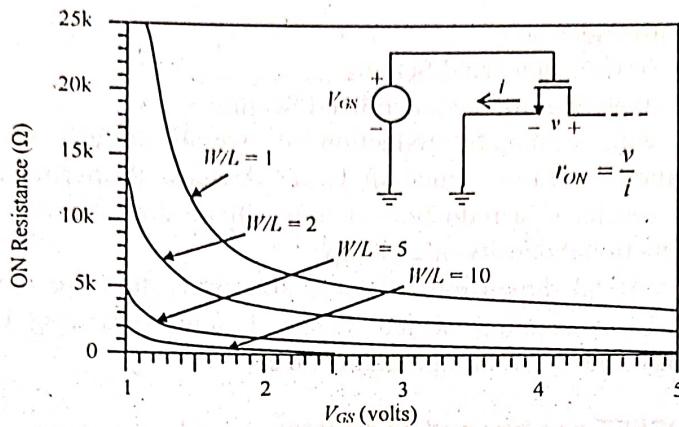


Fig: 1 Illustration of ON resistance of an n-channel

For small signal,  $V_{DS} \approx$  small,

$$\therefore r_{ds} = \frac{1}{K'W(V_{GS} - V_T)} \quad \dots (3)$$

### **MOS Switch as an A.C. Resistor**

An A.C. resistor provides an A.C. voltage for a given (A.C.) current for a zero (D.C.) current. The MOS switch is a good realization of an A.C. resistor since it is almost linear and has no D.C. offset. By controlling the value of voltage between the gate and source, one can get a linear A.C. resistance for values of  $\Delta V$  up to  $1V$ .

**10. What is MOSFET scaling? What is the need of scaling? Compare various types of scaling.** [WBUT 2017]

**Answer:**

**1<sup>st</sup> Part: Refer to Question No. 8(b) of Short Answer Type Questions.**

**2<sup>nd</sup> & 3<sup>rd</sup> Part:**

### **Constant Field Scaling**

The principle of constant field scaling is that device dimensions and device voltages be scaled such that electric fields (both horizontal and vertical) remain essentially constant. To ensure that the reliability of the scaled device is not compromised, the electric fields in the scaled device should remain constant.

Fig 1(a) shows the cross section and parameters of an original NMOS device and fig. 1(b) shows the scaled device, where the scaling parameter is  $k$ . Typically,  $k \approx 0.7$  per generation of a given technology.

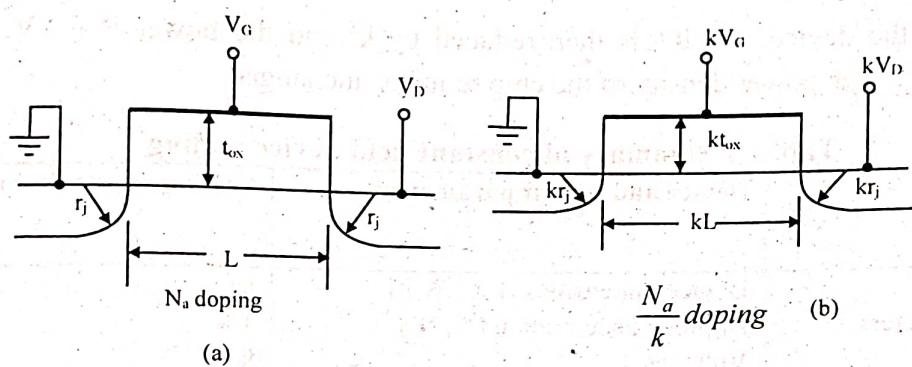


Fig: 1 Cross section of a) original NMOS transistor and  
b) scaled NMOS transistor

As shown in figure, the channel length is scaled from  $L$  to  $kL$ . To maintain a constant horizontal field, the drain voltage must also be scaled from  $V_D$  to  $kV_D$ . The maximum gate voltage will also be scaled from  $V_G$  to  $kV_G$  so that the gate and drain voltages remain compatible. To maintain a constant vertical electric field, the oxide thickness should also be scaled from  $t_{ox}$  to  $kt_{ox}$ .

The maximum depletion width at the drain terminal, for a one sided p-n junctions, is

$$x_D = \sqrt{\frac{2\epsilon(V_{bi} + V_D)}{qN_a}} \quad \dots (1)$$

where  $V_D$  – drain voltage and

$V_{bi}$  – built in potential

$$= \frac{KT}{q} \ln \left( \frac{N_a N_d}{n_i^2} \right)$$

Since the channel length is being reduced, the depletion widths also need to be reduced. If the substrate doping concentration is increased by a factor ( $1/k$ ), then the depletion width is reduced by approximately the factor  $k$  since  $V_D$  is reduced by  $k$ .

The drain current per channel width, for the transistor biased in the saturation region, can be written as

$$\frac{I_D}{W} = \frac{\mu_n \epsilon_{ox}}{2 t_{ox} L} (V_G - V_T)^2 \quad \dots (2)$$

where  $t_{ox}$  – thickness of oxide layer

$W$  &  $L$  – width and length of the channel

$V_G$  &  $V_T$  – gate voltage and threshold voltages.

After scaling, the drift current per channel width remains essentially a constant as

$$t_{ox} \rightarrow kt_{ox}, L \rightarrow kL \text{ and } V_G \rightarrow kV_G.$$

The area of the device,  $A \approx WL$ , is then reduced by  $k^2$  and the power  $P = I \cdot V$  is also reduced by  $k^2$ . The power density of the chip remains unchanged.

**Table: 1 Summary of constant field device scaling**

	Device and circuit parameters	Scaling factor ( $k < 1$ )
Scaled parameters	Device dimensions ( $L, t_{ox}, W, r_j$ ) Doping concentration ( $N_a, N_d$ ) Voltages	$k$ $1/k$ $k$
Effect on device parameters	Electric field Carrier velocity Depletion widths Capacitance ( $C = \epsilon A/t$ ) Drift current	$1$ $1$ $k$ $k$ $k$
Effect on circuit parameters	Device density Power density Power dissipation per device ( $P = VI$ ) Circuit delay time ( $\approx CV/I$ ) Power delay product ( $P_t$ )	$1/k^2$ $1$ $k^2$ $k$ $k^1$

Table 1 summarizes the device scaling and the effect on circuit parameters. The width and length of interconnect lines are also assumed to be reduced by the same scaling factor.

### Constant Voltage Scaling

In constant field scaling, the applied voltages are scaled with the same scaling factor  $k$  as the device dimensions. However, peripheral and interface circuits sometimes require certain voltage levels for inputs and outputs. So, for constant field scaled devices, level shifters and multiple power supply voltages are required. For this reason, constant voltage scaling is preferred in these cases. In addition, other factors that do not scale, such as threshold voltage and sub threshold currents have made the reduction in applied voltages less desirable. As a consequence, electric fields in MOS devices have tended to increase as device dimensions shrink. In constant voltage scaling, all dimensions of the MOSFET are reduced by a factor  $k$ .

The power supply voltages and terminal voltage remain unchanged. Doping densities must be increased by a factor  $1/k^2$  in order to preserve charge field relation.

- I. Gate oxide capacitance per unit area,  $C_{ox}$  increased by  $1/k$ , therefore transconductance increased by  $1/k$ .
- II. The drain current is increased by  $1/k$ , and drain current density is increased by  $1/k^3$  which may cause reliability problem for MOS transistor.

III. Power density per unit is increased by  $1/k^3$ . So, the drawbacks of constant voltage scaling are the reliability problems due to increased drain current density and power density such as electro migrations, hot carrier degradation, oxide breakdown and electrical overstress. As the oxide thickness is reduced and the electric field is increased, gate oxides are closer to breakdown and oxide integrity may be more difficult to maintain. In addition, direct tunneling of carriers through the oxide may be more likely to occur.

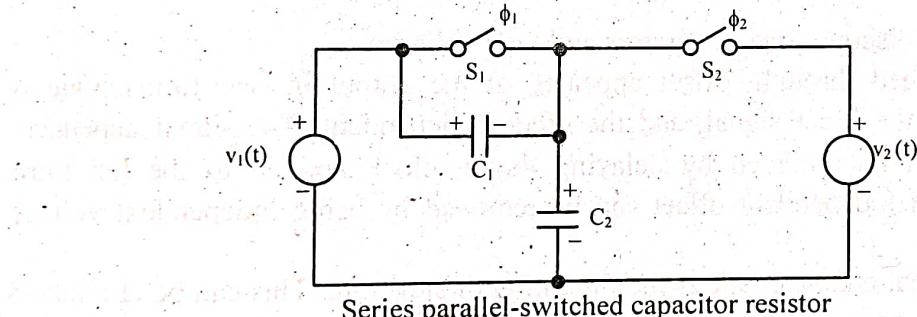
### Long Answer Type Questions

- What do you mean by Series-Parallel switched capacitor circuit? Describe briefly.
- Describe the different types of Switched Capacitor Integrator Circuit. Describe the drawbacks of discrete-time integrator. How do you solve this drawback?
- Design the 1st and 2nd order switched capacitor low-pass filters.

[WBUT 2013, 2015]

**Answer:**

- a) Series-parallel switched capacitor circuit



In this circuit  $i_1(t)$  is flowing during both  $\phi_1$  and  $\phi_2$  clock periods.

Therefore

$$\begin{aligned}
 i_1(av) &= \frac{1}{T} \int_0^T i_1(t) dt = \frac{I}{T} \left[ \int_{T/2}^r i_1(t) dt + \int_{r/2}^{T/2} i_1(t) dt \right] \\
 &= \frac{I}{T} \left[ \int_0^{T/2} dq_1(t) + \int_{T/2}^r dq_1(t) \right] = \frac{q_1(T/2) - q_1(0)}{T} + \frac{q_1(T) - q_1(T/2)}{T} \\
 &= \frac{C_2 [V_{C2}(T/2) - V_{C2}(0)]}{T} + \frac{C_1 [V_{C1}(T) - V_{C1}(T/2)]}{T}
 \end{aligned}$$

$$V_{C2}(T/2) = v_1, V_{C2}(0) = v_2, V_{C1}(T) = v_1 - v_2$$

## POPULAR PUBLICATIONS

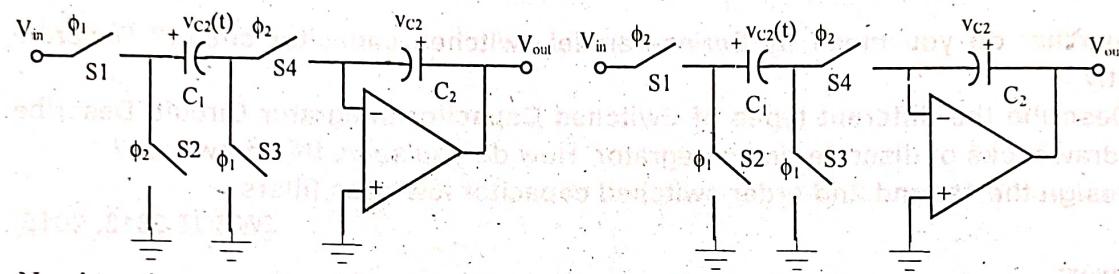
$$V_{C1}(T/2) = 0$$

Therefore,

$$i_1(\alpha v) = \frac{C_2(v_1 - v_2)}{T} + \frac{C_1(v_1 - v_2)}{T} = \frac{1}{T}(v_1 - v_2)(C_2 + C_1) = \frac{v_1 - v_2}{T/C_1 + C_2}$$

$$\text{So; } R = \frac{T}{C_1 + C_2}$$

b) There are two different types of switched capacitor integrator circuits



Non-inverting, stray insensitive integrator

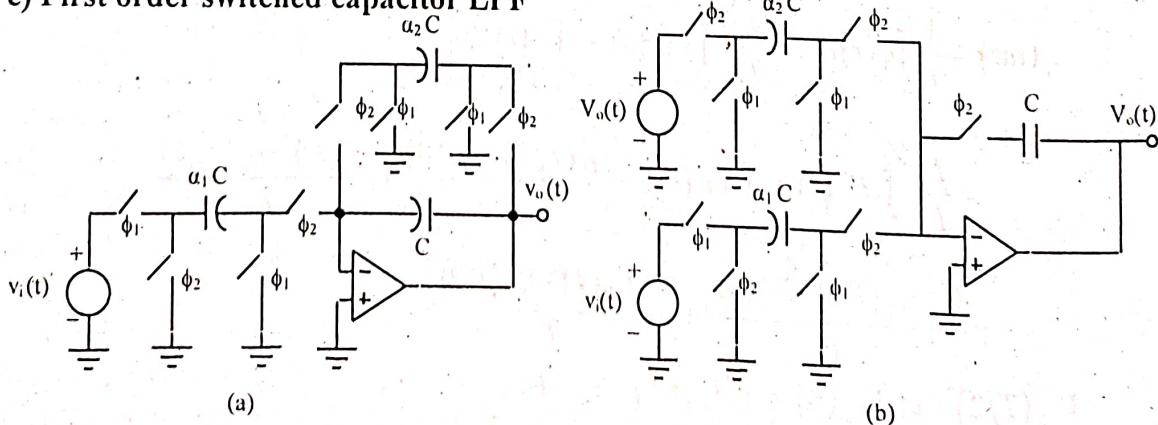
Inverting, stray insensitive integrator

Drawbacks of those discrete time integrators and their solutions:

- Due to clock feed through, offset appeared at the output in two forms. One is independent of the input signal, and the other is dependent. The signal dependent input offset can be removed by delaying the  $\phi_1$  clock applied to the left most switches. Signal independent offset can be removed by using independent voltage sources.
- Another non-ideal characteristic is the limitation of slew rate. This can be eliminated if the following inequality is satisfied.

$$\frac{\Delta V_o(\max)}{SR} < \frac{T}{2}$$

c) First order switched capacitor LPF



Summing currents toward the inverting op-amp terminal gives

$$\alpha_2 C v_o^e(Z) - \alpha_1 C Z^{-\frac{1}{2}} v_i^o(Z) + C(1 - Z^{-1}) v_o^e(Z) = 0$$

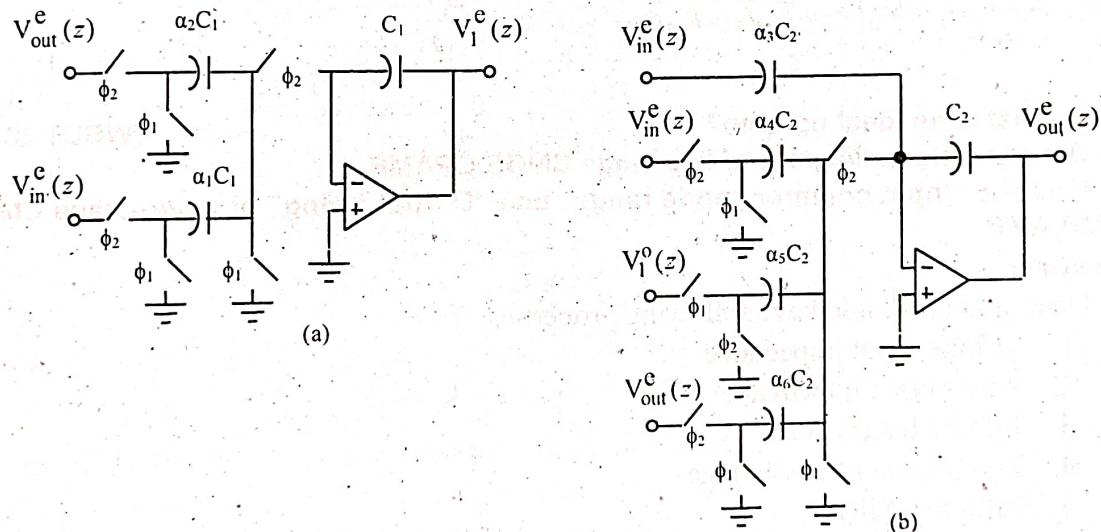
Solving for  $v_o^e(Z)/v_i^o(Z)$ , we get

$$\frac{v_o^e(Z)}{v_i^o(Z)} = \frac{Z^{-\frac{1}{2}} v_o^e(Z)}{v_i^o(Z)} = \frac{\alpha_1 Z^{-\frac{1}{2}}}{1 + \alpha_2 - Z^{-1}} = \frac{\frac{\alpha_1 Z^{-1}}{1 + \alpha_2}}{1 - \frac{z^{-1}}{1 + \alpha_2}}$$

Equating above equation with standard LPF equation, we get,

$$\alpha_1 = \frac{A_0}{B_0} \text{ and } \alpha_2 = \frac{1 - B_0}{B_0}$$

## 2<sup>nd</sup> order switched capacitor LPF



The outputs of the circuits in fig. in Z-domain as

$$V_i^e(Z) = \frac{-\alpha_1}{1 - Z^{-1}} V_{in}^e(Z) - \frac{\alpha_2}{1 - Z^{-1}} V_{out}^e(Z) \quad \dots (1)$$

for fig. (a) and

$$V_o^e(Z) = \alpha_3 V_{in}^e(Z) - \frac{\alpha_4}{1 - Z^{-1}} V_{in}^e(Z) + \frac{\alpha_5 Z^{-1}}{1 - Z^{-1}} V_i^e(Z) - \frac{\alpha_6}{1 - Z^{-1}} V_{out}^e(Z) \quad \dots (2)$$

for fig. (b)

Here,  $V_i^o(Z)$  is multiplied with  $Z^{-\frac{1}{2}}$  to get  $V_i^e(Z)$ . Assuming  $\omega T \ll 1$ , than  $1 - Z^{-1} \equiv ST$ . So, equation (1) and (2) can be approximated as

**POPULAR PUBLICATIONS**

$$\begin{aligned} V_1^e(S) &\equiv -\frac{\alpha_1}{ST} V_{in}^e(S) - \frac{\alpha_2}{ST} V_{out}^e(S) \\ &= -\frac{1}{S} \left[ \frac{\alpha_1}{T} V_{in}^e(S) + \frac{\alpha_2}{T} V_{out}^e(S) \right] \end{aligned} \quad \dots(3)$$

$$\text{and, } V_{out}^e(S) \equiv -\frac{1}{S} \left[ \left( S\alpha_3 + \frac{\alpha_4}{T} \right) V_{in}^e(S) - \frac{\alpha_5}{T} V_1^e(S) + \frac{\alpha_6}{T} V_{out}^e(S) \right] \quad \dots(4)$$

Equation (3) and (1) can be combined to get

$$H^{ee}(S) = \frac{-\left(\alpha_3 S^2 + \frac{\alpha_4}{T} S + \frac{\alpha_1 \alpha_5}{T^2}\right)}{S^2 + \frac{\alpha_6 S}{T} + \frac{\alpha_2 \alpha_5}{T^2}} \quad \dots(5)$$

Comparing equation (5) with standard equation, we get

$$\alpha_1 = \frac{K_0 T}{\omega_0}, \alpha_2 = |\alpha_5| = \omega_0 T, \alpha_3 = k_2, \alpha_4 = K_1 T, \alpha_6 = \frac{\omega_0 T}{Q}$$

2. a) What is an ideal op-amp?

[WBUT 2014]

b) Draw the block diagram of two stage CMOS OPAMP.

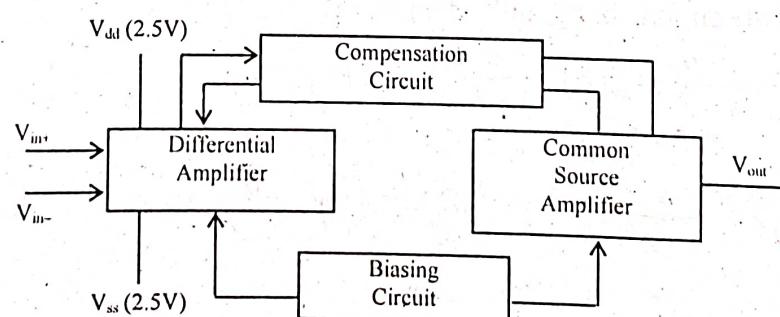
c) Find the "Input common mode range" and "Output swing" of a two-stage CMOS OPAMP.

**Answer:**

a) Ideal op-amp should have following properties:

1. Infinite input impedance
2. Zero output impedance
3. Infinite bandwidth
4. Zero output offset voltage
5. Infinite CMRR

b)



Block diagram of two stage CMOS Op-Amp

c) Input Common Mode Voltage Range is defined as the average voltage at the inverting and non-inverting input pins. If the common mode voltage gets too high or too low, the inputs will shut down and proper operation ceases. The common mode input voltage range, VICR, specifies the range over which normal operation is guaranteed.

Maximum Output Voltage Swing  $V_{OM\pm}$ , is defined as the maximum positive or negative peak output voltage that can be obtained without wave form clipping, when quiescent DC output voltage is zero.  $V_{OM\pm}$  is limited by the output impedance of the amplifier, the saturation voltage of the output transistors, and the power supply voltages.

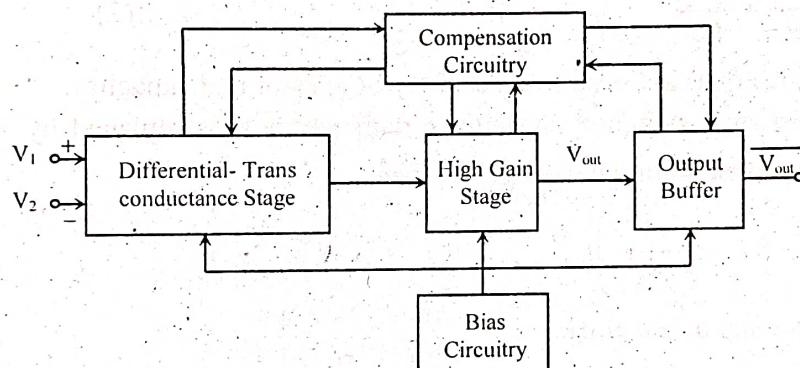
3. a) What is current mirror? Explain with proper circuit diagram. [WBUT 2016]

Answer:

Refer to Question No. 7 of Short Answer Type Questions.

b) Draw the block diagram of a two-stage op-amp and explain the function of each block. [WBUT 2016]

Answer:



Block diagram of a general two stage Op-amp

- **Differential trans-conductance stage:** Forms the input and sometimes provides the differential-to-single-ended conversion.
- **High gain stage:** Provides the voltage gain required by the op amp together with the input stage.
- **Output buffer:** Used if the op amp must drive a low resistance.
- **Compensation:** Necessary to keep the op amp stable when resistive negative feedback is applied.

c) How can you realize the resistor using switched capacitor circuits? [WBUT 2016]

Answer:

The frequency or time precision of an analog signal processing circuit is determined by the accuracy of circuit time constant.

## POPULAR PUBLICATIONS

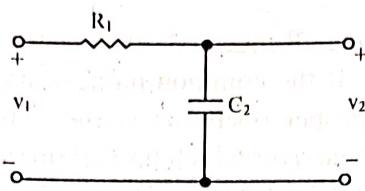


Fig: 1 Continuous time first order low pass circuit

Let us consider the simple low pass first order filter circuit as shown in fig.1.  
The voltage transfer function in the frequency domain is,

$$H(j\omega) = \frac{v_2(j\omega)}{v_1(j\omega)} = \frac{1}{1 + j\omega R_1 C_2} = \frac{1}{1 + j\omega\tau} \quad \dots(1)$$

where  $\tau = R_1 C_2$

To compare the accuracy of a continuous time circuit with a switched capacitor circuit, let us designate  $\tau$  as  $\tau_c$ . The accuracy of  $\tau_c$  can be expressed as

$$\frac{d\tau_c}{\tau_c} = \frac{dR_1}{R_1} + \frac{dC_2}{C_2} \quad \dots(2)$$

So, the accuracy is equal to the sum of accuracy of resistor and capacitor.

Now, let us consider a switched capacitor circuit where  $R$  is replaced by  $T/C_1$ . If we consider the time constant in the case as  $\tau_D$ , then

$$\tau_D = \left(\frac{T}{C_1}\right) C_2 = \left(\frac{1}{f_C C_1}\right) C_2$$

where  $f_C$  is frequency of the clock.

Here the accuracy of  $\tau_D$  can be expressed as

$$\frac{d\tau_D}{\tau_D} = \frac{dC_2}{C_2} - \frac{dC_1}{C_1} - \frac{df_C}{f_C} \quad \dots(3)$$

So the accuracy of the discrete time constant  $\tau_D$  is equal to the relative accuracy of  $C_1$  and  $C_2$  and accuracy of the clock frequency. If the clock frequency is perfectly accurate, the accuracy of  $\tau_D$  can be controlled within 0.1% in standard CMOS technology.

### d) What is switched capacitor filter?

**Answer:**

[WBUT 2016]

The sensitivities of the active RC filters with respect to the changes in the environmental conditions depend on the control of the RC products and the technology. Besides RC resistors and capacitors have large tolerance and poor linearity. Large valued resistors and capacitors also occupy a considerable space during IC fabrication. The switched capacitor filters avoid many of these problems and can be realized in the monolithic form

using CMOS technology. This has the advantage in that the analog filters can be placed with the other digital circuits in the same IC. Most active SC filters are realized by replacing the resistors in the active RC filters with the switched capacitors.

**4. a) Why is reference voltage required in IC? What are the criteria for a good reference voltage source in VLSI circuit?** [WBUT 2017]

**Answer:**

**1<sup>st</sup> Part:**

As ideal current or voltage reference provides a stable current or voltage which is independent of power supply and temperature. Analog circuits require these references to compare or for mirroring and/or amplifying output currents or voltages.

Reference term is used when the current or voltage values have more precision and stability than an ordinary source of current or voltage. The reference should be independent of load. So, to achieve this condition, for heavy load, buffer amplifiers are used to isolate the reference from load.

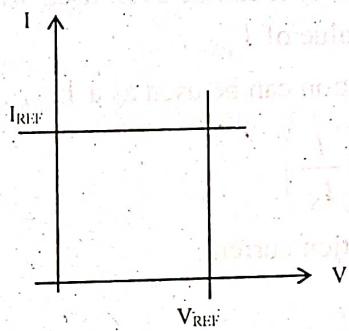


Fig: 1 Characteristics of ideal current and voltage references

The large signal current and voltage characteristics of an ideal current and voltage references are shown in fig (1).

The most simple voltage reference can be made from a voltage divider circuit, where passive or active components can be used as voltage divider elements. But in this circuit, the value  $V_{REF}$  is directly proportional to the power supply.

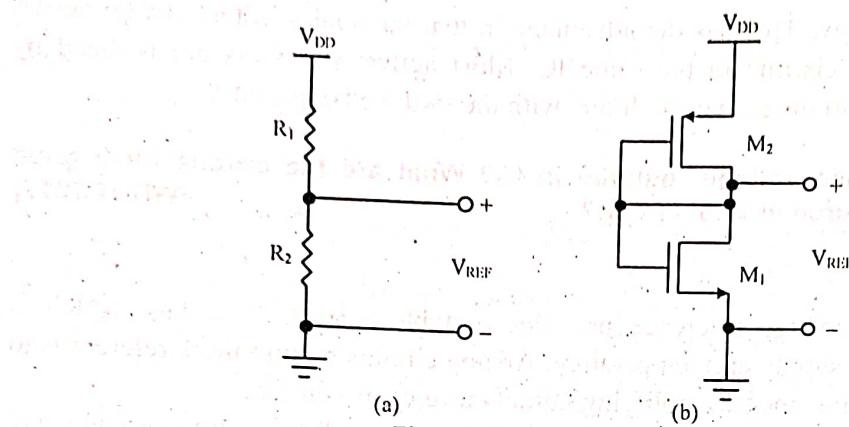
**2<sup>nd</sup> Part:**

The voltage should be independent of the load this is the actual criteria of a good voltage source in a VLSI system.

A measure of this relationship can be done for any  $V_{REF}$  circuit with the concept of sensitivity 'S'. The sensitivity of  $V_{REF}$  to the power supply  $V_{DD}$  can be expressed as

$$S = \frac{V_{REF}}{V_{DD}} = \frac{\partial V_{REF} / V_{REF}}{\partial V_{DD} / V_{DD}} = \frac{V_{DD}}{V_{REF}} \left( \frac{\partial V_{REF}}{\partial V_{DD}} \right) \quad \dots \quad (1)$$

## POPULAR PUBLICATIONS



a) Using passive components and b) Active devices

The performance of the circuit can be improved by connecting the substrate BJT to the power supply through a resistance. It can be even used with the previous passive voltage divider circuit to increase the value of  $V_{PFF}$ .

The voltage across the p-n junction can be used as a  $V_{bias}$ . So,

$$V_{REF} = V_{EB} = \frac{KT}{q} \ln \left( \frac{I}{I_S} \right) \quad \dots \dots (2)$$

where  $I_s$  is the junction saturation current.

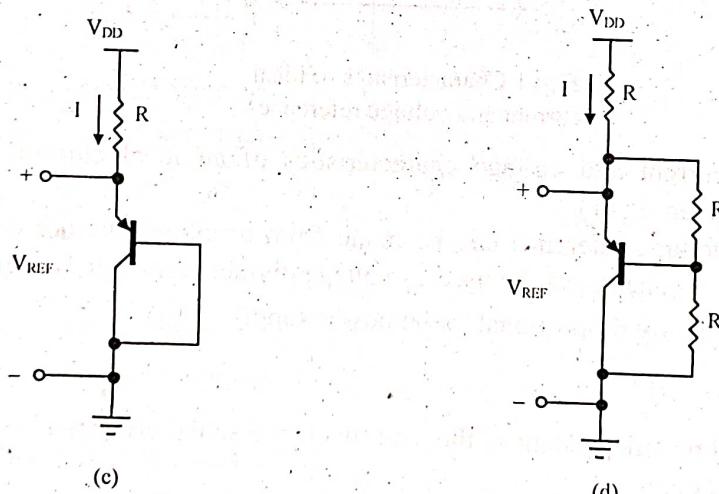


Fig: 2 p-n junction voltage references

If  $V_{DD}$  is much greater than  $V_{REF}$ , then the current I can be obtained as

$$I = \frac{V_{DD} - V_{EB}}{R} \cong \frac{V_{DD}}{R} \quad \dots (3)$$

And the reference voltage of the circuit is approximated to

$$V_{REF} \approx \frac{KT}{q} \ln \left( \frac{V_{DD}}{RI_S} \right) \quad \dots (4)$$

So, the sensitivity of  $V_{REF}$  would become,

$$S_{V_{DD}} V_{REF} = \frac{1}{\ln [V_{DD}/(RI_S)]} = \frac{1}{\ln(I/I_S)} \quad \dots (5)$$

As  $I$  is much greater than  $I_S$ , the sensitivity of  $V_{REF}$  is less than unity (the desired value).

The reference voltage of Fig 2 (b) is

$$V_{REF} \approx V_{EB} \left( \frac{R_1 + R_2}{R_1} \right) \quad \dots (6)$$

So if the common - emitter current gain ( $\beta$ ) and  $R_1$  and  $R_2$  are large, we get larger value of  $V_{REF}$  but the current  $I$  becomes a function of  $V_{REF}$ .

If the BJT in fig 2 be replaced with an enhancement type MOS device,  $V_{REF}$  would become less dependent on  $V_{DD}$ .

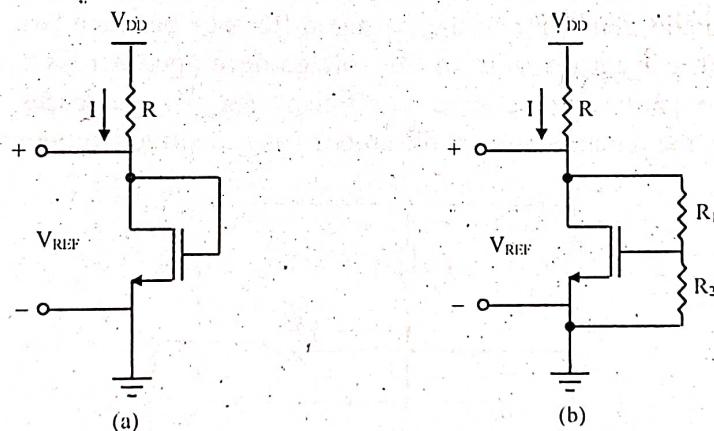


Fig: 3 MOS equivalent circuit of fig 2

Here,

$$V_{GS} = V_r + \sqrt{\frac{2I}{\beta}} \quad \dots (7)$$

Ignoring channel length modulation effect,

$$V_{REF} = V_T - \frac{1}{\beta R} + \sqrt{\frac{2(V_{DD} - V_T)}{\beta R}} + \frac{1}{\beta^2 R^2} \quad \dots (8)$$

## POPULAR PUBLICATIONS

Here,  $\beta$  is the trans conductance of MOS.

So the sensitivity can be found as

$$S = \frac{V_{REF}}{V_{DD}} = \left( \frac{1}{1 + \beta(V_{REF} - V_T)R} \right) \left( \frac{V_{DD}}{V_{REF}} \right) \quad \dots (9)$$

For circuit in Fig. 4.25(b) the reference voltage can be increased and is given by

$$V_{REF} = V_{GS} \left( 1 + \frac{R_2}{R_1} \right) \quad \dots (10)$$

In these circuits, geometry dependent parameters are changed to adjust  $V_{REF}$ .

For BJT, geometry dependent parameter is  $I_S$  and for MOS, geometry dependent parameter is  $W/L$ .

b) Explain the operation of a band gap voltage reference source in a VLSI circuit.

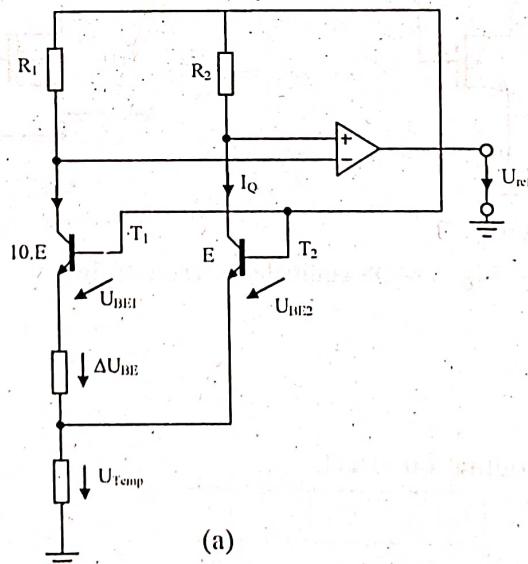
[WBUT 2017]

**Answer:**

### **Bandgap Voltage Reference**

The most common voltage reference circuit used in integrated circuits is the bandgap voltage reference. A bandgap-based reference (commonly just called a 'bandgap') uses analog circuits to add a multiple of the voltage difference between two bipolar junctions biased at different current densities to the voltage developed across a diode. The diode voltage has a negative temperature coefficient (i.e. it decreases with increasing temperature), and the junction voltage difference has a positive temperature coefficient.

### **Operation:**



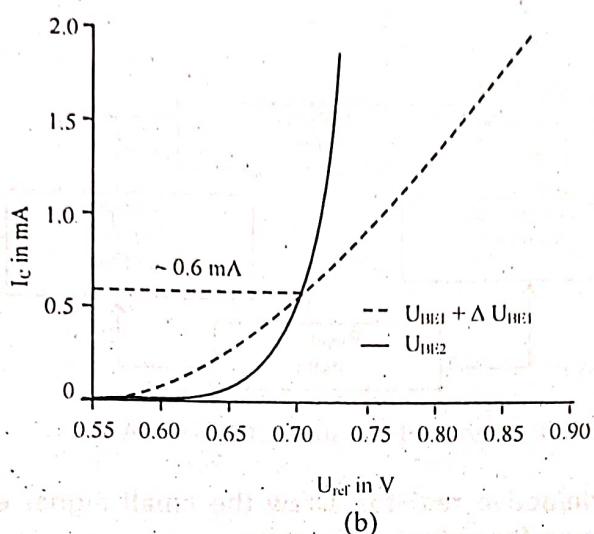


Fig: 1 (a) Circuit of a band gap reference (b) Characteristic and balance point of T1 and T2

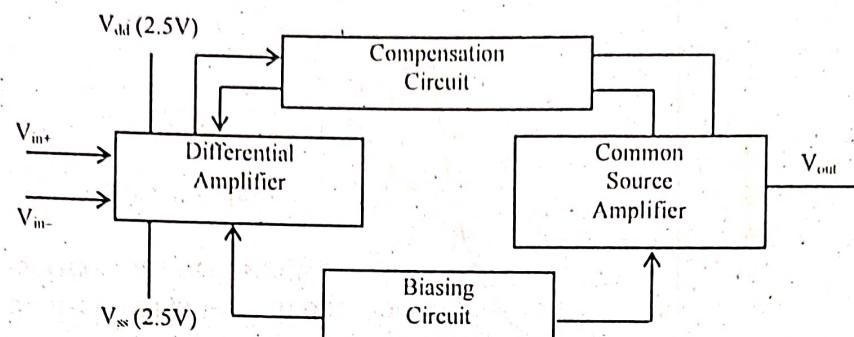
The voltage difference between two diodes, often operated at the same current and of different junction areas, is used to generate a proportional to absolute temperature (PTAT) current in a first resistor. This current is used to generate a voltage in a second resistor. This voltage in turn is added to the voltage of one of the diodes (or a third one, in some implementations). The voltage across a diode operated at constant current, or here with a PTAT current, is complementary to absolute temperature (CTAT—reduces with increasing temperature), with approximately  $-2$  mV/K. If the ratio between the first and second resistor is chosen properly, the first order effects of the temperature dependency of the diode and the PTAT current will cancel out. The resulting voltage is about  $1.2$ – $1.3$  V, depending on the particular technology, and is close to the theoretical  $1.22$  eV band gap of silicon at  $0$  K. The remaining voltage change over the operating temperature of typical integrated circuits is on the order of a few mV. This temperature dependency has a typical parabolic behaviour.

Because the output voltage is by definition fixed around  $1.25$  V for typical band gap reference circuits, the minimum operating voltage is about  $1.4$  V, as in a CMOS circuit at least one drain-source voltage of a FET (field effect transistor) has to be added.

c) Explain briefly different stages of an operational amplifier with the help of a block diagram. [WBUT 2017]

**POPULAR PUBLICATIONS**

**Answer:**



Block diagram of two stage CMOS Op-Amp

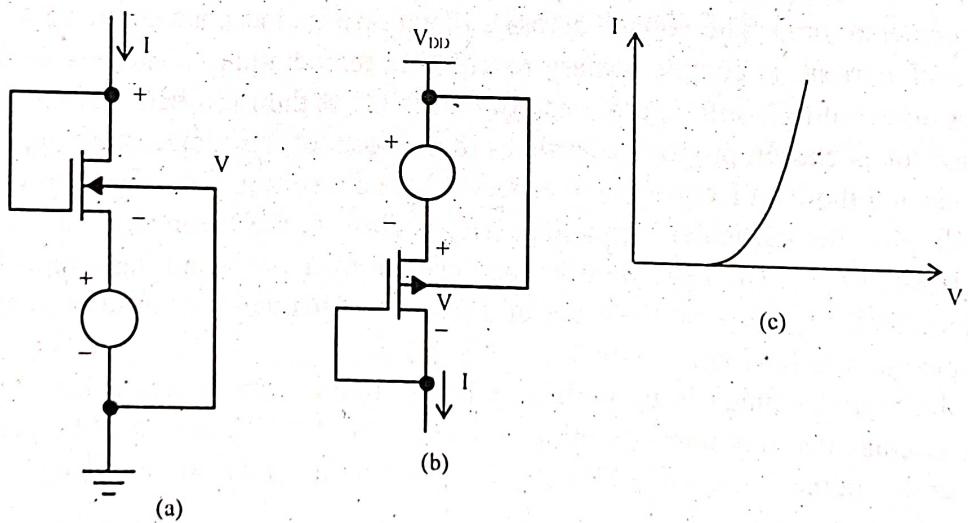
- What is MOS diode/active resistor. Draw the small signal equivalent circuit and find out the expression for output resistance.
- Explain how MOSFET can be used as Current Source or Current Sink?
- Realize a Resistor using Switched Capacitor.

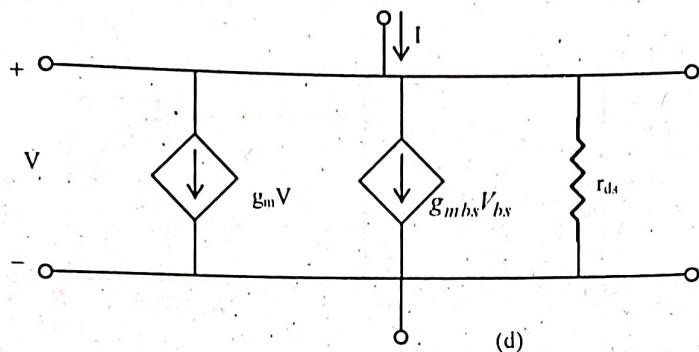
[WBUT 2018]

**Answer:**

- a) MOS Diode/active resistor

When the gate and drain of a MOS transistor are tied together, its I-V characteristics become similar to diode, called MOS diode.





Active resistor/Diode a) n-channel b) p-channel c) V-I characteristics of n-channel d) Small signal model

The drain current condition is described by the saturation current and given by:

$$I = I_D = \frac{K'}{2} \left( \frac{W}{L} \right) [(V_{GS} - V_T)^2] = \frac{\beta}{2} (V_{GS} - V_T)^2$$

where  $\beta$  is considered as trans-conductance as

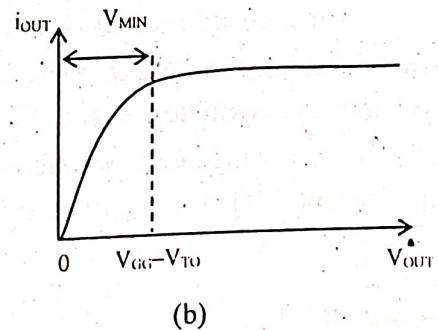
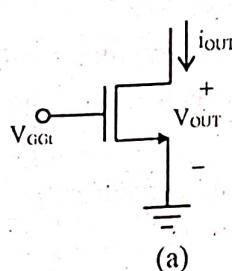
From the small signal model of a MOS diode the small signal resistance of the MOS diode can be calculated as:

$$r_{out} = \frac{1}{g_m + g_{mhs} + g_{ds}} \approx \frac{1}{g_m}$$

$$\text{where } g_m = \frac{\partial I_D}{\partial V_{GS}}, g_{mhs} = \frac{\partial I_D}{\partial V_{BS}}, g_{ds} = \frac{\partial I_D}{\partial V_{DS}}$$

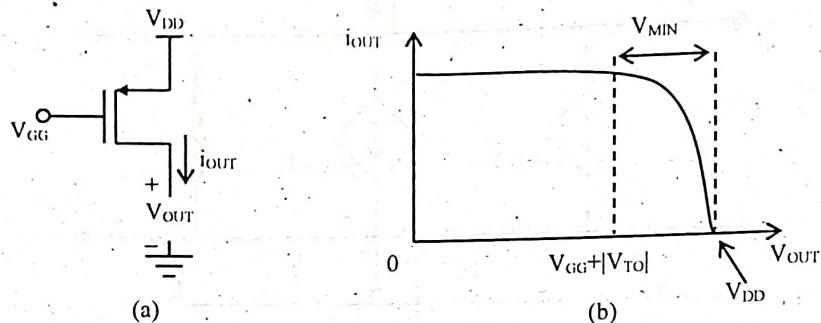
### b) Current Sinks and Sources:

- Current sinks and sources are two terminal components whose current at any instant of time is independent of the voltage across their terminals.
- Through sink and source, current flows from positive node to negative node.
- Current sink typically has a negative node at  $V_{SS}$  and a current source has a positive node at  $V_{DD}$ .
- The gate potential of the MOS should be taken to such level to create desired value of current. The non-saturation region of MOS device is not a good current source.



a) Current sink b) Current – voltage characteristics of current sink

## POPULAR PUBLICATIONS



a) current source b) current-voltage characteristics of current source

To perform current sink properly –

$$V_{OUT} \geq V_{GG} - V_{TO}$$

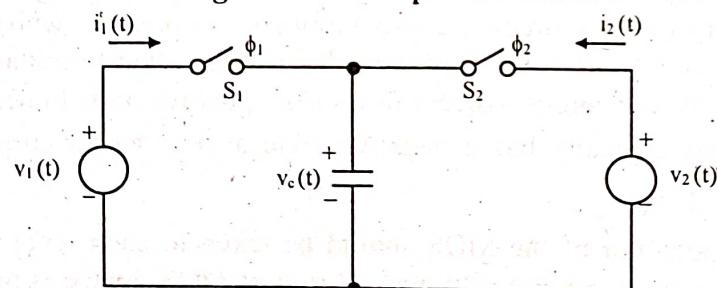
If both source and bulk are connected to ground, then small signal output is given by:-

$$r_{out} = \frac{1 + \lambda V_{DS}}{\lambda I_D} \approx \frac{1}{\lambda I_D}$$

Performance of current sink and source may be improved further by:

- i) Increasing the small signal output resistance, resulting in more constant current over the range of  $V_{OUT}$  values and
- ii). Reducing in  $V_{MIN}$  value, allowing a larger range of  $V_{OUT}$  over which current sink/source works properly.

### c) Realization of a resistor using Switched Capacitor Circuit



Parallel-switched capacitor equivalent resistor

In this type of switched capacitor circuit, there are two independent voltage sources  $v_1(t)$  and  $v_2(t)$ , two controlled switches  $S_1$  and  $S_2$  and a capacitor 'C'. The switches are controlled by the non-overlapping clock waveforms.

The value of average current  $i_1(t)$  flowing from  $v_1(t)$  into the capacitor 'C'. The average current is

$$i_1(av) = \frac{1}{T} \int_0^T i_1(t) dt$$

But  $i_1(t)$  only flows during the time  $0 \leq t \leq T/2$

$$\text{So, } i_1(\text{av}) = \frac{1}{T} \int_0^{T/2} i_1(t) dt$$

$$\text{We know that } i_1(t) = \frac{dq_1(t)}{dt}$$

$$\text{So, we get } i_1(\text{av}) = \frac{1}{T} \int_0^{T/2} dq_1(t) = \frac{q_1(T/2) - q_1(0)}{T}$$

The charge associated with a time domain capacitor is expressed as

$$q_C(t) = C.v_C(t)$$

$$\text{So, } i_1(\text{av}) = \frac{C}{T} [V_C(T/2) - V_C(0)]$$

$$\text{Here } V_C(T/2) = v_1(T/2) \equiv v_1 \quad \text{and} \quad V_C(0) = v_2(0) \equiv v_2$$

Therefore

$$i_1(\text{av}) = \frac{C}{T} [v_1 - v_2]$$

If a resistance 'R' is connected between two voltage sources  $v_1$  and  $v_2$ , the average current through R is

$$i(\text{av}) = \frac{v_1 - v_2}{R}$$

$$\text{Comparing above equations, we get } R = \frac{T}{C}$$

It shows that the parallel-switched capacitor circuit is equivalent to a resistor.

#### 6. Write short notes on the following:

- a) Switch capacitor [WBUT 2013]
- b) Folded-cascade amplifier [WBUT 2014]
- c) Description of the circuit and operation of a two stage OPAMP [WBUT 2015]
- d) MOS capacitor [WBUT 2016]
- e) Constant voltage scaling [WBUT 2017]
- f) Short channel effects [WBUT 2017]

Answer:

#### a) Switched Capacitor Circuits:

Previously analog signal processing was done by continuous time circuits consisting of resistors, capacitors and op-amps. But in modern VLSI circuits, analog sampled data

## POPULAR PUBLICATIONS

techniques are used to replace resistors, resulting in circuits consisting of only MOSFET switches, capacitors and op-amps. These circuits are called switched capacitor circuits. In these circuits we use two non overlapping clocks, which are used to drive MOS transistors as switches. If the clock frequency is perfectly accurate, the accuracy of  $\tau_D$  can be controlled within 0.1% in standard CMOS technology.

There are four different versions of the switched capacitor circuits, which are used to analyze the circuits. They are the parallel switch capacitor, the positive and negative switched capacitors, and a capacitor in series with a switch.

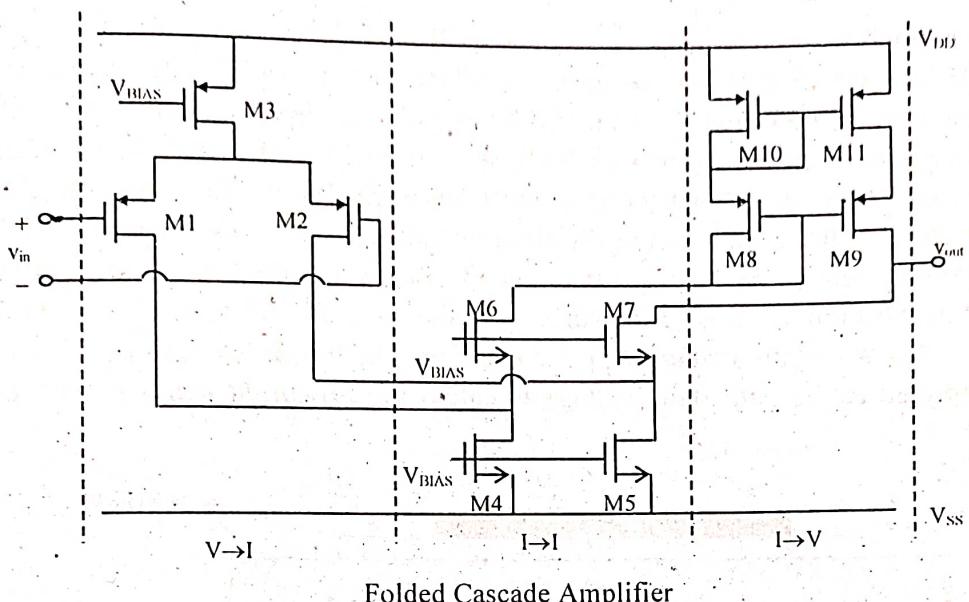
In four different types of two port switched capacitor circuits, the z-domain models contain three types of admittances.

- The first type of admittance is expressed as  
 $I(z) = Y \cdot z^0 V(z) = Y \cdot V(z)$   
Which is interpreted as a current  $I(z)$  of value  $YV(z)$  occurs with no delay when a voltage  $V(z)$  is applied.
- The second type of admittance is represented as  
 $I(z) = Y \cdot z^{-1/2} V(z)$   
Which is interpreted as current  $I(z)$  of value  $YV(z)$  occurring a half period after the voltage  $V(z)$  is applied.
- The third type of admittance is  
 $I(z) = (1 - z^{-1})Y \cdot V(z)$   
Which is interpreted as a current that first appears as a value of  $Y \cdot V(z)$  and at a period later this current is zero.

Switched capacitor circuits are used to design amplifiers, op-amps and filters.

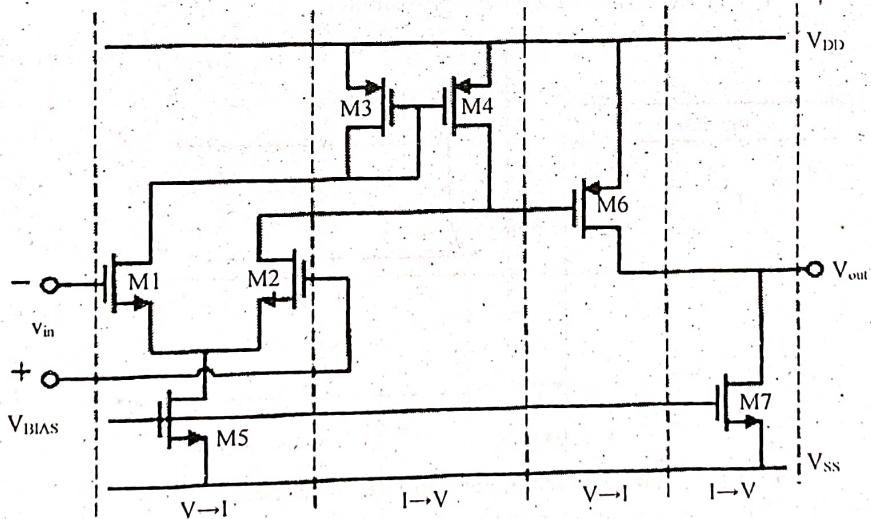
### b) Folded-cascade amplifier:

This type of amplifier has improved common mode range and power supply rejection, compared to conventional amplifier circuits. One of the advantages of the folded cascade op-amp is that it has a push-pull output. That is the op-amp can actively sink or source current from the load. This op-amp uses cascading in the output stage combined with a differential amplifier to achieve good input common mode range. Thus the folded cascade op-amp offers self-compensation, good input common mode range, and the gain of the two stage op-amp.



c) Description of the circuit and operation of a two stage OPAMP:

The two stage op-amp consists of a cascade of V to I and I to V stages as shown in figure below.



The first stage consists of a differential amplifier converting the differential input voltage to differential currents. These differential currents are applied to a current mirror load recovering the differential voltage. The second stage consists of a common source MOSFET converting the second stage input voltage to current. The driver transistor is loaded by current sink load, which converts the current to voltage at the output.

d) MOS capacitor:

In a MOS structure, there is an *oxide layer* below *Gate* terminal. Since oxide is a very good *insulator*, it contributes to an oxide capacitance in the circuit. Normally, the capacitance value of a capacitor doesn't change with values of voltage applied across its terminals. However, this is not the case with *MOS capacitor*. The capacitance of MOS capacitor changes its value with the variation in Gate voltage. This is because application of gate voltage results in the band bending of silicon substrate and hence variation in charge concentration at Si-SiO<sub>2</sub> interfaces. As shown in fig A2.5, the C-V<sub>G</sub> curve splits into two, after a certain voltage, depending upon the frequency (high or low) of AC voltage applied at the gate. This voltage is called the *threshold voltage* (V<sub>th</sub>) of MOS capacitor.

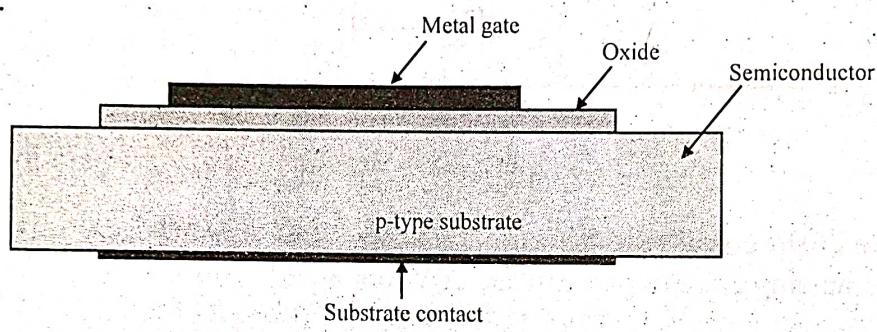


Fig: 1 Structure of a MOS capacitance

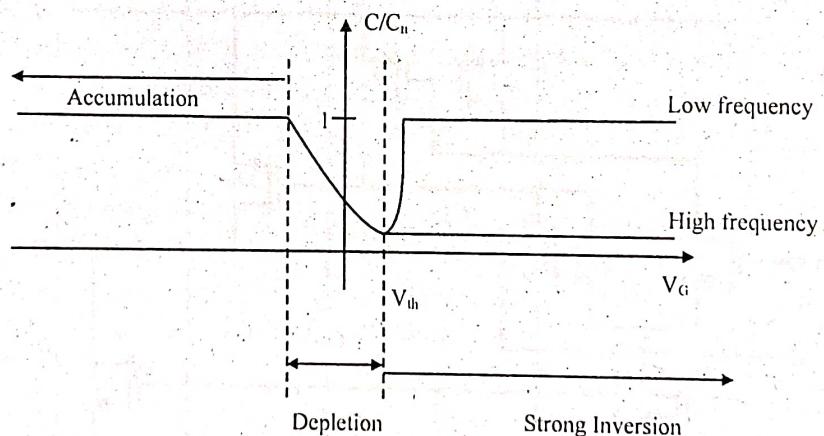


Fig: 2 The low frequency and high frequency C-V characteristics curves of a MOS capacitor

The low frequency or quasi-static measurement maintains thermal equilibrium at all times. This capacitance is the ratio of the change in charge to the change in gate voltage, measured while the capacitor is in equilibrium. A typical measurement is performed with an electrometer, which measures the charge added per unit time as one slowly varies the applied gate voltage.

The high frequency capacitance is obtained from a small-signal capacitance measurement at high frequency. The bias voltage on the gate is varied slowly to obtain the capacitance versus voltage. Under such conditions, one finds that the charge in the inversion layer does not change from the equilibrium value corresponding to the applied DC voltage. The high frequency capacitance therefore reflects only the charge variation in the depletion layer and the (rather small) movement of the inversion layer charge.

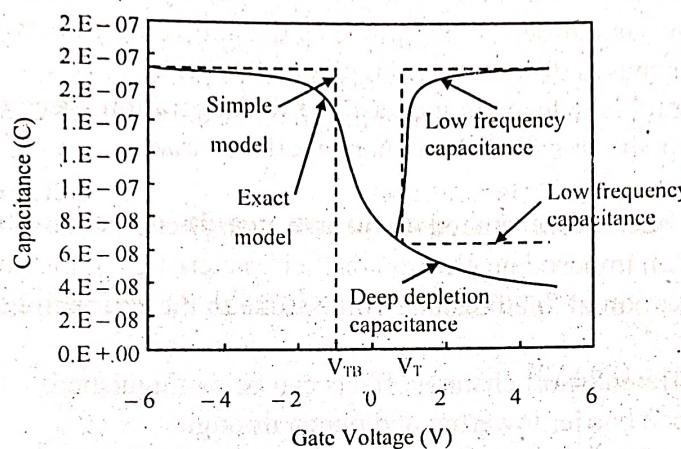


Fig: 3 Measured low and high frequency C-V curves

#### e) Constant voltage scaling:

In constant voltage scaling, all dimensions of the MOSFET are reduced by a factor  $k$ . The power supply voltages and terminal voltage remain unchanged. Doping densities must be increased by a factor  $1/k^2$  in order to preserve charge field relation. Gate oxide capacitance per unit area,  $C_{ox}$  increased by  $1/k$ , therefore transconductance increased by  $1/k$ .

The drain current is increased by  $1/k$ , and drain current density is increased by  $1/k^3$  which may cause reliability problem for MOS transistor.

Power density per unit is increased by  $1/k^3$ . So, the drawbacks of constant voltage scaling are the reliability problems due to increased drain current density and power density such as electro migrations, hot carrier degradation, oxide breakdown and electrical overstress. As the oxide thickness is reduced and the electric field is increased, gate oxides are closer to breakdown and oxide integrity may be more difficult to maintain. In addition, direct tunneling of carriers through the oxide may be more likely to occur.

In constant field scaling, the applied voltages are scaled with the same scaling factor  $k$  as the device dimensions. However, peripheral and interface circuits sometimes require certain voltage levels for inputs and outputs. So, for constant field scaled devices, level shifters and multiple power supply voltages are required. For this reason, constant voltage scaling is preferred in these cases. In addition, other factors that do not scale, such as

## POPULAR PUBLICATIONS

threshold voltage and sub threshold currents have made the reduction in applied voltages less desirable. As a consequence, electric fields in MOS devices have tended to increase as device dimensions shrink.

### f) Short channel effects:

Short channel effect in a MOSFET is an effect whereby the MOSFET in which the channel length is the same order of magnitude as the depletion layer widths of the source and drain junction, behaves differently from other MOSFETs.

As the channel length  $l$  is reduced to increase both the operation speed and the number of components per chip, the so called short channel effects arise.

The short channel effects are attributed on two physical phenomena:

1. The limitation imposed on electron drift characteristics in the channel.
2. The modification of the threshold voltage due to the shortening channel length.

In particular five different short channel effects can be distinguished.

1. Drain induced barrier lowering and punch through.
2. Surface scattering
3. Velocity saturation
4. Impact ionization
5. Hot electron effect.

## **11. Explain with a circuit diagram, operation of a differential amplifier. [WBUT 2017]**

**Answer:**

### **CMOS Differential Amplifier**

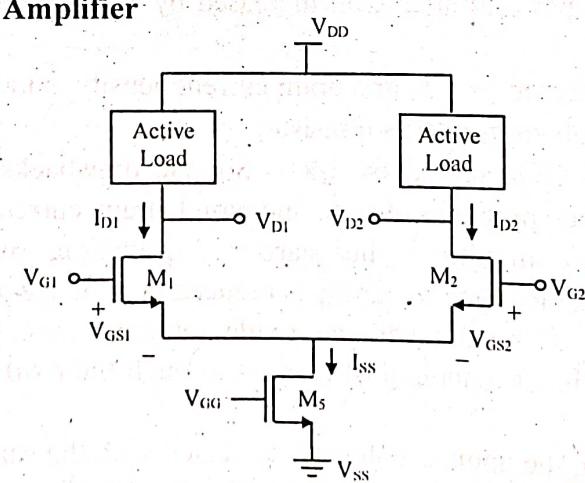


Fig: 1(a) A general MOS differential amplifier configuration

The figure shows the circuit of a general MOS differential amplifier. The active loads consist of MOS transistors at different configurations. The key aspect of the differential amplifier is the input source coupled pair, M<sub>1</sub> and M<sub>2</sub>.

For large signal considerations, let us assume that M<sub>1</sub> and M<sub>2</sub> are in saturation and their threshold voltages are equal. Neglecting channel modulation,

$$V_{ID} = V_{G1} - V_{G2} = V_{GS1} - V_{GS2} = \left| \frac{2I_{D1}}{\beta_1} \right|^{\frac{1}{2}} - \left| \frac{2I_{D2}}{\beta_2} \right|^{\frac{1}{2}} \quad \dots (1)$$

$$\text{and } I_{SS} = I_{D1} + I_{D2} \quad \dots (2)$$

where,  $\beta = K'(W/L)$ . Assuming  $\beta_1 = \beta_2 = \beta$ , from equation (1) and (2), we get

$$I_{D1} = \frac{I_{SS}}{2} + \frac{I_{SS}}{2} \left[ \frac{\beta V_{ID}^2}{I_{SS}} - \frac{\beta^2 V_{ID}^4}{4I_{SS}^2} \right]^{\frac{1}{2}} \quad \dots (3)$$

$$\text{and } I_{D2} = \frac{I_{SS}}{2} - \frac{I_{SS}}{2} \left[ \frac{\beta V_{ID}^2}{I_{SS}} - \frac{\beta^2 V_{ID}^4}{4I_{SS}^2} \right]^{\frac{1}{2}} \quad \dots (4)$$

These relationships are valid only for

$$|V_{ID}| \leq \left| \frac{2I_{SS}}{\beta} \right|^{\frac{1}{2}} \quad \dots (5)$$

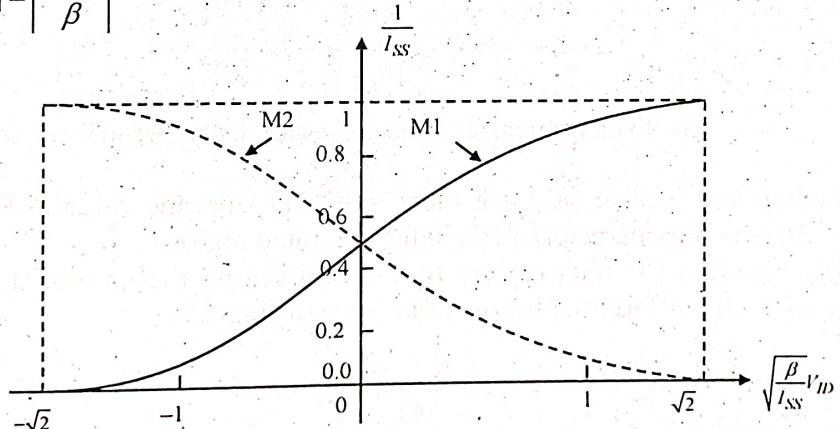


Fig: 1 (b) Large signal transconductance of MOS differential amplifier

The large signal voltage transfer characteristics of the differential amplifier can be found by using the equations (3) and (4) along with voltage current characteristics of the active load devices.

In this VTC, the regions where M<sub>1</sub> and M<sub>2</sub> are in saturation can be found by using equation (1) as  $V_{ID} = V_{G1} - V_{G2}$ , and if we assume symmetry and no common mode excitation, then

## POPULAR PUBLICATIONS

$$V_{G1} = \frac{V_{ID}}{2} \text{ and } V_{G2} = -\frac{V_{ID}}{2} \quad \dots (6)$$

Therefore M<sub>1</sub> is in saturation if,

$$V_{D1} \geq V_{G1} - V_{T,n} = \frac{V_{ID}}{2} - V_{T,n} \quad \dots (7)$$

and M<sub>2</sub> is in saturation if,

$$V_{D2} \geq -V_{G2} - V_{T,n} = -\frac{V_{ID}}{2} - V_{T,n} \quad \dots (8)$$

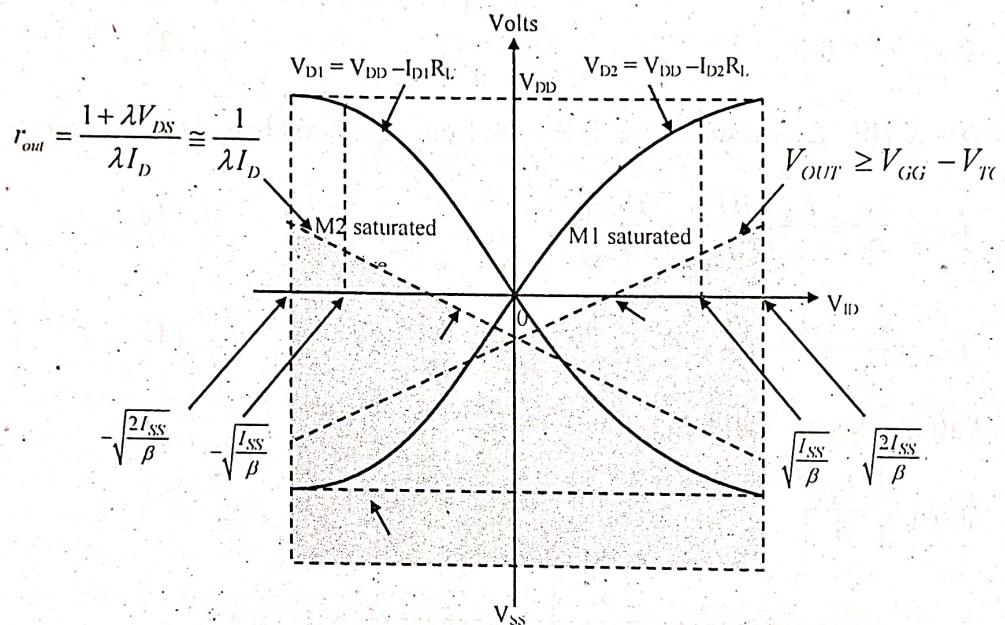


Fig: 3 voltage transfer characteristic of the differential amplifier

Assuming the load as resistive ( $R_L$ ) it is clear in Fig. (1) that the value of  $R_L$  determines how much of the transfer characteristic is in the saturated region.

Differentiating equation (3) with respect to  $V_{ID}$  and setting the quiescent value of  $V_{ID}$  equal to zero gives the differential transconductance of fig. (2) as

$$\begin{aligned} g_m &= \left. \frac{\partial I_D}{\partial V_{ID}} \right|_{V_{ID}=0} = \left( \frac{\beta_1 I_{SS}}{4} \right)^{1/2} = \left( \frac{K' I_{SS} W_1}{4 L_1} \right)^{1/2} \\ &= \left( \frac{K' I_{D1} W_1}{2 L_1} \right)^{1/2} \end{aligned} \quad \dots (9)$$

Here, the trans-conductance is half than the actual value. The reason for this is that only half of the input voltage is applied to M<sub>1</sub> or M<sub>2</sub> resulting in half the output current. This trans-conductance is called the differential-in, single ended output trans-conductance. The differential-out trans-conductance ( $g_{md}$ ), can be found by defining a differential output current  $I_{OD}$  as

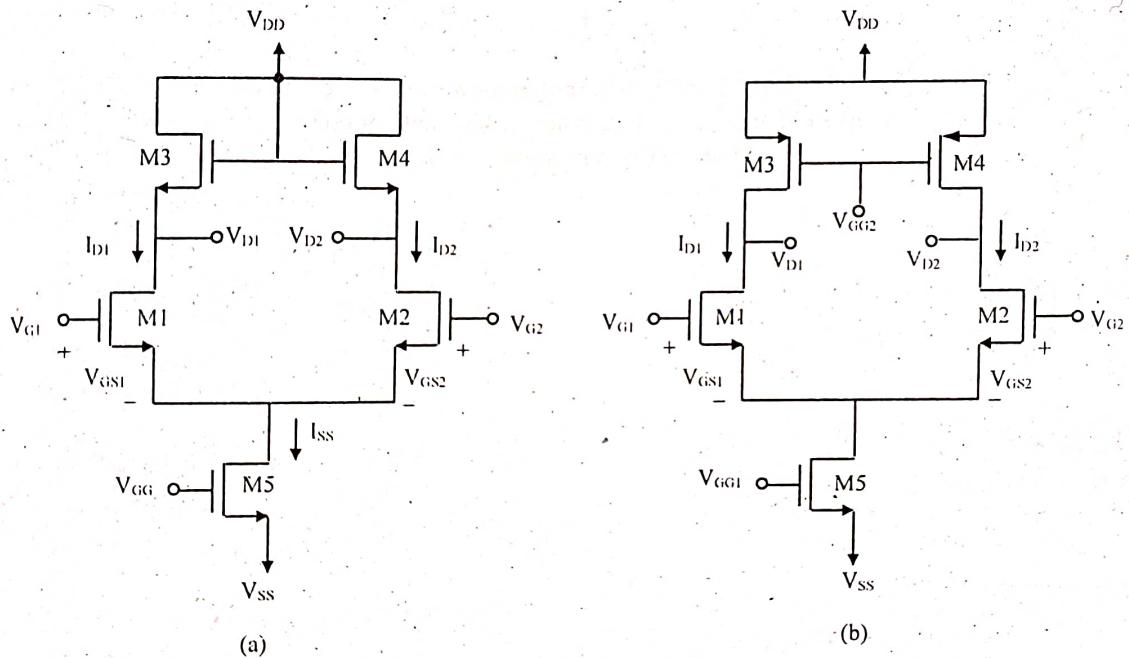
$$I_{OD} = I_{D1} - I_{D2} \quad (10)$$

And  $g_{ml}$  can be written as

$$g_{md} = \left. \frac{\partial I_{OD}}{\partial V_{ID}} \right|_{V_{ID}=0} = (\beta_1 I_{ss})^{\frac{1}{2}} = \left( \frac{K' I_{ss} W_1}{L_1} \right)^{\frac{1}{2}} = \left( \frac{2 K' I_{D1} W_1}{L_1} \right)^{\frac{1}{2}} \quad \dots (11)$$

This trans-conductance is exactly equal to the trans-conductance of common source transistor if  $I_{D1}$  is half of  $I_{SS}$ .

The output voltage of the differential amplifier depends on how the active loads of Fig. (1) are implemented.



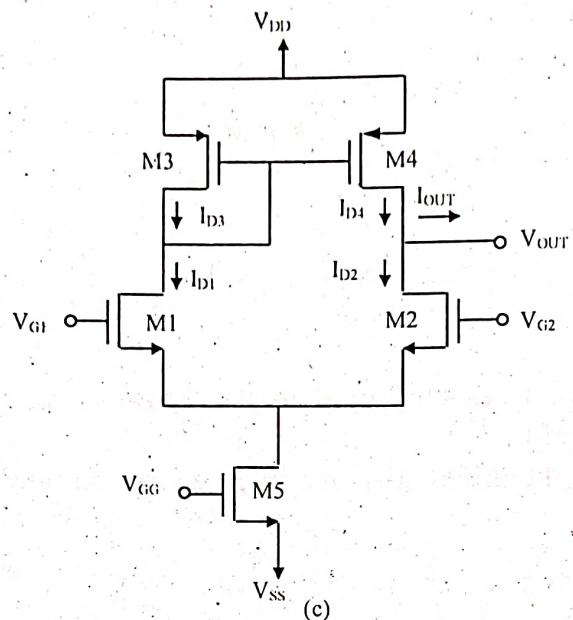


Fig: 4 MOS differential amplifier

- a) Enhancement active loads
- b) Current source loads
- c) Current mirror load

# UNIT PROCESS IN VLSI AND IC FABRICATION

## Multiple Choice Type Questions

1. The output of physical design is [WBUT 2014, 2018]  
 a) Layout      b) Mask      c) RTL      d) Circuit Design

Answer: (b)

2. Material used for the fabrication of gate in modern MOSFET [WBUT 2015]  
 a) highly pure Si      b) highly doped polysilicon  
 c) highly pure SiO<sub>2</sub>      d) none of these

Answer: (b)

3. The difference between a depletion load MOSFET and enhancement MOSFET only is the absence of [WBUT 2015]  
 a) insulated gate      b) electrons  
 c) induced channel      d) none of these

Answer: (c)

4. Czochralski Process is used to form [WBUT 2015]  
 a) wafer      b) ingot      c) IC      d) none of these

Answer: (a)

5. Twin-Tub process is widely used for fabrication of [WBUT 2015]  
 a) CMOS      b) PMOS      c) NMOS      d) none of these

Answer: (a)

6. CMOS logic gates are intrinsically [WBUT 2015]  
 a) inverting      b) non inverting  
 c) neither inverting nor non inverting      d) none of these

Answer: (a)

7. A positive photoresistor is [WBUT 2015]  
 a) initially insoluble and become soluble after exposure to UV  
 b) initially soluble and become insoluble after exposure to UV  
 c) initially opaque and become transparent after exposure to UV  
 d) initially transparent and become opaque after exposure to UV

Answer: (a)

### **POPULAR PUBLICATIONS**

7. What are the faults which can be reduced by  
a) Single transistor dynamic RAM      b) Pseudo-static RAM      c) Static RAM
- Answer: (c)
9. The static power dissipation is NIL for  
a) Single transistor dynamic RAM      b) Pseudo-static RAM      c) Static RAM
- [WBUT 2016]
10. Maximum transistor gate required to design XOR gate in CMOS structure  
a) 6      b) 8      c) 12      d) 10
- [WBUT 2017]
- Answer: (c)
11. Scaling is done for  
a) improving switching capacity      b) decreasing the power dissipation  
c) reduce chip size      d) all of these
- [WBUT 2017]
- Answer: (d)
12. Minimum TTL gates required to design XOR gate is  
a) Six      b) Four      c) Eight      d) Ten
- [WBUT 2018]
- Answer: (b)
13. What are the dominant faults in diffusion layers?  
a) Short circuit faults      b) Open circuit faults  
c) Short Open circuit faults      d) Power supply faults
- [WBUT 2018]
- Answer: (a)
14. What is minimum poly width in layout design rule?  
a)  $1\lambda$       b)  $2\lambda$       c)  $3\lambda$       d)  $4\lambda$
- [WBUT 2018]
- Answer: (b)

### **Short Answer Type Questions**

1. Describe photo-lithography process.

[WBUT 2013]

Answer:

Photolithography is the process of transferring geometric shapes on a mask to the surface of a silicon wafer. The steps involved in the photolithographic process are, wafer cleaning; barrier layer formation; photo-resist application; soft baking; mask alignment; exposure and development; and hard-baking. It uses light to transfer a geometric pattern from a photo mask to a light sensitive chemical "photoresist", on the substrate. A series of chemical treatment then either engraves the exposure pattern into, or enables deposition of a new material in the desired pattern upon, the material underneath the photo resist.

2. Describe the N-well CMOS fabrication process.

[WBUT 2015, 2017]

**Answer:**

*Refer to Question No. 4(b) of Long Answer Type Questions.*

**3. What are the differences in between diffusion and ion implantation? [WBUT 2017]**

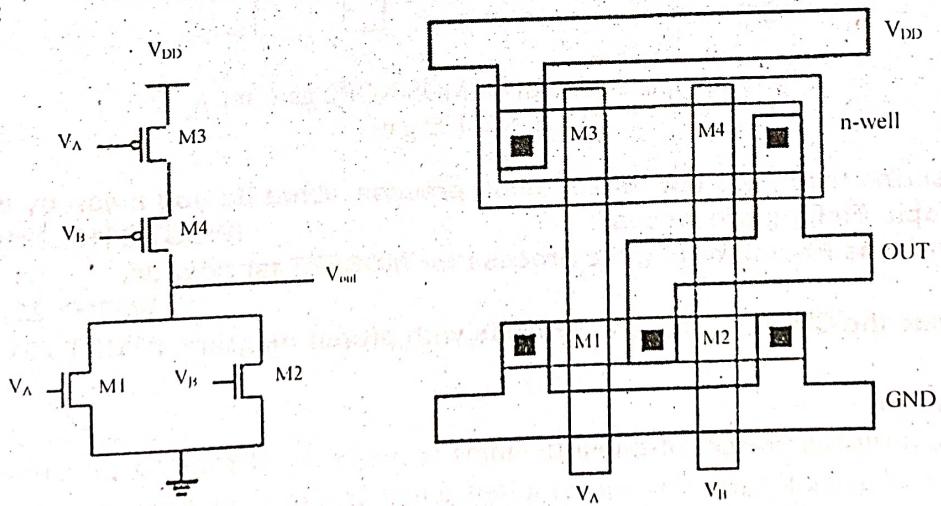
**Answer:**

Ion Implantation	Diffusion
It is a low temperature process used to change the chemical and physical properties of a material.	Diffusion can be defined as the motion of the impurities within a substance.
Isotropic and very directional.	Isotropic and mainly includes lateral diffusion.
Done at low temperature.	Done at high temperature.
Amount of dopant can be controlled.	Amount of dopant cannot be controlled.
May sometimes damage the surface of the target.	Does not damage the surface of the target.
Expensive because it requires specific equipment.	Comparatively less expensive.

### Long Answer Type Questions

**1. What do you mean by 'Lambda Rule' & 'Micron Rule'? Draw the Layout & Schematic diagram of a static CMOS NAND/NOR gate & identify the corresponding components in the two drawing. [WBUT 2013, 2015, 2017]**

**Answer:**



The VLSI design rules are usually described in two ways:

- Micron rules, in which the layout constraints such as minimum feature sizes and minimum allowable feature separations, are stated in terms of absolute dimensions in micrometers, or,

## POPULAR PUBLICATIONS

- Lambda rules, which specify the layout constraints in terms of a single parameter ( $\lambda$ ) and, thus, allow linear, proportional scaling of all geometrical constraints. Lambda-based layout design rules were originally devised to simplify the industry-standard micron-based design rules and to allow scaling capability for various processes.

Figure below shows the sample layouts of a two-input NOR gate and a two-input NAND gate, using single-layer poly-silicon and single-layer metal. Here, the p-type diffusion area for the pMOS transistors and the n-type diffusion area for the nMOS transistors are aligned in parallel to allow simple routing of the gate signals with two parallel poly-silicon lines running vertically. The two mask layouts show a very strong symmetry, due to the fact that the NAND and the NOR gate have a symmetrical circuit topology.

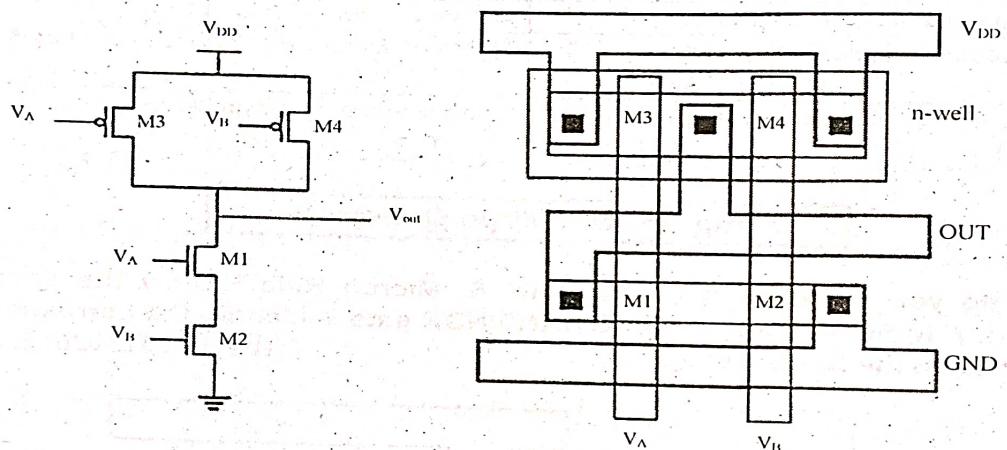


Fig: Sample layouts of a CMOS NOR2 gate and a CMOS NAND2 gate

2. a) Describe the Fick's law for Diffusion process. What do you mean by Isotropic & Anisotropic Etching processes? [WBUT 2013, 2015, 2017]  
 b) Describe the Photolithographic process for MOSFET fabrication.

- c) Describe the CMOS fabrication process with proper diagram. [WBUT 2013, 2015]  
 Answer:

a) **Fick's Law**

The basic diffusion process of impurity atoms is similar to that of charge carriers. Let  $F$  be the flux of dopant atoms traversing through a unit area in a unit time, and

$$F = -D \frac{\delta C}{\delta x}$$

where  $D$  is the diffusion coefficient,  $C$  is the dopant concentration, and  $x$  is the distance in one dimension. The equation imparts that the main driving force of the diffusion process is the concentration gradient,  $\delta C / \delta x$ . In fact, the flux is proportional to the concentration gradient, and the dopant atoms will diffuse from a high-concentration

region toward a low-concentration region. The negative sign on the right-hand-side of the equation states that matter flows in the direction of decreasing dopant concentration, that is, the concentration gradient is negative.

According to the law of conservation of matter, the change of the dopant concentration with time must be equivalent to the local decrease of the diffusion flux, in the absence of

a source or a sink. Thus  $\frac{\delta C}{\delta t} = -\frac{\delta F}{\delta x} = \frac{\delta}{\delta x} \left( D \frac{\delta C}{\delta x} \right)$

When the concentration of the dopant is low, the diffusion constant at a given temperature can be considered as a constant and above equation can be written as:

$$\frac{\delta C}{\delta t} = D \frac{\delta^2 C}{\delta x^2}$$

This equation is often referred as Fick's Second Law of diffusion.

Etching is selective removal of thin film resulting in a desired thin film pattern. Etching can be done either wet or dry environment. Measure of relative etch rates in different directions gives us two different types of etching processes. In isotropic etching, etch rates are same in all directions. It is usually related to chemical processes. Highly directional etching with different etch rates gives us anisotropic etching. It is usually related to physical processes such as ion bombardment and sputtering.

b) Photolithographic process for MOSFET fabrication & c) CMOS Fabrication process

Upto step 5, it is photolithographic process for MOS fabrication.

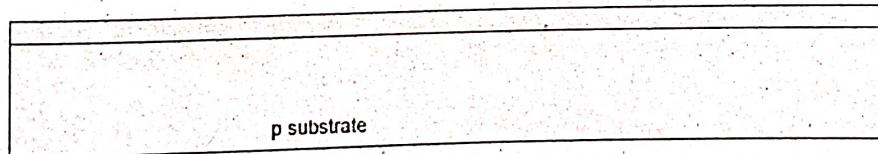
#### **Fabrication Steps (n-well process)**

**Step 1:** Start with p-type substrate



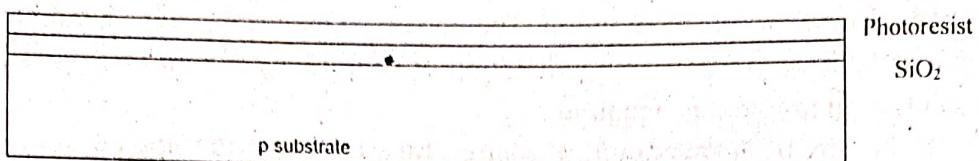
**Step 2:** Oxidation - Exposing to high-purity oxygen and hydrogen at approx. 1000°C in oxidation furnace.

$\text{SiO}_2$

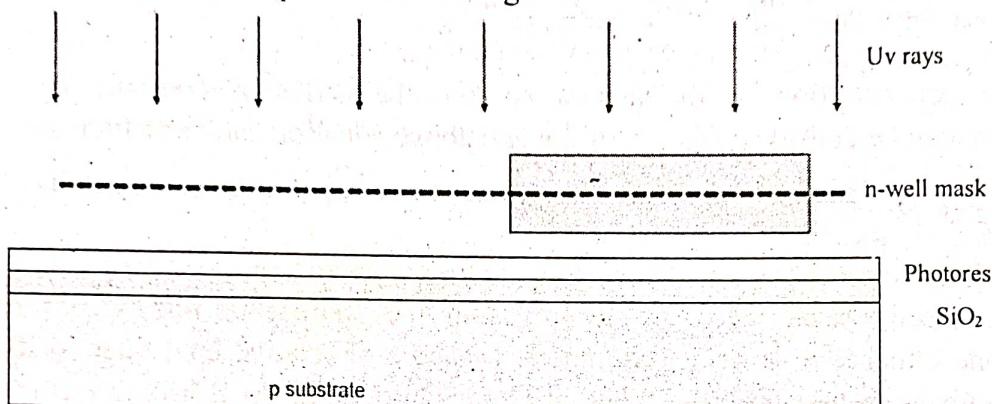


**Step 3:** Photoresist coating – The light sensitive organic polymer softens when exposed to light.

## POPULAR PUBLICATIONS



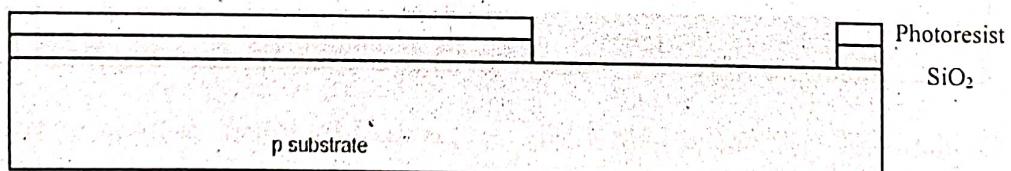
**Step 4:** Masking – Expose photoresist through n-well mask



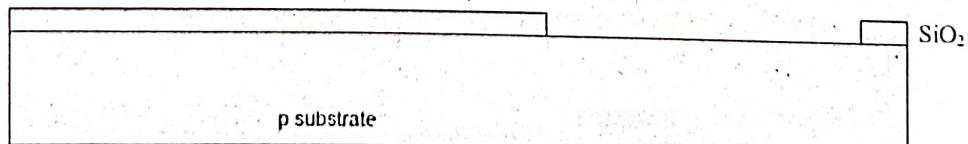
**Step 5:** Removal of photoresist – Photoresists are removed by treating the wafer with acidic or base solutions.



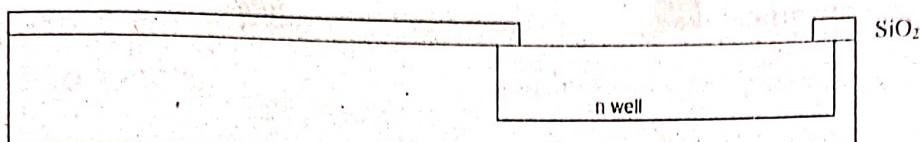
**Step 6:** Acid etching – SiO<sub>2</sub> is selectively removed from areas of wafer that are not covered by photoresist by using hydrofluoric acid.



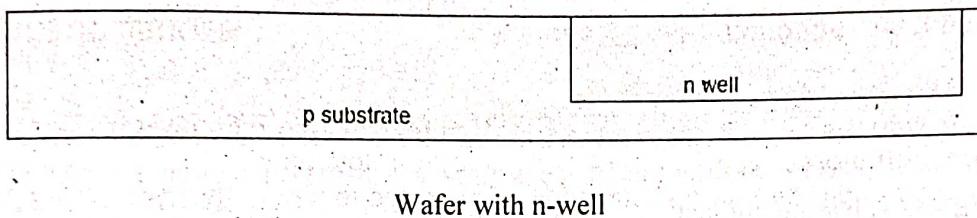
**Step 7:** Removal of photoresist – Strip off the remaining photoresist.



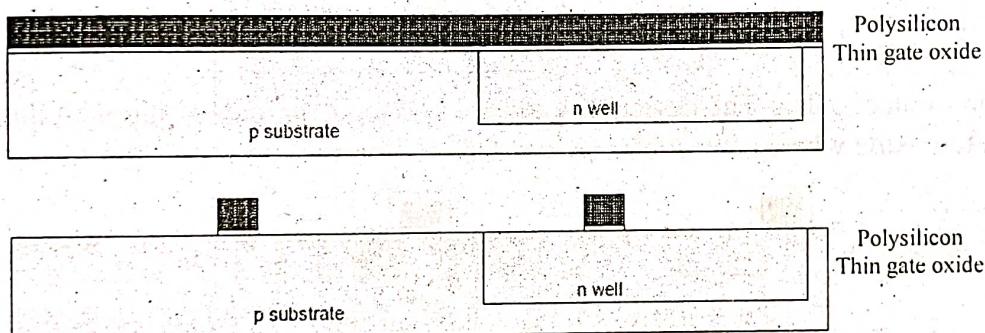
**Step 8:** Formation of n-well – n well is formed with diffusion or ion implantation.



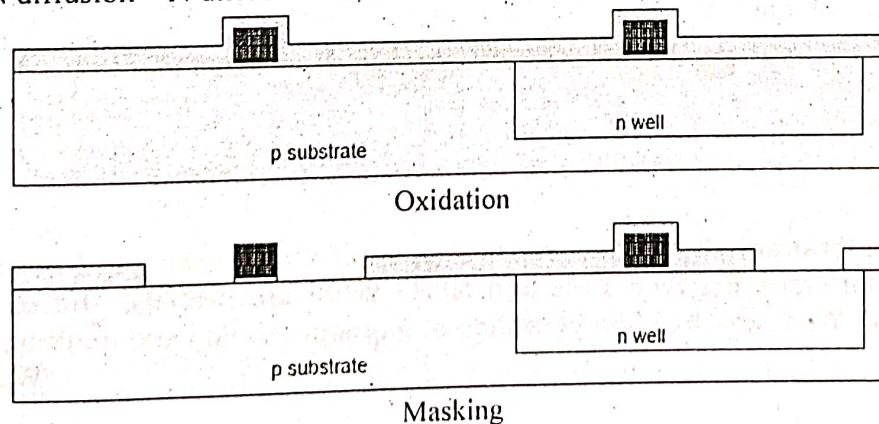
**Step 9:** Removal of SiO<sub>2</sub> – Strip off the remaining oxide using hydrofluoric acid.



**Step 10:** Polysilicon deposition – Deposit very thin layer of gate oxide and a polysilicon layer using Chemical Vapour Deposition (CVD) process.

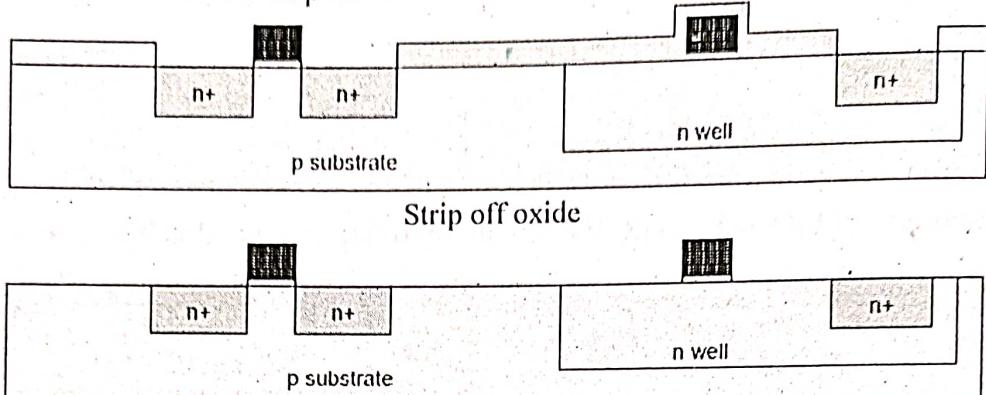


**Step 11:** N diffusion – N diffusion form n MOS source, drain and n well contact.

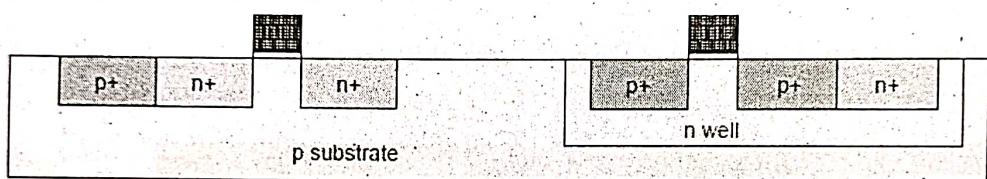


## POPULAR PUBLICATIONS

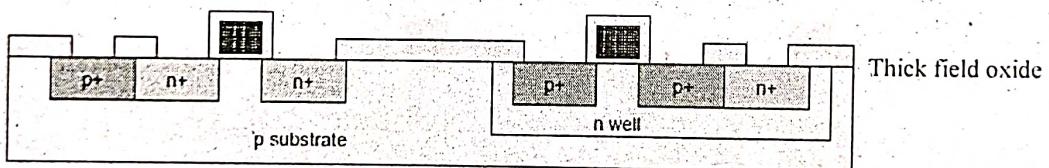
Dopants are diffused or ion implanted



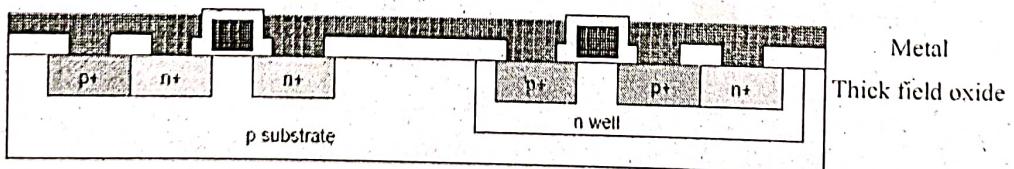
**Step 12:** P diffusion – Similar set of steps form p+ diffusion regions for pMOS source and drain and substrate contact.



**Step 13:** Contact cuts – The devices are to be wired together, cover chip with thick field oxide, etch oxide where contact cuts are needed.



**Step 14:** Metallization – Sputter on aluminium over whole wafer. Pattern to remove excess metal, leaving wires.



**3. What is lithography? Mention various types of lithography used in VLSI. What are the differences between PPR and NPR? What are prebake and postbake in lithography? What are the characteristics of exposure tools used in lithography?**

[WBUT 2014]

**Answer:**

**1<sup>st</sup> & 2<sup>nd</sup> Part:**

**Lithography:** The fabrication of an integrated circuit (IC) requires a variety of physical and chemical processes performed on a semiconductor (e.g., silicon) substrate. In general,

the various processes used to make an IC fall into three categories: film deposition, patterning, and semiconductor doping. Films of both conductors (such as polysilicon, aluminum, and more recently copper) and insulators (various forms of silicon dioxide, silicon nitride, and others) are used to connect and isolate transistors and their components. Selective doping of various regions of silicon allow the conductivity of the silicon to be changed with the application of voltage. By creating structures of these various components millions of transistors can be built and wired together to form the complex circuitry of a modern microelectronic device. Fundamental to all of these processes is lithography, i.e., the formation of three-dimensional relief images on the substrate for subsequent transfer of the pattern to the substrate.

In VLSI, lithography processes are used with positive and negative photoresists.

### **3<sup>rd</sup> Part:**

#### **Positive and Negative Photoresist**

There are two types of photoresist: positive and negative. For positive resists, the resist is exposed with UV light wherever the underlying material is to be removed. In these resists, exposure to the UV light changes the chemical structure of the resist so that it becomes more soluble in the developer. The exposed resist is then washed away by the developer solution, leaving windows of the bare underlying material. In other words, "whatever shows, goes." The mask, therefore, contains an exact copy of the pattern which is to remain on the wafer.

Negative resists behave in just the opposite manner. Exposure to the UV light causes the negative resist to become polymerized, and more difficult to dissolve. Therefore, the negative resist remains on the surface wherever it is exposed, and the developer solution removes only the unexposed portions. Masks used for negative photoresists, therefore, contain the inverse (or photographic "negative") of the pattern to be transferred. The figure below shows the pattern differences generated from the use of positive and negative resist.

Negative resists were popular in the early history of integrated circuit processing, but positive resist gradually became more widely used since they offer better process controllability for small geometry features. Positive resists are now the dominant type of resist used in VLSI fabrication processes.

## POPULAR PUBLICATIONS

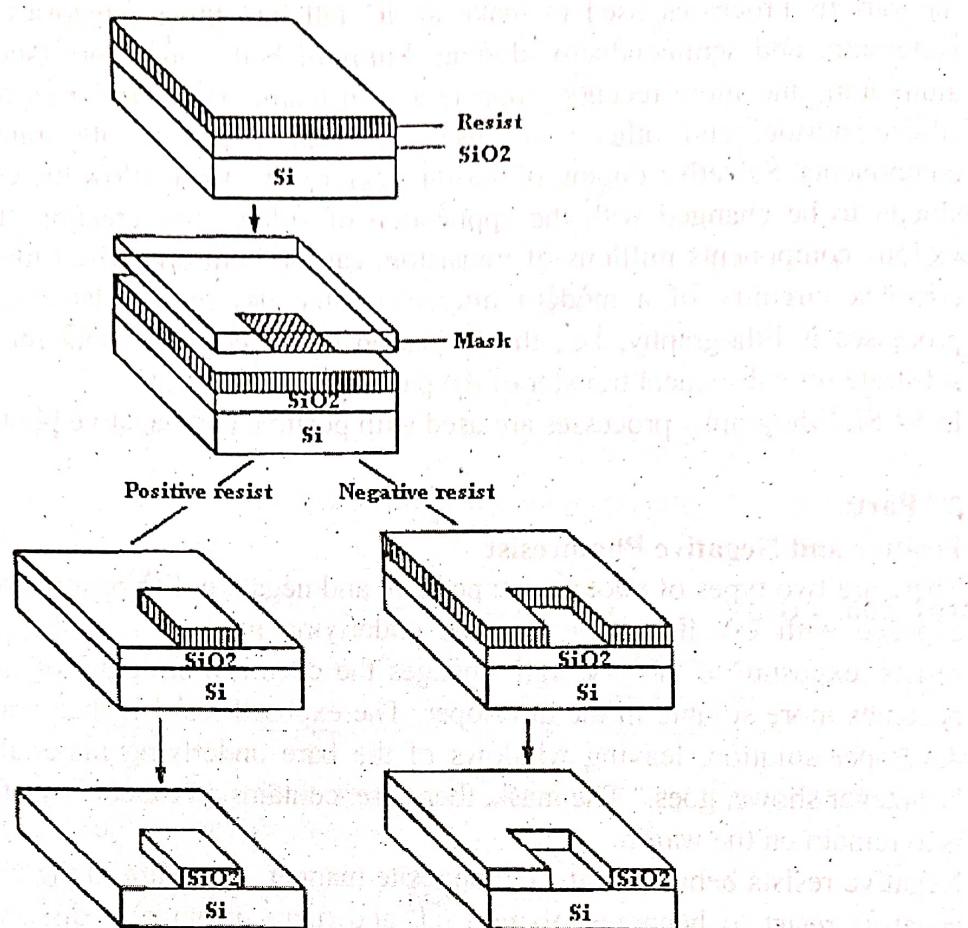


Fig.: Pattern structures using positive and negative photoresists

## **4<sup>th</sup> Part:**

**Pre bake & Post Bake:** In lithographic process, prebake (soft bake) is used to evaporate the coating solvent and to densify the resist after spin coating. Commercially, microwave oven or IR lamps are used for that. 900°C to 1000°C it is dried for 20 minutes. After that at 75°C to 85°C, it is placed on a hot plate for 45 seconds. Hot plating of resist is usually faster, more controllable, and does not trap solvents.

Post bake (hard bake) is used to stabilize and harden the developed photoresist prior to processing steps that the resist will mask. Post bake removes any remaining traces of the coating solvent or developer. This eliminates the solvent burst effects in vacuum processing. Postbake introduces some stress into the photoresist. Some shrinkage of the photoresist may occur. Longer or hotter postbake makes resist removal much more difficult. Firm post bake is needed for acid etching.

**5<sup>th</sup> Part:**

Characteristics of exposure tools used in photolithography are:

- Resolution
- Registration
- Throughput
- Near field and far field diffraction limits.
- Depth of focus
- Numerical aperture

4. Explain the following dominant CMOS fabrication process with neat diagrams.

- a) P-well process
- b) N-well process
- c) Twin tub process
- d) Silicon on insulator

[WBUT 2014]

**Answer:**

**a) P-well process:**

The common approach to p-well CMOS fabrication has been to start with a moderately doped n-type substrate (wafer), create the p-type well for a n-channel devices and build the p-channel transistor on substrate. This diffusion must be carried out with a special care since the p-well doping concentration and depth will affect the threshold voltages as well as the breakdown voltages of n-transistors. To achieve low threshold voltage (0.6 – 0.7V) we need either deep well diffusion or high well resistivity. However, deep wells require larger spacing between the n- and p-type transistors and wires because of lateral diffusion and therefore a larger chip area.

The p-wells act as substrates for the n-devices within the parent n-substrate and provided that voltage polarity restrictions are observed, the two areas are electrically isolated.

The masking, patterning and diffusion process is similar to nMOS fabrication. The typical processing steps are:

**Mask 1** – defines the areas in which the deep p-well diffusions are to take place.

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

**Mask 3** – used to pattern polysilicon layer which is deposited after the thin oxide.

**Mask 4** – A p-plus mask is now used (to be in effect ‘Anded’ with Mask 2) to define all areas where p-diffusion is to take place.

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

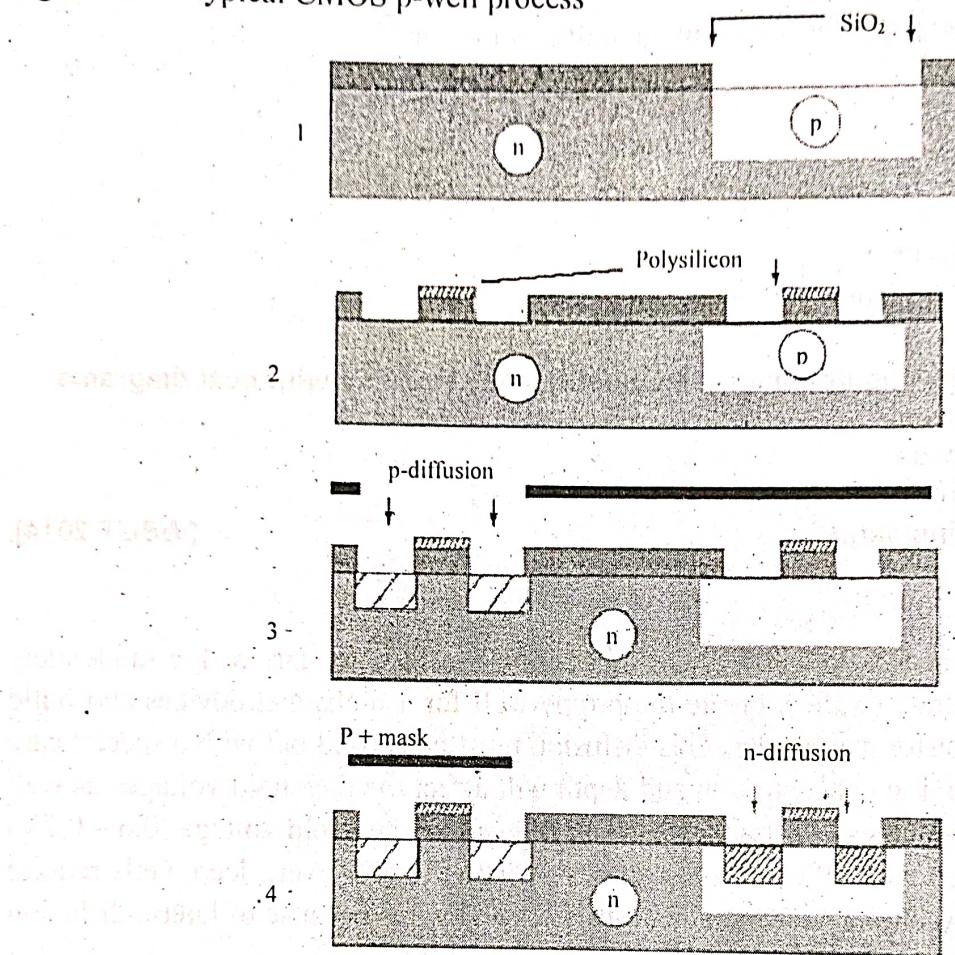
**Mask 6** – Contact cuts are now defined.

**Mask 7** – The metal layer pattern is defined by this mask.

**Mask 8** – An overall passivation (over glass) layer is now applied and Mask 8 is needed to define the openings for access to bonding pads.

## POPULAR PUBLICATIONS

Figure shows typical CMOS p-well process



### b) The n-well process:

N-well CMOS circuits are also superior to p-well because of the lower substrate bias effects on transistor threshold voltage and inherently lower parasitic capacitances associated with source and drain regions. Typical n-well fabrication steps are similar to p-well process except that n-well is used.

The typical processing steps are

- **Mask 1** – Defines the areas in which the deep n-well diffusions are to take place.
- **Mask 2** – Defines the nMOS and pMOS active areas.
- **Mask 3** – Defines the thinox regions, namely those areas where the thick oxide is to be stripped and thin oxide grown to accommodate p- and n- transistors and diffusion wires.
- **Mask 4** – Used to pattern polysilicon layer which is deposited after the thin oxide.
- **Mask 5** – A p-plus mask is now used to define all areas where p – diffusion is to take place.
- **Mask 6** – Defines those areas where n-type diffusion is to take place.

- **Mask 7** – Contact cuts are now defined.
- **Mask 8** – The metal layer pattern is defined by this mask.
- **Mask 9** – An overall passivation (over glass) layer is now applied and Mask 8 is needed to define the openings for access to bonding pads.

### c) Twin-Tub (Twin-Well) CMOS Process:

This technology provides the basis for separate optimization of the nMOS and pMOS transistors, thus making it possible for threshold voltage, body effect and the channel transconductance of both types of transistors to be tuned independently. Generally, the starting material is a n+ or p+ substrate, with a lightly doped epitaxial layer on top. This epitaxial layer provides the actual substrate on which the n-well and the p-well are formed. Since two independent doping steps are performed for the creation of the well regions, the dopant concentrations can be carefully optimized to produce the desired device characteristics.

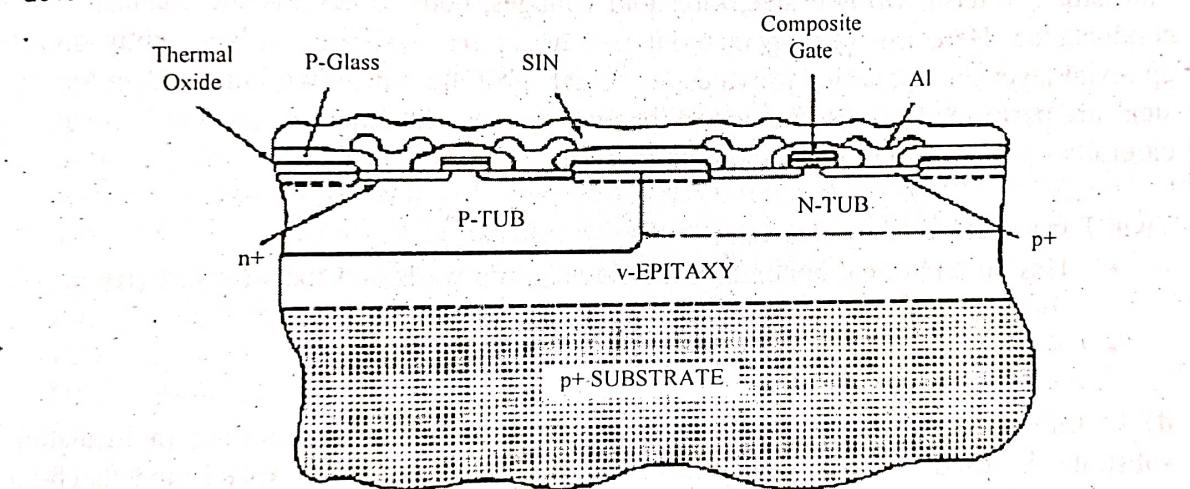


Fig. 1: Cross-section of nMOS and pMOS transistors in twin-tub CMOS process

In the conventional n-well CMOS process, the doping density of the well region is typically about one order of magnitude higher than the substrate, which, among other effects, results in unbalanced drain parasitics. The twin-tub process (Fig. 1) also avoids this problem.

#### *Advantages of Twin-tub CMOS process:*

1. Provide separate optimization of the n-type and p-type transistors.
2. Make it possible to optimize "V<sub>t</sub>", "Body effect", and the "Gain" of n, p devices, independently.
3. Steps:
  - A. Starting material: an n+ or p+ substrate with lightly doped -> "epitaxial" or "epi" layer -> to protect "latch up"

## POPULAR PUBLICATIONS

### B. Epitaxy"

- Grow high-purity silicon layers of controlled thickness
- With accurately determined dopant concentrations
- Electrical properties are determined by the dopant and its concentration in Si

### C. Process sequence

- Tub formation
- Thin-Oxide construction
- Source & drain implantations
- Contact cut definition
- Metallization

In twin-tub process, balanced performance of n and p devices can be constructed.

This process gives the advantage of separate optimization of the n-MOS and p-MOS transistors, in terms of their size, threshold voltages, body effect and the channel transconductance. Here the starting material is a n+ or p+ substrate, with a lightly doped epitaxial layer on top, which provides the actual substrate. Since two independent doping steps are performed for the creation of the well regions, the dopant concentrations can be carefully optimized to produce the desired device characteristics.

Twin Tub structure:

- Has an additional epitaxial layer between n/p wells and the substrate (n+ or p+) for protection against latch-up.
- Similar to p-well process (with both p-well and n-well).

d) In this process, rather than using silicon as the substrate material, an insulating substrate is used to improve process characteristics such as speed and latch-up susceptibility. The SOI CMOS technology allows the creation of independent, completely isolated nMOS and pMOS transistors virtually side-by-side on an insulating substrate (for example: sapphire). The main advantages of this technology are the higher integration density (because of the absence of well regions), complete avoidance of the latch-up problem, and lower parasitic capacitances compared to the conventional n-well or twin-tub CMOS processes. A cross-section of nMOS and pMOS devices in created using SOI process is shown in Fig.

The SOI CMOS process is considerably more costly than the standard n-well CMOS process. Yet the improvements of device performance and the absence of latch-up problems can justify its use, especially for deep-sub-micron devices.

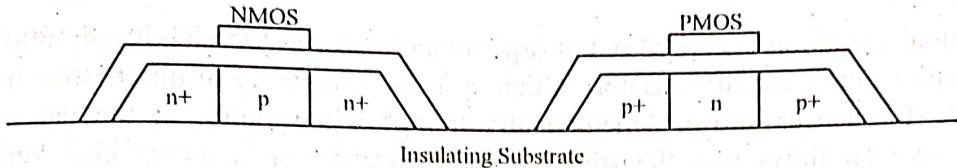


Fig: Cross-section of nMOS and pMOS transistors in SOI CMOS process

**5. What is oxidation? With a diagram, explain silicon thermal oxidation process. Determine the ratio of silicon consumed to the thickness of grown  $\text{SiO}_2$  layer over the wafer. If a  $\text{SiO}_2$  layer of  $1200 \text{ \AA}$  is to be grown, what would be the thickness of used silicon? Given, molecular weight of  $\text{SiO}_2 = 61.09 \text{ g/mol}$ , density of  $\text{SiO}_2 = 2.25 \text{ g/cc}$ , atomic weight of Si =  $28.12 \text{ g/mol}$ , density of Si =  $2.45 \text{ g/cc}$ .**

[WBUT 2014]

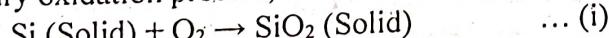
**Answer:****1<sup>st</sup> Part:**

**Oxidation:** Oxidation involves the growth of a layer of silicon dioxide on a silicon wafer. The oxide layer plays several important roles in both BJT and MOSFET technology as follows: (1) thermal oxide of silicon is used as a masking layer during diffusion and ion implantation of impurities into silicon to form junctions in selected portions of the silicon wafer. (2) It also serves as a junction passivant layer, so that the junction formed underneath the oxide is always protected against contamination from outside world. (3) Thermal oxide of silicon is chosen as the dielectric layer for the MOS portion of the MOSFET, which is the heart of the present day digital integrated circuits. Precise control of the thickness of the oxide layer is of great importance in silicon planar technology. In addition, the oxidation process plays a key role in determining the electrical characteristics of the junctions and MOSFETs.

**2<sup>nd</sup> Part:**

Thermal oxidation of silicon is carried out with two approaches: (1) dry oxidation and (2) wet oxidation.

In the dry oxidation process, the oxide is grown by the reaction of oxygen with silicon,



In the wet oxidation process, the oxide is formed by the reaction of water vapour with silicon:



In both the processes, silicon is consumed to convert it to the oxide. From the densities and molecular weights it can be shown that in the growth of an oxide layer of thickness,  $t_{ox}$ , a silicon layer of thickness  $0.44t_{ox}$  is consumed. Thus the thickness of the oxide layer formed is 2.27 times the thickness of consumed silicon.

## POPULAR PUBLICATIONS

The thermal oxidation is carried out at high temperatures in the vicinity of  $1000^{\circ}\text{C}$ . This is achieved in the oxidation reactor, which is usually made up of quartz tube heated by resistance furnace as shown schematically in following figure. In the dry oxidation process, oxygen flows past the silicon wafers mounted on a quartz boat kept in the reactor. Whereas in the wet oxidation process, water vapours are carried by nitrogen gas bubbled through de-ionized water heated to close to  $100^{\circ}\text{C}$ . Gas flow velocities are usually about  $1\text{ cm/sec}$  which ensures that the flow is laminar.

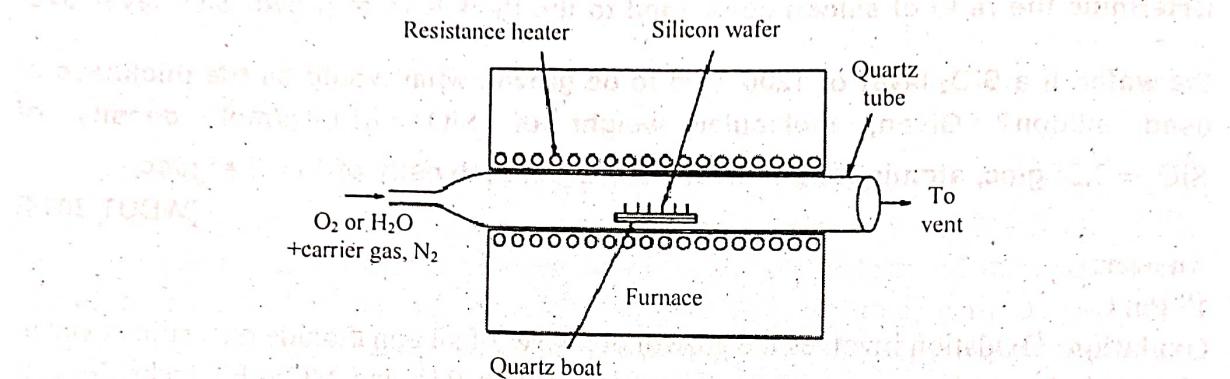


Fig: A schematic diagram showing the oxidation furnace

The oxidation process initially proceeds by the direct reaction of the oxidizing species with silicon. Once a few layers of oxide has grown, further oxidation takes at the oxide-silicon interface by the oxidizing species reaching the interface by diffusion through the growing oxide layer. This is governed by the diffusivity and solid solubility of the oxidizing species ( $\text{O}_2$  in dry oxidation process and  $\text{H}_2\text{O}$  in wet oxidation process) in the oxide. The oxide growth rate is higher in wet oxidation than in the dry oxidation process mainly due to the high solid solubility of water molecules in the oxide. Hence it is used in most planar processes for field oxides and as masking layers for diffusion of dopant impurities. As the controllability of thin oxide is achieved easily with the dry oxidation process, it is invariably used for growing gate oxides for MOSFET.

### 3<sup>rd</sup> Part:

$$\text{Vol. of 1 mole of } \text{SiO}_2 = \frac{\text{molecular weight}}{\text{density}} = \frac{61.09}{2.25} = 27.15$$

$$\text{Vol. of 1 mole of Si} = \frac{28.12}{2.45} = 11.477$$

$$\frac{\text{Vol. of 1 mole of Si}}{\text{Vol. of 1 mole of } \text{SiO}_2} = \frac{\text{Thickness of Si} \times \text{area}}{\text{Thickness of } \text{SiO}_2 \times \text{area}}$$

$$\text{Area of } \text{SiO}_2 = \text{area of Si}$$

$$\text{So, } \frac{11.477}{27.15} = \frac{\text{thickness of Si layer}}{1200},$$

$$\text{thickness of Si layer} = \frac{11.477}{27.15} \times 1200 = 507.276$$

$$\text{Ratio of thickness of Si and } \text{SiO}_2 = \frac{507.276}{1200} = 0.4227$$

6. a) With suitable diagram briefly describe the p-well fabrication process of a CMOS inverter. [WBUT 2016]

Answer:

Refer to Question No. 4(a) of Long Answer Type Questions.

b) What is micron rule?

[WBUT 2016]

Answer:

Micron rule is a design rule used in MOS based design. The features are:

- All minimum sizes and spacing specified in microns
- Rules do not have to be multiples of  $\lambda$ .
- Can result in 50% reduction in area over  $\lambda$  based rules.
- Standard in industry.

7. a) Distinguish between diffusion and ion-implantation technique. [WBUT 2016]

Answer:

Diffusion and ion implantation are two methods of introducing impurities to semiconductors to control the majority type of the carrier and the resistivity of layers. In diffusion, dopant atoms move from surface into Silicon by means of the concentration gradient. It is via substitution or interstitial diffusion mechanisms. In ion implantation, dopant atoms are added forcefully into Silicon by injecting an energetic ion beam. Diffusion is a high-temperature process while ion implantation is a low-temperature process. Dopant concentration and the junction depth can be controlled in ion implantation, but it cannot be controlled in the diffusion process. Diffusion has an isotropic dopant profile whereas ion implantation has an anisotropic dopant profile.

- **Advantages:** Diffusion creates no damage and batch fabrication is also possible. Ion implantation is a low-temperature process. It allows you to control the precise dose and the depth. Ion implantation is also possible through the thin layers of oxides and nitrides. It also includes short process times.
- **Disadvantages:** Diffusion is limited to solid solubility and it is a high-temperature process. Shallow junctions and low dosages are difficult the process of diffusion. Ion implantation involves an additional cost for annealing process.
- Diffusion has an isotropic dopant profile whereas ion implantation has an anisotropic dopant profile.

## POPULAR PUBLICATIONS

b) What is photolithography? Explain.

[WBUT 2016]

Answer:

Photolithography is the process of transferring geometric shapes on a mask to the surface of a silicon wafer. The steps involved in the photolithographic process are wafer cleaning; barrier layer formation; photo-resist application; soft baking; mask alignment; exposure and development; and hard-baking. Photolithography, also termed optical lithography or UV lithography, is a process used in micro fabrication to pattern parts of a thin film or the bulk of a substrate. It uses light to transfer a geometric pattern from a photo mask to a light sensitive chemical "photoresist", or simply "resist," on the substrate. A series of chemical treatment then either engraves the exposure pattern into, or enables deposition of a new material in the desired pattern upon, the material underneath the photo resist.

Photolithography shares some fundamental principles with photography in that the pattern in the etching resist is created by exposing it to light, either directly (without using a mask) or with a projected image using an optical mask. This procedure is comparable to a high precision version of the method used to make printed circuit boards. Subsequent stages in the process have more in common with etching than with lithographic printing. It is used because it can create extremely small patterns (down to a few tens of nanometers in size), it affords exact control over the shape and size of the objects it creates, and because it can create patterns over an entire surface cost-effectively. Its main disadvantages are that it requires a flat substrate to start with, it is not very effective at creating shapes that are not flat, and it can require extremely clean operating conditions.

A single iteration of photolithography combines several steps in sequence. Modern clean rooms use automated, robotic wafer track systems to coordinate the process. The procedure described here omits some advanced treatments, such as thinning agents or edge-bead removal.

### ***Preparation***

The wafer is initially heated to a temperature sufficient to drive off any moisture that may be present on the wafer surface. Wafers that have been in storage must be chemically cleaned to remove contamination. A liquid or gaseous "adhesion promoter", such as Bis(trimethylsilyl) amine ('hexamethyldisilazane', HMDS), is applied to promote adhesion of the photoresist to the wafer. The surface layer of silicon dioxide on the wafer reacts with HMDS to form tri-methylated silicon-dioxide, a highly water repellent layer not unlike the layer of wax on a car's paint. This water repellent layer prevents the aqueous developer from penetrating between the photoresist layer and the wafer's surface, thus preventing so-called lifting of small photoresist structures in the (developing) pattern.

### ***Wafer Cleaning, Barrier Formation and Photoresist Application***

In the first step, the wafers are chemically cleaned to remove particulate matter on the surface as well as any traces of organic, ionic, and metallic impurities. After cleaning, silicon dioxide, which serves as a barrier layer, is deposited on the surface of the wafer. After the formation of the  $\text{SiO}_2$  layer, photoresist is applied to the surface of the wafer. High-speed centrifugal whirling of silicon wafers is the standard method for applying photoresist coatings in IC manufacturing. This technique, known as "Spin Coating," produces a thin uniform layer of photoresist on the wafer surface.

### ***Positive and Negative Photoresist***

There are two types of photoresist: positive and negative. For positive resists, the resist is exposed with UV light wherever the underlying material is to be removed. In these resists, exposure to the UV light changes the chemical structure of the resist so that it becomes more soluble in the developer. The exposed resist is then washed away by the developer solution, leaving windows of the bare underlying material. In other words, "whatever shows, goes." The mask, therefore, contains an exact copy of the pattern which is to remain on the wafer.

Negative resists behave in just the opposite manner. Exposure to the UV light causes the negative resist to become polymerized, and more difficult to dissolve. Therefore, the negative resist remains on the surface wherever it is exposed, and the developer solution removes only the unexposed portions. Masks used for negative photoresists, therefore, contain the inverse (or photographic "negative") of the pattern to be transferred. The figure below shows the pattern differences generated from the use of positive and negative resist.

## POPULAR PUBLICATIONS

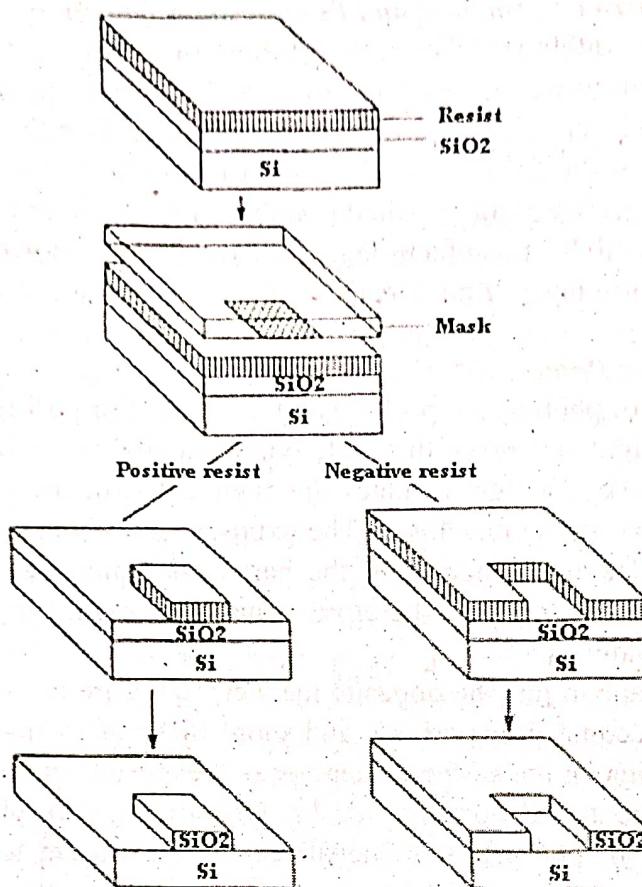


Fig. 1: Pattern structures using positive and negative photoresists

Negative resists were popular in the early history of integrated circuit processing, but positive resist gradually became more widely used since they offer better process controllability for small geometry features. Positive resists are now the dominant type of resist used in VLSI fabrication processes.

c) What is meant by Molecular Beam Epitaxy (MBE)?

[WBUT 2016]

**Answer:**

Molecular Beam Epitaxy is a non-CVD epitaxial process that uses an evaporation method. In MBE, a source material is heated to produce an evaporated beam of particles. These particles travel through a very high vacuum ( $10^{-8}$  Pa; practically free space) to the substrate, where they condense. MBE has lower throughput than other forms of epitaxy. This technique is widely used for growing III-V semiconductor crystals.

- In MBE, 'molecular beams' of semiconductor material are deposited onto a heated crystalline substrate to form thin epitaxial layers.
- Deposited films are crystalline
- Special thermal evaporation or sputter tool required
- Growth rates = a few  $\frac{\text{A}^\circ}{\text{s}}$

- Substrate temperatures can exceed 900°C

**8. What are the different types of lithography process? Describe photolithography with diagram.**

**Answer:**

i) In VLSI technology, two types of lithography processes are used. Microlithography and nanolithography refer specifically to lithographic patterning methods capable of structuring material on a fine scale. Typically, features smaller than 10 micrometers are considered microlithographic, and features smaller than 100 nanometers are considered nanolithography. Photo lithography is one of these methods, often applied to semiconductor manufacturing of microchips. Photolithography is also commonly used for fabricating Micro Electromechanical Systems (MEMS) devices. Photolithography generally uses a pre-fabricated photomask or reticle as a master from which the final pattern is derived.

ii) Photolithography is the process of transferring geometric shapes on a mask to the surface of a silicon wafer. The steps involved in the photolithographic process are wafer cleaning; barrier layer formation; photo-resist application; soft baking; mask alignment; exposure and development; and hard-baking. **Photolithography**, also termed **optical lithography** or **UV lithography**, is a process used in micro fabrication to pattern parts of a thin film or the bulk of a substrate. It uses light to transfer a geometric pattern from a photo mask to a light sensitive chemical "photoresist", or simply "resist," on the substrate. A series of chemical treatment then either engraves the exposure pattern into or enables deposition of a new material in the desired pattern upon, the material underneath the photo resist.

iii)

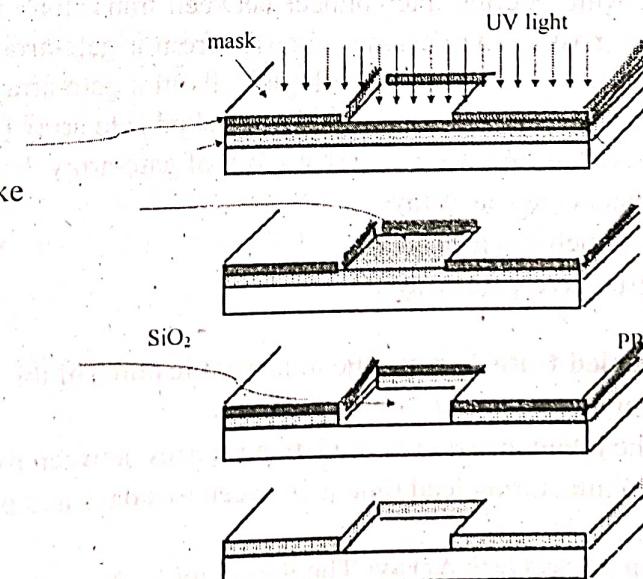
1. Oxidation
2. Photoresist (PR) coating
3. Stepper exposure
4. Photoresist development and bake
5. Acid etching

    Unexposed (negative PR)  
    Exposed (positive PR)

6. Processing step

    Ion implantation  
    Plasma etching  
    Metal deposition

8. Photoresist removal (ashing)

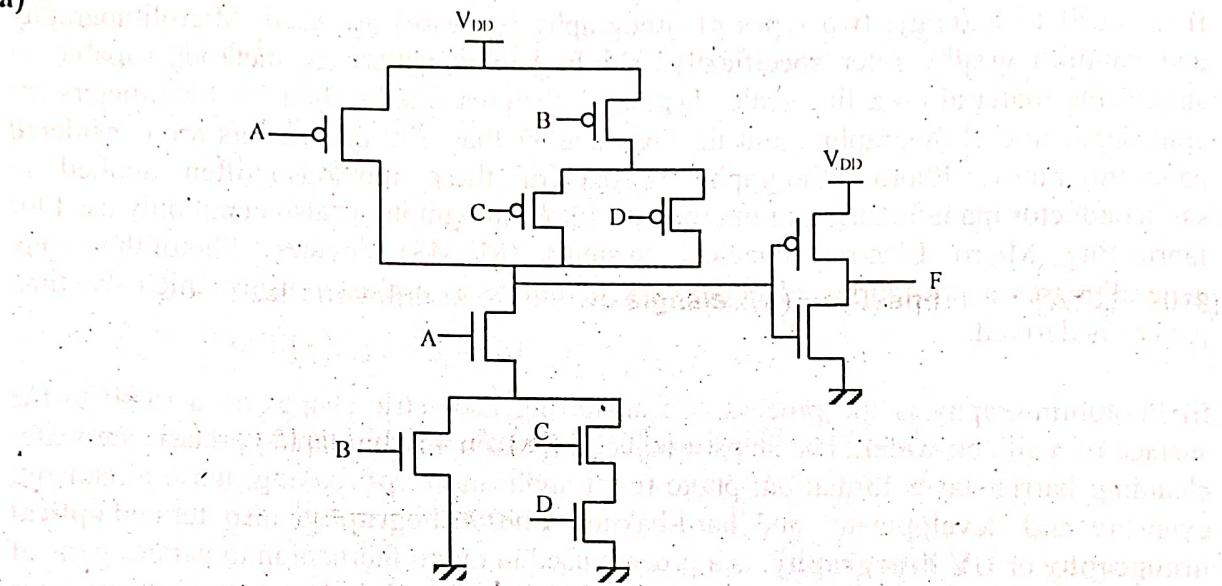


## POPULAR PUBLICATIONS

9. a) Implement the Boolean function  $F = [A(B+CD)]$  using CMOS logic. [WBUT 2018]  
b) Explain the gate array based VLSI system design.  
c) Discuss the layout design rules.

**Answer:**

a)



b) In a **gate array** (sometimes abbreviated to GA) or gate-array-based ASIC the transistors are predefined on the silicon wafer. The predefined pattern of transistors on a gate array is the **base array**, and the smallest element that is replicated to make the base array is the **base cell** (sometimes called a **primitive cell**). Only the top few layers of metal, which define interconnect between transistors, are defined by the designer using custom masks. The designer chooses from a gate-array library of predesigned and pre-characterized logic cells. The logic cells in a gate-array library are often called **macros**. That is why they are often called as masked gate array (MGA). There are the following different types of gate-array-based ASICs:

- Channeled gate arrays.
- Channel-less gate arrays.
- Structured gate arrays.

**Channeled Gate Array:** The important features of this type of MGA are:

- Only interconnects are customized.
- The interconnect uses predefined spaces between rows of base cells.
- Manufacturing lead time is between two days and two weeks.

**Channel-less Gate Array:** The important features of this type of MGA are as follows:

- Only some (the top few) mask layers are customized—interconnect.
- Manufacturing lead time is between two days and two weeks.

The key difference between a channel-less gate array and channeled gate array is that there are no predefined areas set aside for routing between cells on a channel-less gate array. Instead we route over the top of the gate-array devices. When we use an area of transistors for routing in a channel-less array, we do not make any contacts to the devices lying underneath; we simply leave the transistors unused.

**Structured Gate Array:** An **embedded gate array** or **structured gate array** (also known as **master slice** or **master image**) combines some of the features of CBICs and MGAs.

The important features of this type of MGA are the following:

- Only interconnects are customized.
- Custom blocks (the same for each design) can be embedded.
- Manufacturing lead time is between two days and two weeks.

**c) Layout Design Rules:**

The physical mask layout of any circuit to be manufactured using a particular process must conform to a set of geometric constraints or rules, which are generally called layout design rules. These rules usually specify the minimum allowable line widths for physical objects on-chip such as metal and poly-silicon interconnects or diffusion areas, minimum feature dimensions, and minimum allowable separations between two such features. The main objective of design rules is to achieve a high overall yield and reliability while using the smallest possible silicon area, for any circuit to be manufactured with a particular process.

The design rules are usually described in two ways:

- **Micron rules**, in which the layout constraints such as minimum feature sizes and minimum allowable feature separations, are stated in terms of absolute dimensions in micrometers, or,
- **Lambda rules**, which specify the layout constraints in terms of a single parameter ( $\lambda$ ) and, thus, allow linear, proportional scaling of all geometrical constraints

**10. Write short notes on the following:**

- a) VLSI interconnects [WBUT 2014]
- b) LIGA process [WBUT 2014]
- c) Clean room [WBUT 2014]
- d) High K dielectric materials for IC fabrication [WBUT 2014]
- e) CVD Technique [WBUT 2015]
- f) Design Rule Checker (DRC) [WBUT 2015]
- g) Ion implantation process for MOSFET fabrication [WBUT 2015, 2016]
- h) Czochralski technique for crystal growth [WBUT 2016]
- i) Comparator [WBUT 2017]

## POPULAR PUBLICATIONS

### j) Basic Steps of Fabrication Process

[WBUT 2018]

**Answer:**

#### a) VLSI interconnects:

In high density CMOS processes, several metal layers are used for creating the interconnections between the transistors and for routing the power supply, signal and clock lines on the chip surface. These layers are isolated from neighbouring layers by a dielectric, allows higher integration densities for realization of complex structures. The electrical connections between layers are made by via that are placed wherever needed. Local signal connections are preferably made with metal lines as much as possible and metal-polysilicon contacts are used to provide the electrical connection between the two layers, wherever necessary.

The final step in the mask layout is the local interconnections in metal for the output node,  $V_{DD}$  and GND contacts. The dimensions of metal lines in a mask layout are usually dictated by the minimum metal width and the minimum metal separation.

The conventional delay estimation approaches seek to classify three main components of the output load. They are: i) internal parasitic capacitance of the transistors, ii) interconnect capacitances, and iii) input capacitances of the fan-out gates. Of these three components, the load conditions imposed by the interconnection lines present serious problems, especially in sub-micron circuits.

#### b) LIGA process:

LIGA is a German acronym for *Lithographie, Galvanoformung, Abformung* (Lithography, Electroplating, and Molding) that describes a fabrication technology used to create high-aspect-ratio microstructures.

The LIGA consists of three main processing steps; lithography, electroplating and molding. There are two main LIGA-fabrication technologies, **X-Ray LIGA**, which uses X-rays produced by a synchrotron to create high-aspect ratio structures, and **UV LIGA**, a more accessible method which uses ultraviolet light to create structures with relatively low aspect ratios.

LIGA was one of the first major techniques to allow on-demand manufacturing of high-aspect-ratio structures (structures that are much taller than wide) with lateral precision below one micrometer.

In the process, an X-ray sensitive polymer photoresist, typically PMMA, bonded to an electrically conductive substrate, is exposed to parallel beams of high-energy X-rays from a synchrotron radiation source through a mask partly covered with a strong X-ray absorbing material. Chemical removal of exposed (or unexposed) photoresist results in a three-dimensional structure, which can be filled by the electrodeposition of metal. The resist is chemically stripped away to produce a metallic mold insert. The mold insert can be used to produce parts in polymers or ceramics through injection molding.

The LIGA technique's unique value is the precision obtained by the use of deep X-ray lithography (DXRL). The technique enables microstructures with high aspect ratios and

high precision to be fabricated in a variety of materials (metals, plastics, and ceramics). Many of its practitioners and users are associated with or are located close to synchrotron facilities.

**c) Clean room:**

All fabrication process steps are implemented in a clean room. Here the air is maintained at a well-controlled temperature and humidity level and is continuously filtered and recirculated. Particulates must be avoided on wafers since they cause improperly defined features, undesirable surface topography, leakage through insulating layers, and other deleterious effects. Air in the room is monitored and classified with respect to particulates. A "class 100" environment has a maximum of 100 particles per cubic foot with particle size larger than  $0.5\mu\text{m}$ , and a maximum of 10 particles per cubic feet with particle size larger than  $5.0\mu\text{m}$ .

The filtered air flows from ceiling to floor at more than 85 linear feet per minute. Particulates can emanate from process equipment as well as from humans; personnel must wear proper clothing to protect the wafers.

**d) High K dielectric materials for IC fabrication:**

Scaling down of MOSFET gate length require scaling of gate insulator thickness. To attain large current drive in a MOSFET, thickness of the dielectric should be small which in turn increase quantum mechanical tunnelling.

To improve the performance of the device, high K dielectrics (such as Hafnium Oxide  $\text{HfO}_2$ ) are used in place of  $\text{SiO}_2$ . High K gate insulators offer two main benefits: it can provide a sub- $10\text{\AA}$  equivalent oxide thickness. It improves the drive current while significantly reducing the gate leakage current as compared to  $\text{SiO}_2$ .

Materials with high dielectric constant show a tendency to be associated with a lower bandgap, which reduces the effectiveness of these materials in suppressing gate leakage current.

**e) CVD Technique:**

**Chemical vapor deposition (CVD)** is a chemical process used to produce high quality, high-performance, solid materials. The process is often used in the semiconductor industry to produce thin films. In typical CVD, the wafer (substrate) is exposed to one or more volatile precursors, which react and/or decompose on the substrate surface to produce the desired deposit. Frequently, volatile by-products are also produced, which are removed by gas flow through the reaction chamber.

Microfabrication processes widely use CVD to deposit materials in various forms, including: monocrystalline, polycrystalline, amorphous, and epitaxial. These materials include: silicon (oxide, carbide, nitride, oxy-nitride), carbon (fiber, nanofibers, nanotubes,

## POPULAR PUBLICATIONS

graphene) fluorocarbons, filaments (tungsten, titanium nitride) and various high-K dielectrics.

### **Types of CVD Processes:**

CVD is practiced in a variety of formats. These processes generally differ in the means by which chemical reactions are initiated.

#### **Classified by operating pressure:**

- a) Atmospheric pressure CVD (APCVD) – CVD at atmospheric pressure.
- b) Low-pressure CVD (LPCVD) – CVD at sub-atmospheric pressures. Reduced pressures tend to reduce unwanted gas-phase reactions and improve film uniformity across the wafer.
- c) Ultrahigh vacuum CVD (UHVCVD) – CVD at very low pressure, typically below 10<sup>-6</sup> Pa (~10<sup>-8</sup> torr).

#### **Classified by physical characteristics of vapor:**

- a) Aerosol assisted CVD (AACVD) – CVD in which the precursors are transported to the substrate by means of a liquid/gas aerosol, which can be generated ultrasonically. This technique is suitable for use with non-volatile precursors.
- b) Direct liquid injection CVD (DLICVD) – CVD in which the precursors are in liquid form (liquid or solid dissolved in a convenient solvent). Liquid solutions are injected in a vaporization chamber towards injectors (typically car injectors). The precursor vapors are then transported to the substrate as in classical CVD. This technique is suitable for use on liquid or solid precursors. High growth rates can be reached using this technique.

#### **Plasma methods:**

- a) Microwave plasma-assisted CVD (MPCVD),
- b) Plasma Enhanced CVD (PECVD) – CVD that utilizes plasma to enhance chemical reaction rates of the precursors. PECVD processing allows deposition at lower temperatures, which is often critical in the manufacture of semiconductors. The lower temperatures also allow for the deposition of organic coatings, such as plasma polymers, that have been used for nanoparticle surface functionalization.
- c) Remote plasma-enhanced CVD (RPECVD) – Similar to PECVD except that the wafer substrate is not directly in the plasma discharge region. Removing the wafer from the plasma region allows processing temperatures down to room temperature.
- d) Atomic-layer CVD (ALCVD) – Deposits successive layers of different substances to produce layered crystalline films.
- e) Combustion Chemical Vapor Deposition (CCVD) – Combustion Chemical Vapor Deposition or flame pyrolysis is an open-atmosphere, flame-based technique for depositing high-quality thin films and nanomaterials.

f) **Design Rule Checker (DRC):**

Design Rules are a series of parameters provided by semiconductor manufacturers that enable the designer to verify the correctness of a mask set. Design rules are specific to a particular semiconductor manufacturing process. A design rule set specifies certain geometric and connectivity restrictions to ensure sufficient margins to account for variability in semiconductor manufacturing processes, so as to ensure that most of the parts work correctly.

The main objective of design rule checking (DRC) is to achieve a high overall yield and reliability for the design. If design rules are violated the design may not be functional. To meet this goal of improving die yields, DRC has evolved from simple measurement and Boolean checks, to more involved rules that modify existing features, insert new features, and check the entire design for process limitations such as layer density.

DRC products define rules in a language to describe the operations needed to be performed in DRC. For example, Mentor Graphics uses Standard Verification Rule Format (SVRF) language in their DRC rules files and Magma Design Automation is using Tcl-based language. A set of rules for a particular process is referred to as a run-set, rule deck, or just a deck. Usually DRC checks will be run on each sub-section of the ASIC to minimize the number of errors that are detected at the top level.

Some examples of DRC's in IC design include:

- Active to active spacing
- Well to well spacing
- Minimum channel length of the transistor
- Minimum metal width
- Metal to metal spacing
- Metal fill density (for processes using CMP)
- Poly density
- ESD and I/O rules
- Antenna effect

g) **Ion implantation process for MOSFET fabrication:**

Ion Implantation is an alternative to a deposition diffusion and is used to produce a shallow surface region of dopant atoms deposited into a silicon wafer. This technology has made significant roads into diffusion technology in several areas. In this process a beam of impurity ions is accelerated to kinetic energies in the range of several tens of kV and is directed to the surface of the silicon. As the impurity atoms enter the crystal, they give up their energy to the lattice in collisions and finally come to rest at some average penetration depth, called the projected range expressed in micro meters. Depending on the impurity and its implantation energy, the range in a given semiconductor may vary from a few hundred angstroms to about 1 micro meter. Typical distribution of impurity along the projected range is approximately Gaussian. By performing several

implantations at different energies, it is possible to synthesize a desired impurity distribution, for example a uniformly doped region.

### ***Ion Implantation System***

A typical ion-implantation system is shown in the figure below. A gas containing the desired impurity is ionized within the ion source. The ions are generated and repelled from their source in a diverging beam that is focused before it passes through a mass separator that directs only the ions of the desired species through a narrow aperture. A second lens focuses this resolved beam which then passes through an accelerator that brings the ions to their required energy before they strike the target and become implanted in the exposed areas of the silicon wafers. The accelerating voltages may be from 20 kV to as much as 250 kV. In some ion implanters, the mass separation occurs after the ions are accelerated to high energy. Because the ion beam is small, means are provided for scanning it uniformly across the wafers. For this purpose the focused ion beam is scanned electro statically over the surface of the wafer in the target chamber. Repetitive scanning in a raster pattern provides exceptionally uniform doping of the wafer surface. The target chamber commonly includes automatic wafer handling facilities to speed up the process of implanting many wafers per hour.

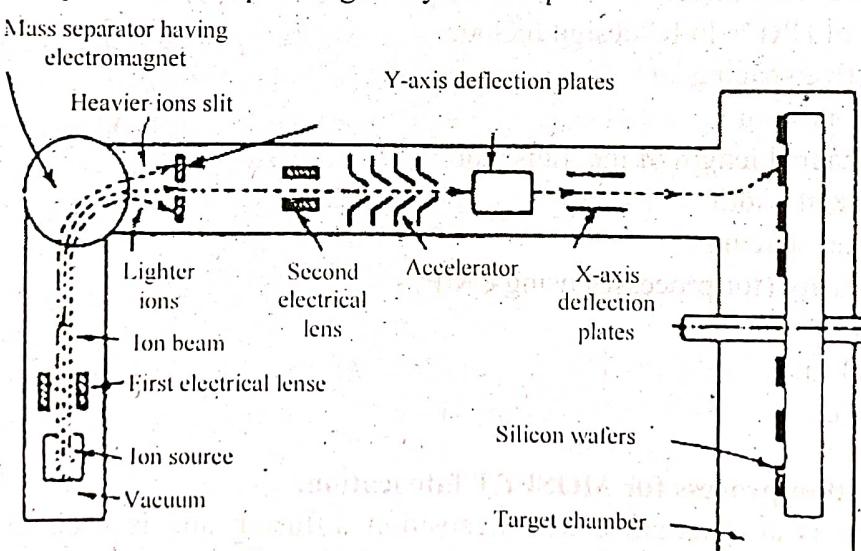


Fig: Ion implantation system

### ***h) Czochralski technique for crystal growth:***

A substantial amount of the silicon crystals for semiconductor industry are prepared by Czochralski technique.

The **Czochralski process** is a method of crystal growth used to obtain single crystal of semiconductors. The process is named after Polish scientist Jan Czochralski, who invented the method in 1916 while investigating the crystallization rates of metals. The most important application may be the growth of large cylindrical ingots, or boules, of single crystal silicon.

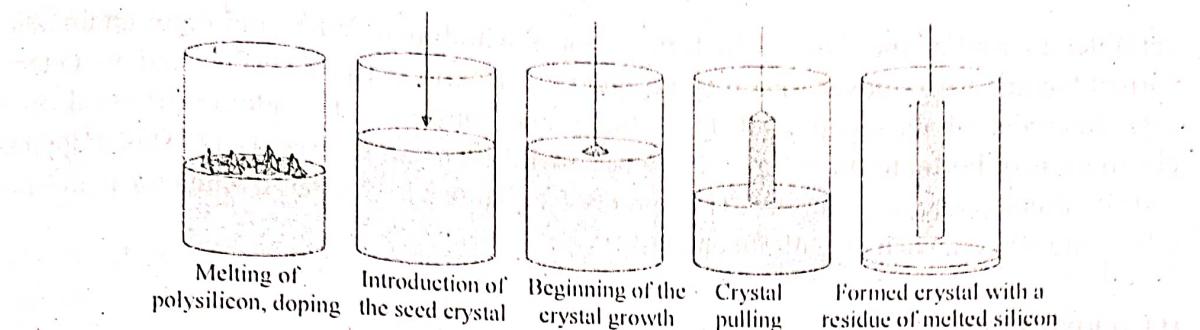


Fig: The Czochralski Process

High-purity, semiconductor-grade silicon (only a few parts per million of impurities) is melted in a crucible, usually made of quartz. Dopant impurity atoms such as boron or phosphorus can be added to the molten silicon in precise amounts to dope the silicon, thus changing it into p-type or n-type silicon, with different electronic properties. A precisely oriented rod-mounted seed crystal is dipped into the molten silicon. The seed crystal's rod is slowly pulled upwards and rotated simultaneously. By precisely controlling the temperature gradients, rate of pulling and speed of rotation, it is possible to extract a large, single-crystal, cylindrical ingot from the melt. Occurrence of unwanted instabilities in the melt can be avoided by investigating and visualizing the temperature and velocity fields during the crystal growth process. This process is normally performed in an inert atmosphere, such as argon, in an inert chamber, such as quartz.

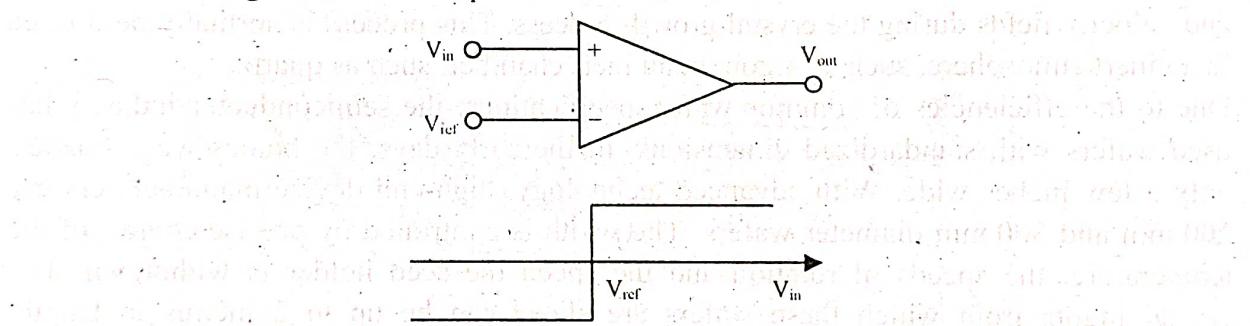
Due to the efficiencies of common wafer specifications, the semiconductor industry has used wafers with standardized dimensions. In the early days, the boules were smaller, only a few inches wide. With advanced technology, high-end device manufacturers use 200 mm and 300 mm diameter wafers. The width is controlled by precise control of the temperature, the speeds of rotation and the speed the seed holder is withdrawn. The crystal ingots from which these wafers are sliced can be up to 2 metres in length, weighing several hundred kilo-grams. Larger wafers allow improvements in manufacturing efficiency, as more chips can be fabricated on each wafer, so there has been a steady drive to increase silicon wafer sizes. The next step up, 450 mm, is currently scheduled for introduction in 2012. Silicon wafers are typically about 0.2–0.75 mm thick, and can be polished to great flatness for making integrated circuits or textured for making solar cells.

The process begins when the chamber is heated to approximately 1500 degrees Celsius, melting the silicon. When the silicon is fully melted, a small seed crystal mounted on the end of a rotating shaft is slowly lowered until it just dips below the surface of the molten silicon. The shaft rotates counter clockwise and the crucible rotates clockwise. The rotating rod is then drawn upwards very slowly, allowing a roughly cylindrical ingot to be formed. The ingot can be from one to two metres, depending on the amount of silicon in the crucible. The diameter of the silicon ingot is determined by the temperature

variables as well as the rate at which the ingot is withdrawn. When the ingot attains the correct length, it is removed, and then ground to a uniform external surface and diameter. The electrical characteristics of the silicon are controlled by adding material like phosphorus or boron to the silicon before it is melted. The added material is called dopant and the process is called doping. This method is also used with semiconductor materials other than silicon, such as gallium arsenide.

### i) Comparator:

A comparator is the basic component mainly used in analog-to-digital converters. Ideally, it generates an output logic signal as an instant response to the sign of an analog input (voltage or current). Obviously, a real circuit doesn't achieve the ideal function. The most important limits are the finite sensitivity, the offset and the finite speed. The electrical function of a comparator is to generate an output voltage which value is high or low depending on whether the sign of the input is positive or negative. We can have two different types of input: voltage or current. In the former case the input voltage is measured with respect to a given reference level. Therefore, the comparator determines whether the amplitude of the input is higher or not than a reference. When the current is the input variable the comparator determines whether the input current is flowing in or out the input terminal.



A logic signal denotes the output. The amplitude of electrical representation of the high or the low state should match the convention used in the associated digital logic to clearly distinguish between a logic 1 and a logic 0. When a comparator is used in a sampled data system a clock controls the action of the circuit. The comparator provides the output with a given periodicity synchronous with the clock. Therefore, a given time interval is available to achieve the result. Often, the fast variation of the input signal and the defined speed of the circuit used determine the need to separate the two functions inherent to the comparison process: to "catch" the value of the input signal and to generate the logic output. A sampled data system favors this disjunction: the clock period can be divided into two (or more) phases: one completes the sampling of the input and the other transforms the result into the logic signal. The latter non-linear operation can take advantage of the use of a latch. A latch effects a

regenerative amplification of the input and, thanks to its positive feedback, preserves the achieved output.

### j) Basic Steps of Fabrication Process:

*Refer to Question No. 2(b) of Long Answer Type Questions.*

**P-well CMOS process:** The fabrication of p-well CMOS process is similar to n-well process except that p-wells acts as "substrate" for the n-devices within the parent n-substrate.

### Advantages of n-well process:

- n-well CMOS are superior to p-well process because of lower substrate bias effects on transistor threshold voltage.
- Lower parasitic capacitances associated with source and drain regions.
- Latch-up problem can be considerably reduced by using a lower resistive type epitaxial p-type substrate, though n-well process degrades the performance of poorly performing p-type transistor.

as bondon of egation well on n-dream wello

## CMOS LOGIC CIRCUITS

### **Multiple Choice Type Questions**

1. The threshold voltage of an enhancement transistor is [WBUT 2013]  
a) greater than 0 V      b) less than 0 V  
c) equal to 0 V      d) none of these

Answer: (a)

2. Noise margin for low voltage is defined as [WBUT 2013]  
a)  $NM_L = V_{IL} - V_{OL}$       b)  $NM_L = V_{IL} - V_{IH}$   
c)  $NM_L = V_{OIL} - V_{OL}$       d)  $NM_L = V_{IH} - V_{IL}$

Answer: (a)

3. In the VTC curve of an inverter, critical voltages are obtained where the shape of the curve ( $dV_{out}/dV_{in}$ ) is [WBUT 2013]  
a) 1      b) -1      c) 0      d) none of these

Answer: (b)

4. Slant in ( $I_D - V_{DS}$ ) occurs due to [WBUT 2013]  
a) body effect      b) velocity saturation  
c) channel length modulation      d) mobility degradation

Answer: (c)

5. The unit of  $\mu_n C_{OX}$  is [WBUT 2013]  
a)  $A/V^2$       b)  $V^{-1}$       c) ohm      d)  $(\text{ohm})^{-1}$

Answer: (a)

6. The model parameter LAMDA in MOS structure stands for [WBUT 2013]  
a) flicker noise      b) transit time  
c) channel length modulation      d) transconductance

Answer: (c)

7. For an n-channel MOSFET  $I_{D(SAT)} = 0.2mA$ ,  $V_{DS} = 5V$  and  $V_{th} = 0.6V$ , the Gate voltage is [WBUT 2013]  
a) 4.8 V      b) 5.6 V      c) 4.4V      d) 5V

Answer: (b)

8. The equivalent ( $W/L$ ) of two NMOS transistors with  $(W_1/L)$  and  $(W_2/L)$  connected in parallel is [WBUT 2013]

- a)  $(W_1/L) + (W_2/L)$
- b)  $(W_1/L) \times (W_2/L)$
- c)  $(W_1/L)/(W_2/L)$
- d) none of these

Answer: (a)

9. How many transistors are required to design function  $F = (A.B + C.D)$ ? [WBUT 2013]

- a) 4
- b) 6
- c) 8
- d) 10

Answer: (d)

10. The main advantage of pre-charge-evaluate dynamic logic is [WBUT 2013]

- a) lesser number of transistor required
- b) high speed
- c) low power consumption
- d) all of these

Answer: (d)

11. Which design is more efficient? [WBUT 2013]

- a) pull-up & pull-down design
- b) TG design
- c) per-charge & evaluate logic

Answer: (c)

12. Dynamic logic requires periodic clock signals in order to [WBUT 2013]

- a) improve performance
- b) synchronization
- c) increasing
- d) charge refreshing

Answer: (d)

13. The unit of trans-conductance parameter of MOSFET is [WBUT 2014]

- a)  $\mu A^2/V^2$
- b)  $\mu A/V^2$
- c)  $\mu A^2/V$
- d)  $\mu A/V$

Answer: (b)

14. VHDL is a [WBUT 2014, 2017]

- a) multithreaded program
- b) a programming language like C
- c) single user program
- d) sequential program

Answer: (a)

15. The body effect occurs due to potential difference between [WBUT 2014, 2017]

- a) source and body
- b) body and drain
- c) gate and body
- d) none of these

Answer: (a)

16. In channel length modulation, the drain current [WBUT 2014, 2017]

- a) increase
- b) decrease
- c) constant
- d) zero

## POPULAR PUBLICATIONS

Answer: (a)

17. PMOS are wider than NMOS transistor

- a) mobility of holes is less than electrons
- b) mobility of holes is greater than electrons

[WBUT 2014, 2017]

Answer: (a)

18. Disadvantage of Domino logic is that,

- a) only non inverting structure can be realized
- b) it suffers from charge sharing
- c) static inverter is required
- d) all of these

Answer: (d)

19. In which device has the highest carrier mobility?

- a) NMOS
- b) PMOS
- c) CMOS

[WBUT 2016]

Answer: (a)

20. Which device acts as good switch?

- a) NMOS
- b) PMOS
- c) CMOS

[WBUT 2016]

Answer: (c)

21. Which has highest noise margin?

- a) active load inverter
- b) resistive load inverter
- c) CMOS inverter

[WBUT 2016]

Answer: (c)

22. One of the disadvantage of pass transistor logic

- a) less number of transistor
- b) poor noise margin
- c) only NMOS are used
- d) none of these

[WBUT 2017]

Answer: (c)

23. Low power logic family is

- a) TTL
- b) CMOS

[WBUT 2017]

Answer: (b)

c) ECL

d) none of these

24. Switch logic is designed using

- a) Complementary switches
- b) Silicon plates
- c) Conductors
- d) Resistors

[WBUT 2018]

Answer: (a)

b) Non-volatile type

25. Memory of SRAM based FPGA is

- a) Volatile type

[WBUT 2018]

- b) Non-volatile type

c) Dynamic type

d) None of these

Answer: (a)

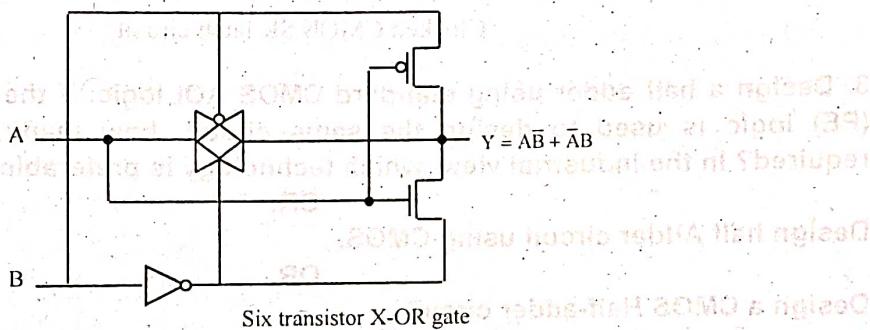
### Short Answer Type Questions

1. What are the advantages of TG logic design style? Explain with neat sketch the construction and operation of an XOR gate using TG design style. [WBUT 2013]

Answer:

*Advantages of using TG logic design styles are:*

- i) It has the advantage of being simple and fast: Complex gates are implemented with the minimum number of transistors (the reduced parasitic capacitance results in fast circuits)
- ii) Transmission gate has better noise margin than Pass transistor.



2. Explain the operation of clocked CMOS S-R latch circuit. [WBUT 2013]

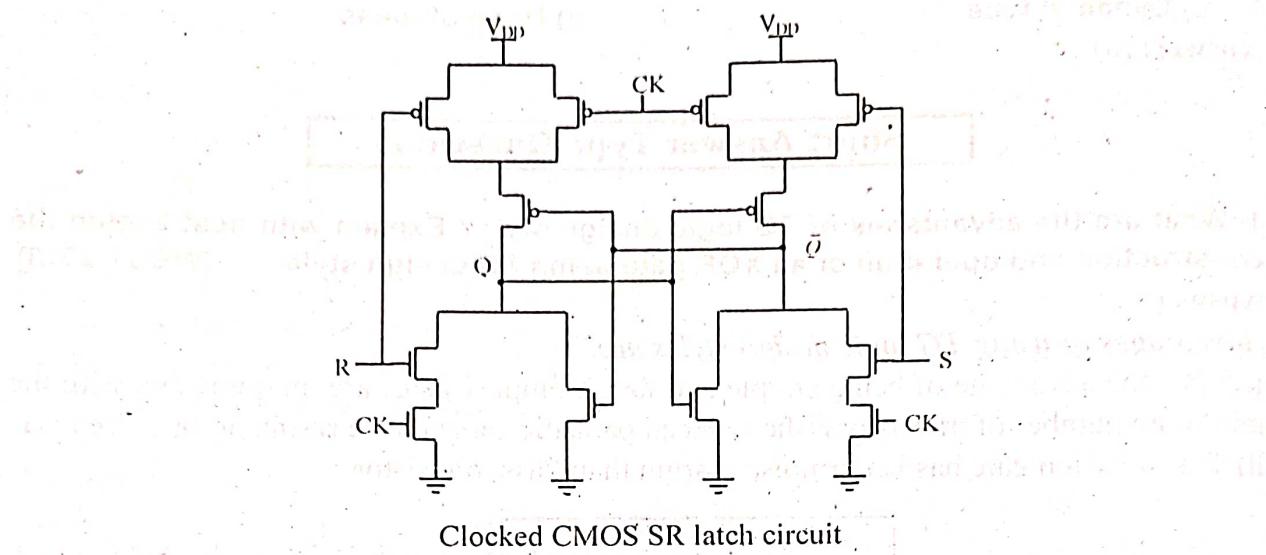
Answer:

*Clocked CMOS S-R latch*

The basic latch circuit is a NOR based S-R latch. It can be observed that if the clock (CK) is equal to logic '0', the input signals have no influence upon the circuit response. The SR latch hold its current state regardless of the 'S' and 'R' input signals.

When the clock input goes to logic '1', the logic levels applied to the S and R inputs are transferred to the SR latch and the outputs change accordingly.

## POPULAR PUBLICATIONS



Clocked CMOS SR latch circuit

3. Design a half adder using standard CMOS AOI logic. If the Precharge-Evaluate (PE) logic is used to design the same circuit, how many transistors will be required? In the Industrial view, which technology is preferable? [WBUT 2014]

OR,

Design half Adder circuit using CMOS.

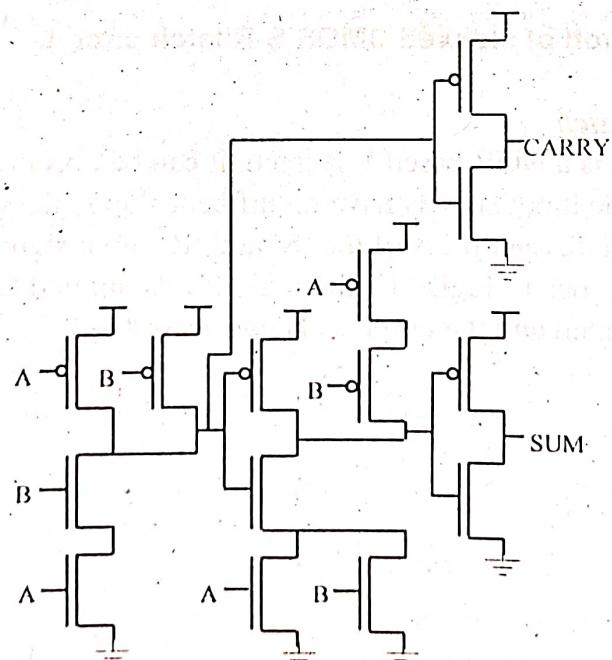
[WBUT 2018]

OR,

Design a CMOS Half-adder circuit.

[WBUT 2018]

Answer:



CMOS Half Adder Circuit

In precharge-evaluate logic, number of transistor would be half of that of CMOS logic plus two.

In industry, precharge-evaluate logic is preferable.

**4. Compare between static logic and dynamic logic. Explain the operation of Domino-logic to design any CMOS circuit.** [WBUT 2014]

**Answer:**

**1<sup>st</sup> Part:**

The difference between static and dynamic logic is illustrated in fig. (i) for a 2-input NAND gate. The static gate uses complementary p-channel and n-channel blocks to ensure that a conducting path, either to ground or  $V_{DD}$ , exists for valid input logic states. The dynamic gate uses the parasitic capacitance at the output node to temporarily store a pre-charge (high logic level for  $\phi = 0V$ ); when the gate evaluates ( $\phi = +5V$ ) this charge either remains high ('1' output) or is discharged ('0' output) depending upon the state of the middle two n-channel transistors.

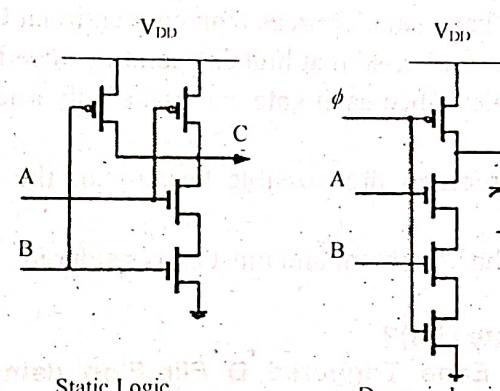


Fig. (i): A 2-input NAND gate

**2<sup>nd</sup> Part:**

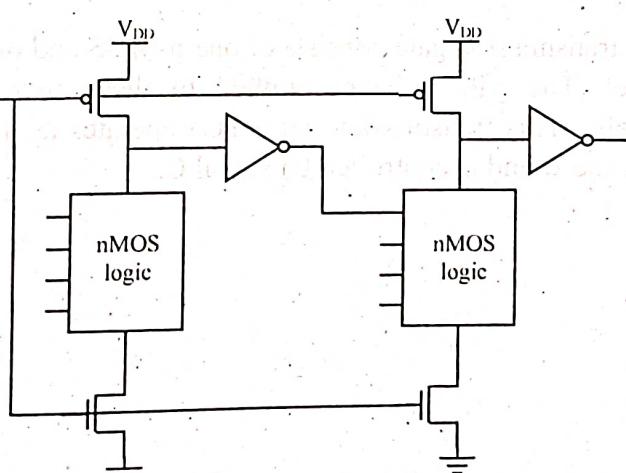


Fig. (ii): Cascaded domino CMOS logic gates

## POPULAR PUBLICATIONS

In this type of dynamic CMOS logic, the output is connected with a static CMOS inverter. The addition of this inverter makes this dynamic CMOS logic to be cascaded. Here, during pre-charge phase the output of the dynamic CMOS stage is pre-charged to a high logic level and the output of the inverter to a low logic level. Now according to external inputs, during evaluation, the output of the inverter is restricted to only the transition from '0' to '1', but from logic '1' to '0' is never possible. So, when these logic blocks are cascaded, all input transistors in subsequent logic blocks will be turned off during the pre-charge phase, since all buffer outputs are equal to '0'. During the evaluation phase, each buffer output can make at most one transistor (0 to 1) and thus each input of all subsequent logic stages can also make at most one transition. In a cascade structure consisting of several such stages, the evaluation of each stage ripples the next stage evaluation, similar to a domino falling one after the other. The structure is hence called domino CMOS logic.

For design consideration, following are the important features:

1. Such logic structures can have smaller areas than conventional CMOS logic.
2. Parasitic capacitances are smaller so that higher operating speeds are possible.
3. Operation is free of glitches since each gate can make only one logic '1' to logic '0' transition.
4. Only non-inverting structures are possible because of the presence of inverting buffer.
5. Charge distribution may be a problem and must be considered.

### 5. What is Transmission Gate (TG)?

[WBUT 2014, 2017]

Explain the operation of Edge Triggered D Flip-Flop using CMOS TG gates.

Implement the expression using CMOS TG logic.  $Z = XY' + X'Y$ . ( $X'$  = complement of  $X$ ). [WBUT 2014]

Answer:

1<sup>st</sup> Part: The CMOS transmission gate consists of one n-MOS and one p-MOS transistor, connected in parallel. The gate voltages applied to these two transistors are also complementary signals. This transmission gate (TG) operates as a bidirectional switch between the nodes A and B and is controlled by signal C.

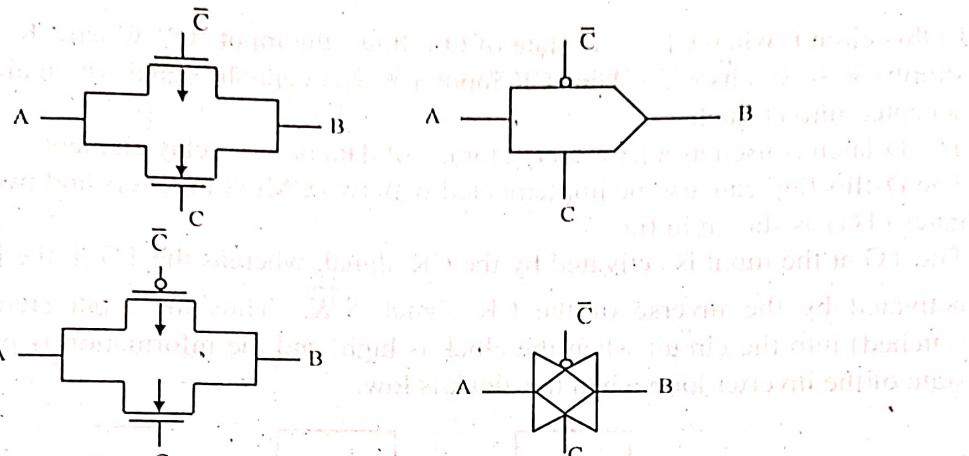


Fig: Symbols of CMOS transmission gate

### Rest part: CMOS D - Latch

Direct CMOS implementations of conventional circuits require a large no. of transistors. But use of CMOS transmission gates make the circuits simpler and require fewer transistors. A simple D-latch circuit is shown in fig. 1 and fig.2.

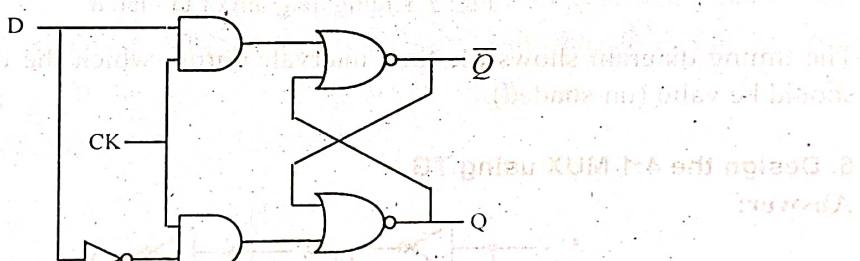


Fig: 1 gate level schematic of D-latch

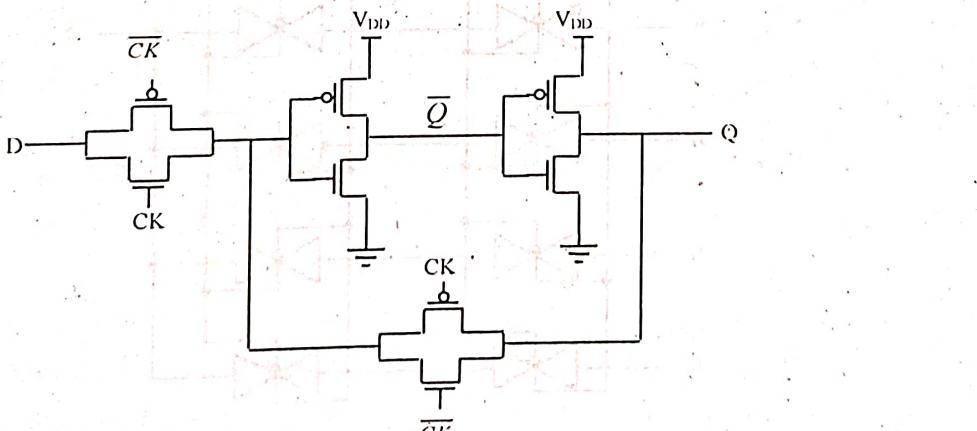


Fig: 2 CMOS implementation of D-latch

## POPULAR PUBLICATIONS

In this circuit, when  $CK = '1'$  state of  $Q$  follows the input ' $D$ '. When  $CK = '0'$ , the output simply preserves its state. Thus  $CK$  input acts as an enable signal which allows data to be accepted into D-latch.

The D-latch is used as a temporary storage of data or as a delay element.

The D-flip-flop can also be implemented with two CMOS inverters and two transmission gates (TG) as shown in fig.

The TG at the input is activated by the  $CK$  signal, whereas the TG in the inverter loop is activated by the inverse of the  $CK$  signal,  $\overline{CK}$ . Thus the input signal is accepted (latched) into the circuit when the clock is high, and the information is preserved as the state of the inverter loop when the clock is low.

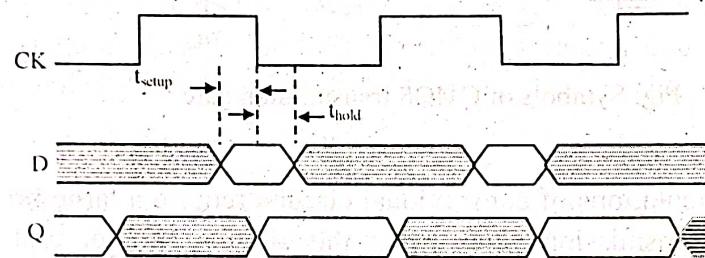


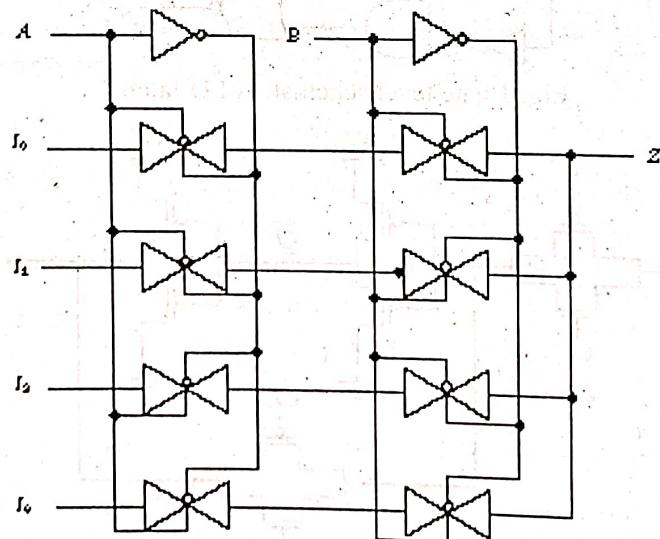
Fig: 3 Timing diagram of D - latch

The timing diagram shows the time intervals during which the input and output signals should be valid (un-shaded).

### 6. Design the 4:1 MUX using TG.

[WBUT 2015, 2017]

Answer:



Here  $I_0, I_1, I_2$ , and  $I_3$  are four inputs and  $A$  &  $b$  are the control lines.

7. a) Explain the working principle of a CMOS inverter. [WBUT 2016]  
 b) Draw the VTC curve of a simple CMOS inverter circuit and clearly define the different operating regions of NMOS and PMOS.

Answer:

a)

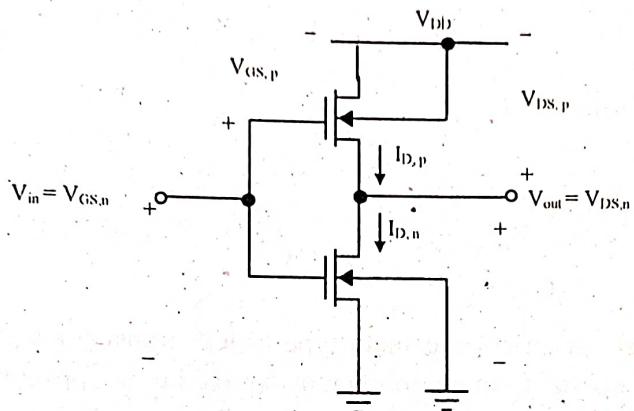


Fig: 1 CMOS inverter circuit

#### Circuit operation:

Here, both the transistors are driven directly by  $V_{in}$ . In order to reverse bias the source and drain junctions, substrate of pMOS is connected to  $V_{DD}$ , and substrate nMOS to ground.

So, as for both the transistors,  $V_{BS} = 0$  there is no substrate bias effect.

According to the circuit diagram,

$$\begin{aligned} V_{GS,n} &= V_{in} \\ \& V_{DS,n} = V_{out} \end{aligned} \quad \dots(1)$$

And also,

$$\begin{aligned} V_{GS,p} &= -(V_{DD} - V_{in}) \\ \& V_{DS,p} = -(V_{DD} - V_{out}) \end{aligned} \quad \dots(2)$$

#### Case (I).

When  $V_{in} < V_{TO,n}$ , nMOS is in cut off region and the pMOS is in linear region.

As  $I_{D,n} = I_{D,p} = 0$ ,  $V_{DS,p} = 0$  &  $V_{out} = V_{OH} = V_{DD}$

#### Case (II)

When  $V_{in} > V_{DD} + V_{TO,p}$ , pMOS turned off.

Through nMOS operating in linear region,  $V_{DS,n} = 0V$ , because  $I_{D,n} = I_{D,p} = 0$

Therefore, output wattage  $V_{out} = V_{OL} = 0V$ .

## POPULAR PUBLICATIONS

Let us examine the operating modes of nMOS and pMOS as functions of input-output voltages. nMOS operates in saturation if  $V_{in} > V_{TO,n}$  and if the following conditions are satisfied.

$$V_{DS,n} \geq V_{GS,n} - V_{TO,n}$$

$$\text{i.e., } V_{out} \geq V_{in} - V_{TO,n}$$

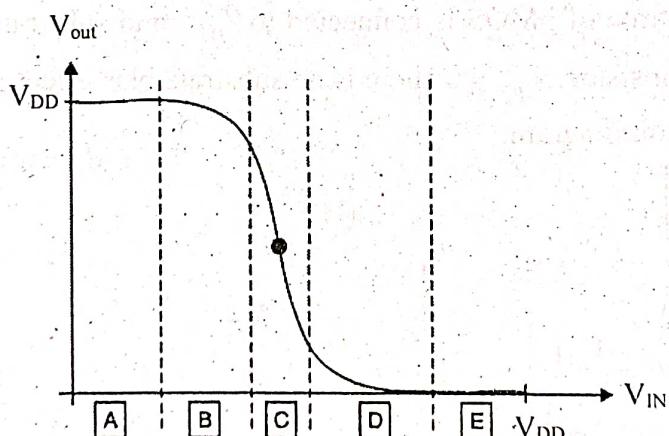
and pMOS operates in saturation if

$$V_{in} < V_{DD} + V_{TO,p}$$

$$V_{DS,p} \leq V_{GS,p} - V_{TO,p}$$

$$\text{i.e., } V_{out} \leq V_{in} - V_{TO,p}$$

- b) This inverter consists of an enhancement type nMOS transistor and an enhancement type pMOS transistor, operating in complimentary mode i.e. in complimentary push-pull in the sense that for high input, the nMOS transistor drives (pull down) the output node while the pMOS transistor acts as a load and for low input the pMOS transistor drives (pull-up) output node while the nMOS transistor acts as a load. That is why; the configuration is called complimentary MOS or CMOS structure.



Region	$V_{in}$	$V_{out}$	nMOS	pMOS
A	$< V_{TO,n}$	$V_{OH}$	cut-off	linear
B	$V_{II}$	high $\approx V_{OH}$	saturation	linear
C	$V_{II}$	$V_{II}$	saturation	saturation
D	$V_{III}$	low $\approx V_{OL}$	linear	saturation
E	$> (V_{DD} + V_{TO,p})$	$V_{OL}$	linear	cut-off

8. Explain dynamic CMOS logic and Domino CMOS logic with suitable diagram.

[WBUT 2017]

**Answer:**

### Dynamic CMOS Logic (Pre-charge – Evaluate Logic)

This type of dynamic CMOS circuit technique reduces the number of transistors used to implement any logic function.

The operation of the circuit is based on two distinct phases – pre-charge and evaluation. First, the circuit is pre-charging the output node capacitance and then evaluating the output level according to the applied inputs.

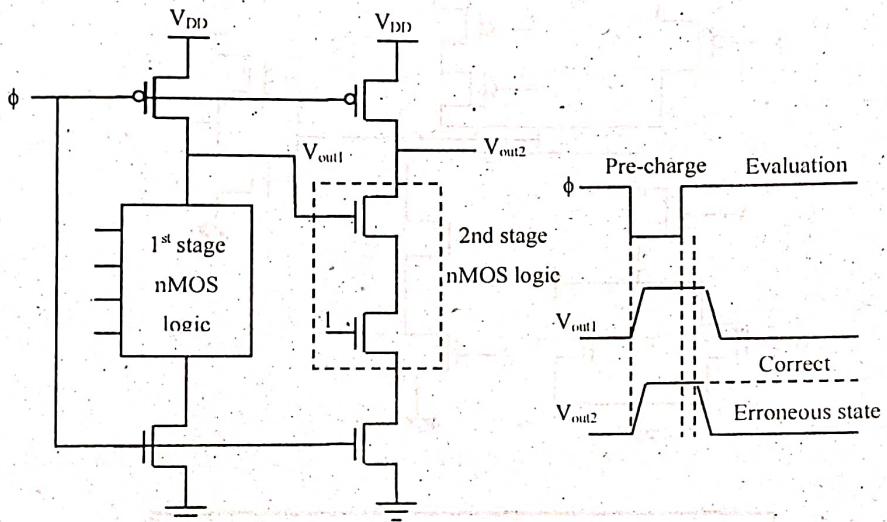


Fig: 1 Dynamic CMOS logic with precharge and evaluation

Both of these operations are scheduled by a single clock signal, which drives one nMOS and one pMOS transistor in each dynamic stage. When the clock signal is low (pre-charge phase), the pMOS pre-charge transistor is conducting, while the complementary nMOS transistor is off. The parasitic output capacitance of the circuit is charged up through the conducting pMOS transistor to a logic high level of  $V_{out} = V_{DD}$ . The input voltages are also applied during this phase, but they have no influence as the n channel path is floating. When the clock signal becomes high (evaluate phase), the pre-charge transistor turns off and nMOS transistor turns on. Now, the output node voltage depends on the input logic levels. If the input signals create a conducting path between the output node and ground, the output node capacitor is discharged and the voltage level comes down to  $V_{out} = 0V$ , otherwise it remains as it is.

### Domino CMOS logic:

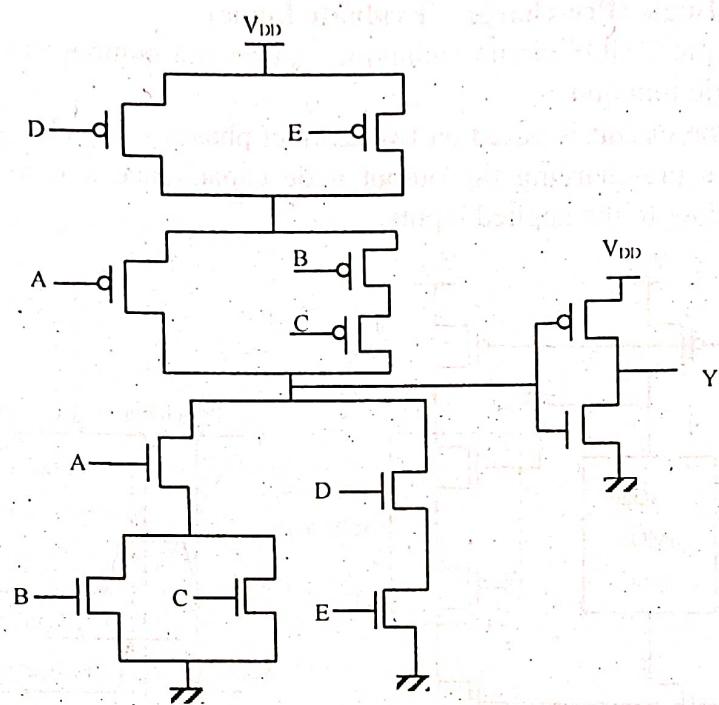
Refer to Question No. 7(a) of Long Answer Type Questions.

9. Design logic circuit using CMOS,  $Y=A(B+C)+DE$ .

[WBUT 2018]

## POPULAR PUBLICATIONS

**Answer:**

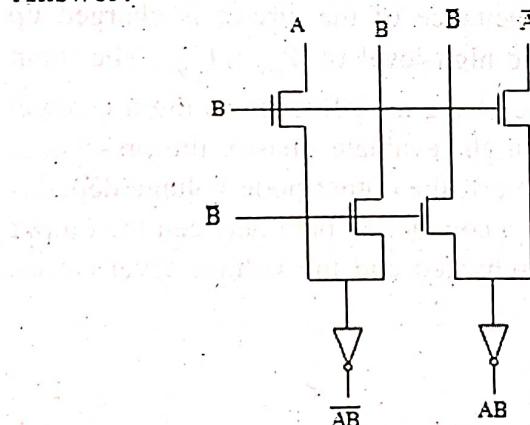


## **Long Answer Type Questions**

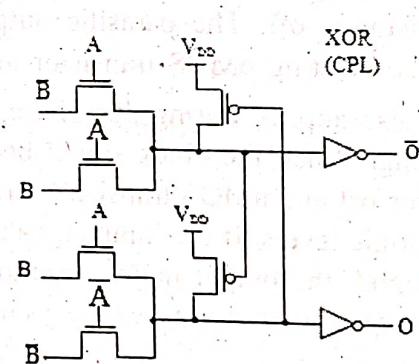
1. a) Design AND/NAND, XOR/XNOR gates using Pass Transistor Logic.

[WBUT 2013, 2017]

**Answer:**



a) AND/NAND



b) XOR/XNOR

b) Describe the Logic '0' and logic '1' transfer mechanism of a Pass Transistor.

[WBUT 2013, 2017]

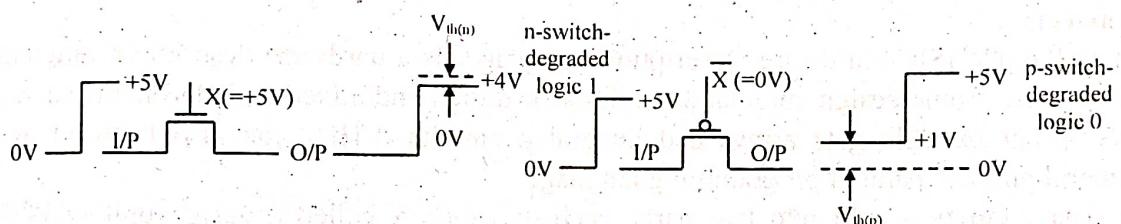
OR,

Design a D-FF using transmission gate technology.

[WBUT 2018]

Answer:

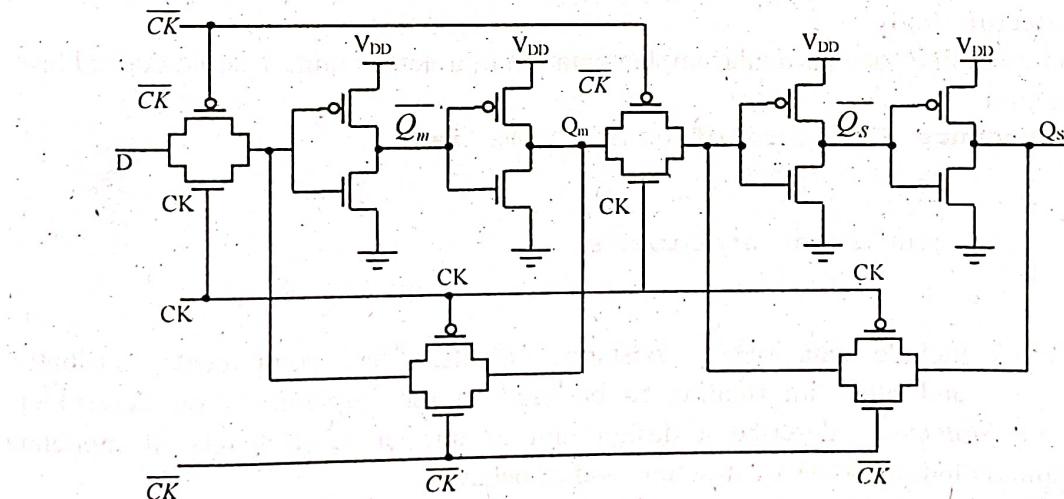
Switch logic circuits can be designed using pass transistor where no static power is consumed from power supply. Switches and switch logics may be formed from simple n- or p- pass transistors. But when logic is passed through the pass transistor, there is an undesirable threshold voltage effects which give rise to the loss of logic levels in pass transistors. For n-MOS switch logic the output voltage is always below by threshold voltage if logic '1' is applied at the input. Similarly, when logic '0' is applied to a p-MOS transistor, the output is always threshold voltage above the zero volt.



c) Design a CMOS Master Slave D flip-flop and describe its operation.

[WBUT 2013, 2017]

Answer:



CMOS negative edge-triggered master slave D flip flop

Here two CMOS D flip-flops are connected in cascaded form. The clock of the second stage is the complemented form of the first one. In this circuit, when CK = '1' state of Q follows the input 'D'. When CK='0', the output simply preserves its state. Thus CK input

## POPULAR PUBLICATIONS

acts as an enable signal which allows data to be accepted into D-latch. The D-flip-flop is implemented with two CMOS inverters and two transmission gates (TG) as shown in fig. The TG at the input is activated by the CK signal, whereas the TG in the inverter loop is activated by the inverse of the CK signal,  $\overline{CK}$ . Thus the input signal is accepted (latched) into the circuit when the clock is high, and the information is preserved as the state of the inverter loop when the clock is low.

2. a) What do you mean by VHDL? What do you mean by "Entity" and "Architecture" in a VHDL?
- b) Discuss about the different styles of describing the architecture in VHDL with example.
- c) Write a program of MOD-6 Counter by VHDL. [WBUT 2015]

Answer:

a) VHDL (VHSIC Hardware Description Language) is a hardware description language used in electronic design automation to describe digital and mixed-signal systems such as field-programmable gate arrays and integrated circuits. VHDL can also be used as a general purpose parallel programming language.

A design entity is split into two parts, each of which is called a design unit in VHDL. The entity declaration represents the external interface to the design entity. The architecture body represents the internal description of the design entity its behavior, its structure, or a mixture of both.

### b) Architecture body

Architecture defines one particular implementation of a design unit, at some desired level of abstraction.

```
architecture arch_name of entity_name is
    .... declarations ...
begin
    ... concurrent statements ...
end
```

*Declarations* include data types, constants, signals, files, components, attributes, subprograms, and other information to be used in the implementation description. *Concurrent statements* describe a design unit at one or more levels of modeling abstraction, including dataflow, structure, and/or behavior.

- Concurrent behavioral description
- Sequential behavioral description
- Structural description

The architecture body looks as follows,

```
architecture architecture_name of
    NAME_OF_ENTITY is
    -- Declarations
        -- components declarations
        -- signal declarations
        -- constant declarations
        -- function declarations
        -- procedure declarations
        -- type declarations
    :
begin
    -- Statements
    :
end architecture_name;
```

c) VHDL code for MOD 6 Counter:

```
entity mod6_counter is
    port(
        clk : in STD_LOGIC;
        reset : in STD_LOGIC;
        dout : out STD_LOGIC_VECTOR(2 downto 0)
    );
end mod6_counter;
architecture mod6_counter_arch of mod6_counter is
begin
    counter : process (clk,reset) is
    variable m : integer range 0 to 7 := 0;
    begin
        if (reset='1') then
            m := 0;
        elsif (rising_edge (clk)) then
            m := m + 1;
        end if;
        if (m=6) then
            m := 0;
        end if;
        dout <= conv_std_logic_vector (m,3);
    end process counter;
end mod6_counter_arch;
```

POPULAR PUBLICATIONS

3. a) Design AND/NAND, XOR/XNOR gates using pass transistor logic. [WBUT 2015]  
b) Design a CMOS Master Slave D flip-flop and describe its operation.  
c) What do you mean by DCVSL Design? Design XOR/XNOR using DCVSL.

Answer:

a) Refer to Question No. 1(a) of Long Answer Type Questions.

b) Refer to Question No. 1(c) of Long Answer Type Questions.

c) DCVSL- Differential Cascode Voltage Switch Logic. This logic substantially reduces area and parasitic capacitances and increase speed compared to CMOS circuit. An example of this logic is as given below.

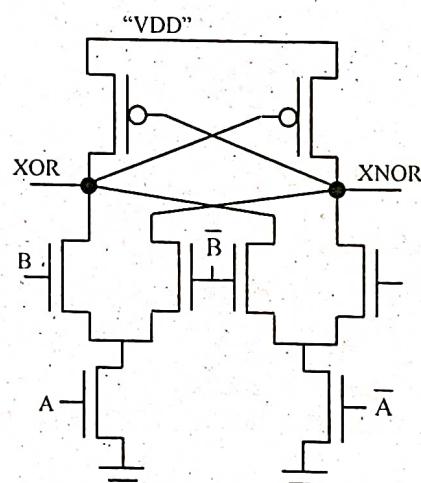
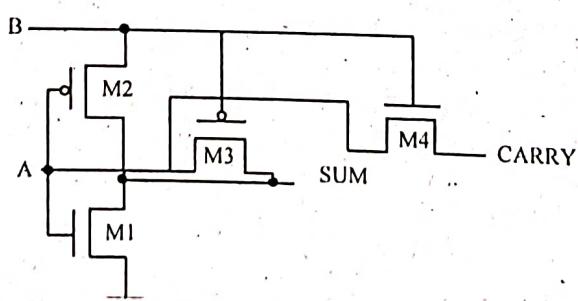


Fig: N-Block Logic Tree

4. a) Design a CMOS half adder using smallest possible number of transistors.

[WBUT 2016]

Answer:



CMOS Half-Adder using minimum number of transistors

b) Draw a clocked D flip-flop using CMOS and explain.

[WBUT 2016]

**Answer:**

Direct CMOS implementations of conventional circuits require a large no. of transistors. But use of CMOS transmission gates make the circuits simpler and require fewer transistors. A simple D-latch circuit is shown in fig. 1 and fig.2.

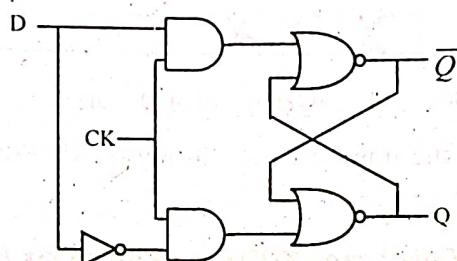


Fig: 1 gate level schematic of D-latch

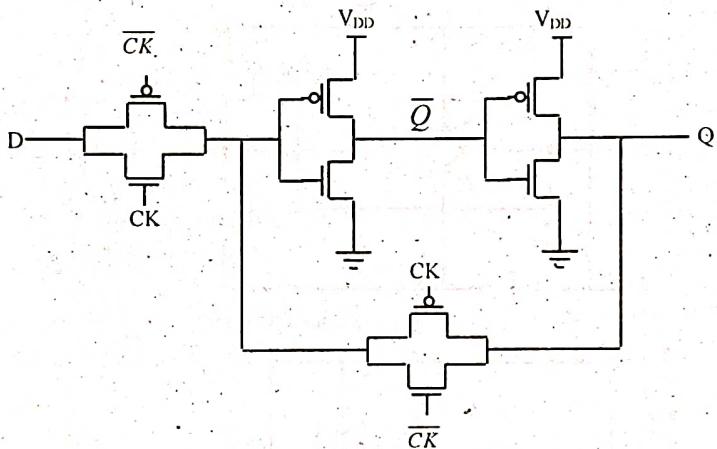


Fig: 2 CMOS implementation of D-latch

In this circuit, when CK = '1' state of Q follows the input 'D'. When CK='0', the output simply preserves its state. Thus CK input acts as an enable signal which allows data to be accepted into D-latch.

The D-latch is used as a temporary storage of data or as a delay element.

The D-flip-flop can also be implemented with two CMOS inverters and two transmission gates (TG) as shown in fig. 2.

The TG at the input is activated by the CK signal, whereas the TG in the inverter loop is activated by the inverse of the CK signal, CK̄. Thus the input signal is accepted (latched) into the circuit when the clock is high, and the information is preserved as the state of the inverter loop when the clock is low.

POPULAR PUBLICATIONS

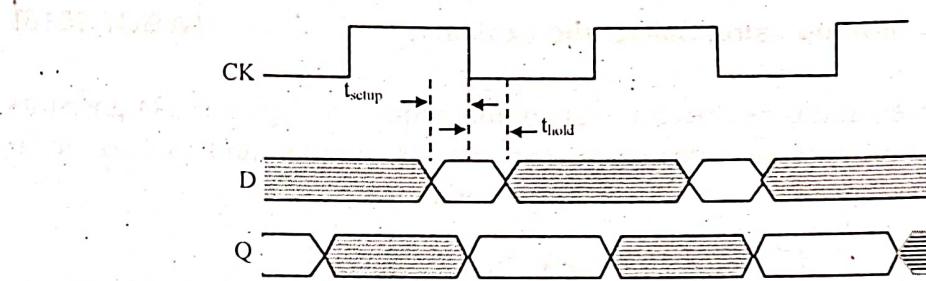


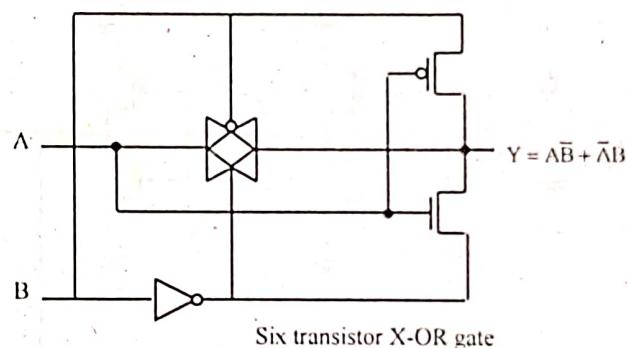
Fig: 3. Timing diagram of D - latch

The timing diagram shows the time intervals during which the input and output signals should be valid (un-shaded).

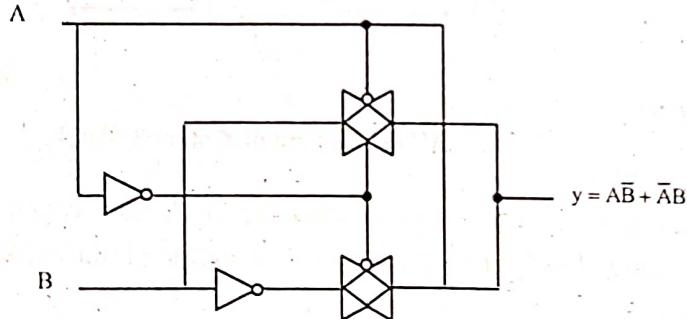
c) Design a transmission Gate based XNOR gate using six transistors.

[WBUT 2016]

**Answer:**



Six transistor X-NOR gate



Eight transistor X OR gate

d) What is pseudo-NMOS logic?

[WBUT 2016]

**Answer:**

In complex CMOS logic gates, large no. of transistors are used since complementary transistors, one nMOS and one pMOS transistor are used for each input. So, it is not suitable for high density design. In place of total pMOS transistor block, if we use a single pMOS transistor with its gate terminal connected to ground, as a load device, we can realize any Boolean junction, but the no. of transistors can be reduced considerably (almost half).

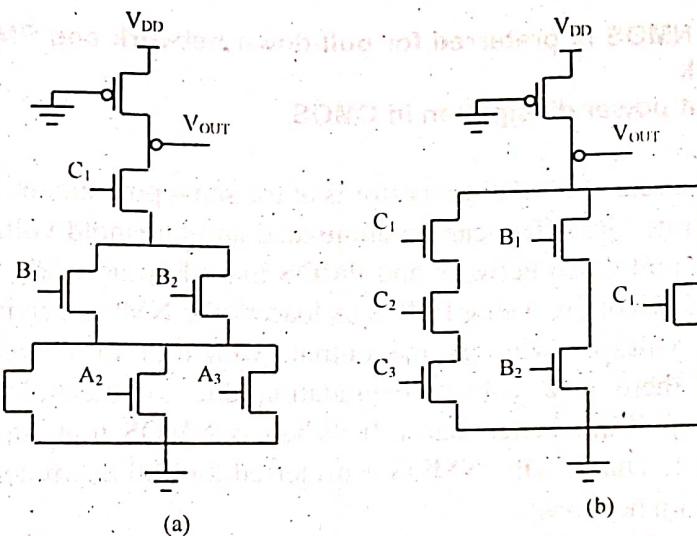


Fig: 1 The pseudo nMOS implementation of (a) OAI gate (b) AOI gate

Though the number of components are reduced, as well as the space requirement, the most significant disadvantage of a pseudo n- MOS gate instead of a full CMOS gate is the non zero static power dissipation, since the pMOS transistor load is always on and conducts a steady state current when the output voltage is lower than V<sub>DD</sub>. The ratio of transconductances of p- MOS load and the nMOS pull down block determine the value of V<sub>OL</sub> and noise margin of the logic gate.

**5. Draw the layout and schematic diagram of a 2-input static CMOS NAND gate and clearly identify the corresponding components in the two drawings. [WBUT 2016]**

**Answer:**

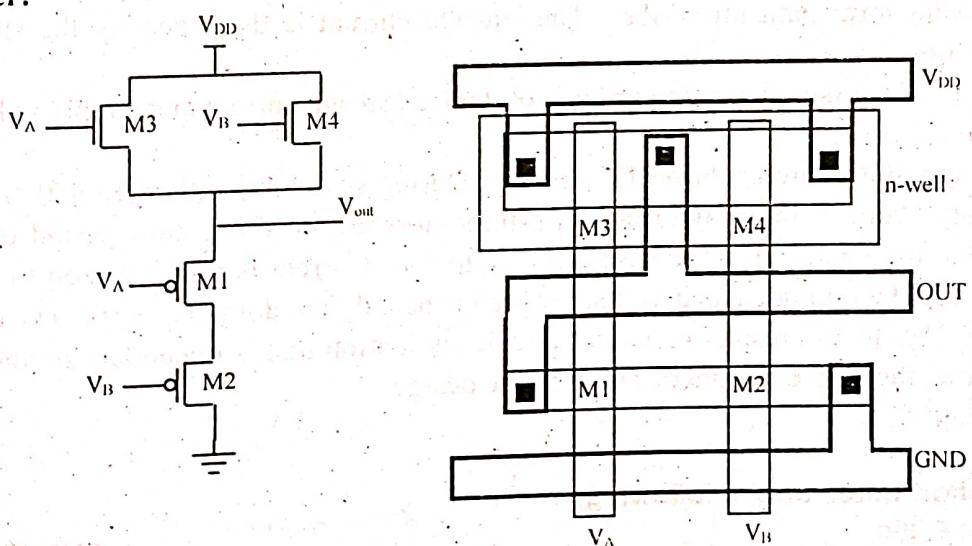


Fig: 1 Layout diagram of CMOS NAND2 gate

## POPULAR PUBLICATIONS

6. a) Explain why NMOS is preferred for pull-down network and PMOS is preferred for pull-up network.

b) Explain different power dissipation in CMOS.

[WBUT 2017]

**Answer:**

a) If the bulk or substrate of a MOS transistor is at the same potential of the source of that transistor, the substrate bias effect can be eliminated and threshold voltages are reduced. So using NMOS in pull down network and PMOS in pull up network, we can eliminate substrate bias effect. Secondly, using PMOS as load of the NMOS driving transistors, we can maximize the voltage swing at the output, which in turn maximize the noise immunity because there is a voltage degradation due to threshold voltage. PMOS transistor pass a 1 (pull-up) better than a 0. Whereas NMOS transistor pass a 0 (pull-down) better than a 1. That is why, NMOS is preferred for pull down network and PMOS is preferred for pull up network.

b) There are three main sources of power dissipation:

- Static power dissipation (PS)
- Dynamic power dissipation (DS)
- Short circuit power dissipation (PSC)

Total power dissipation is the sum of those three.

**Static Power Dissipation:** When CMOS is in static condition, there is no connection between  $V_{DD}$  and GND. However, there is some small static dissipation due to reverse bias leakage between diffusion regions and the substrate. In addition, subthreshold conduction can contribute to the static dissipation. A simple model that describes the parasitic diodes for a CMOS is formed between the source-drain diffusions and the n-well diffusion form parasitic diodes. The leakage current is described by the standard diode equation.

The static power dissipation is the product of device leakage current and supply voltage.

$$P_S = i_{leakage} \times V_{DD}$$

**Dynamic and Short Circuit Power Dissipation:** During switching, either from '0' to '1' or, alternatively, from '1' to '0', both n- and p-transistors are on for a short period of time. This results in a short current pulse from  $V_{DD}$  to  $V_{SS}$ . Current is also required to charge and discharge the output capacitive load, which is usually the dominant term. The current pulse from  $V_{DD}$  to  $V_{SS}$  results in a short circuit dissipation that is dependent on the input rise/fall time, the load capacitance and the gate design.

$$P_D = C_L \cdot V_{DD}^2 \cdot f.$$

7. Write short notes on the following:

a) Domino logic

[WBUT 2013]

b) Design of  $M \times N$  bit SRAM

[WBUT 2013]

c) CMOS NORA logic

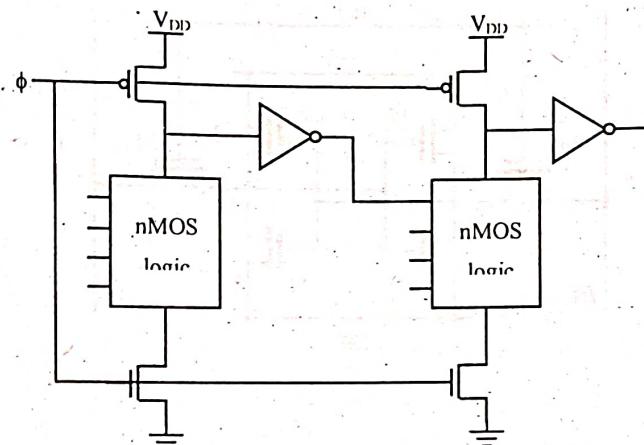
[WBUT 2013]

d) Propagation delay time ( $\Gamma_{PHL}$ ) for CMOS inverter

[WBUT 2015]

**Answer:**

**a) Domino Logic:**



Cascaded domino CMOS logic gates

In this type of dynamic CMOS logic, the output is connected with a static CMOS inverter. The addition of this inverter makes this dynamic CMOS logic to be cascaded. Here, during pre-charge phase the output of the dynamic CMOS stage is pre-charged to a high logic level and the output of the inverter to a low logic level. Now according to external inputs, during evaluation, the output of the inverter is restricted to only the transition from '0' to '1', but from logic '1' to '0' is never possible.

So, when these logic blocks are cascaded, all input transistors in subsequent logic blocks will be turned off during the pre-charge phase, since all buffer outputs are equal to '0'. During the evaluation phase, each buffer output can make at most one transistor (0 to 1) and thus each input of all subsequent logic stages can also make at most one transition. In a cascade structure consisting of several such stages, the evaluation of each stage ripples the next stage evaluation, similar to a domino falling one after the other. The structure is hence called **domino CMOS logic**.

**b) Static RAM (SRAM):**

- Static Random Access Memory
  - **Static:** Data value is retained as long as  $V_{DD}$  is present.
  - **Random Access:** Any location can read at a point in time.(Doesn't need sequential addresses)

SRAM can be built using either:

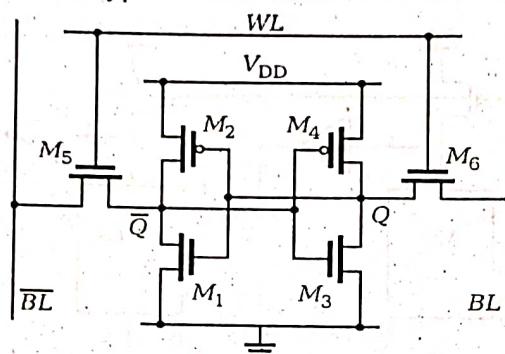
- D-type latch
- 6-transistor CMOS RAM cell

In D-type latch, each BIT would need 16 transistors (NAND gate = 4 transistors). For large SRAM modules this is not very efficient. 1-MB SRAM  $\rightarrow$  8-Mb  $\rightarrow$  128 Million transistors

## POPULAR PUBLICATIONS

### 6-Transistor Cell (Cross Coupled Inverter)

For larger SRAM modules this type of basic blocks are used.



#### TO READ:

- BIT lines are charged high
- Enable line WL is pulled high, switching access transistors M5 and M6 on
- If value stored in  $\bar{Q}$  is 0, value is accessed through access transistor M5 on /BL.
- If value stored in Q is 1, charged value of Bit line BL is pulled up to  $V_{DD}$ .
- Value is 'sensed' on BL and /BL.

#### TO WRITE:

- Apply value to be stored to Bit lines BL and /BL
- Enable line WL is triggered and input value is latched into storage cell
- BIT line drivers must be stronger than SRAM transistor cell to override previous values

While Enable line is held low, the inverters retain the previous value. Could use tri-state WE line on BIT to drive into specific state. Transistor count per bit is only 6 + (line drivers & sense logic)

#### Addressed SRAM

Can view RAM as N-bit by M-word black box:

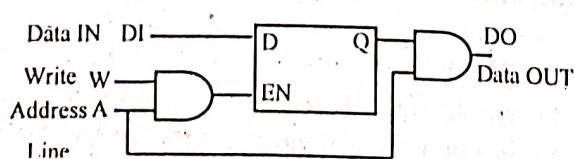
N input lines  $D_{IN}$

N output lines  $D_{OUT}$

A address lines ( $2^A = M$ ) A

WE write enable line WE

Single SRAM Bit



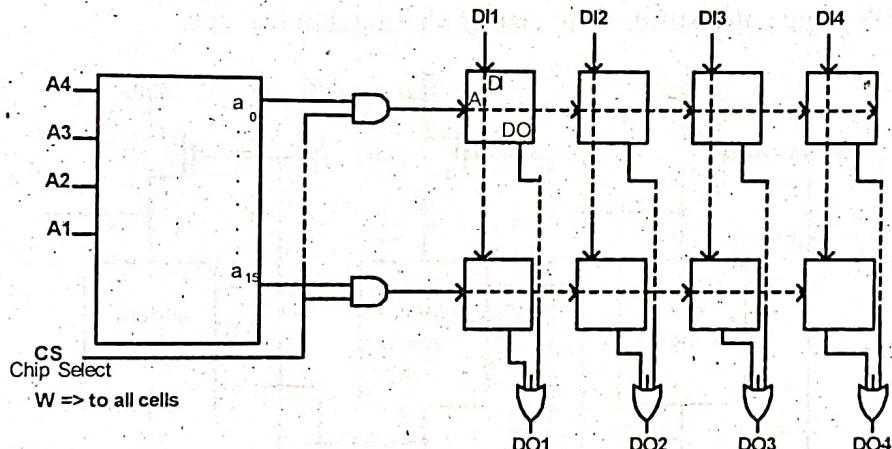
When A = 0,

- Latch Enable is off.
  - Data cannot be written into the D-type latch
  - $D_{OUT} = 0$ .

When A = 1

- Latch is Enabled
  - If W = 1 (Data-Write)
    - Data at  $D_{IN}$  can be written into the D-type latch
    - Output gate is enabled
  - IF W = 0
    - New value on  $D_{IN}$  is not stored.
    - Output gate is enabled.
- Not very efficient since 1-bit address line can access 2 memory locations.
- This memory is 1-bit X 1-word RAM. Stores one 1-bit data value

4-bit  $\times$  16-word SRAM



When CS = 1 AND A4 A3 A2 A1 = 0000

- Address decoder decodes A4-A1 to
  - 1000000000000000 ( $a_0 = 1, a_1-a_{15} = 0$ )
- Data at DI1 DI2 DI3 DI4 is written to address 0 when W = 1
- If W = 0, No new data is stored and address0 drives the output bus
- Contents of memory address 0 appear at output

Address decoder maps input address bits to row control signals

- Should only set one bit for every possible input
  - $2^A$  states where A is the number of address lines

The CS (chip select) line allows the memory to be doubled with only one inverter [+ OR gates.]

c) CMOS NORA logic:

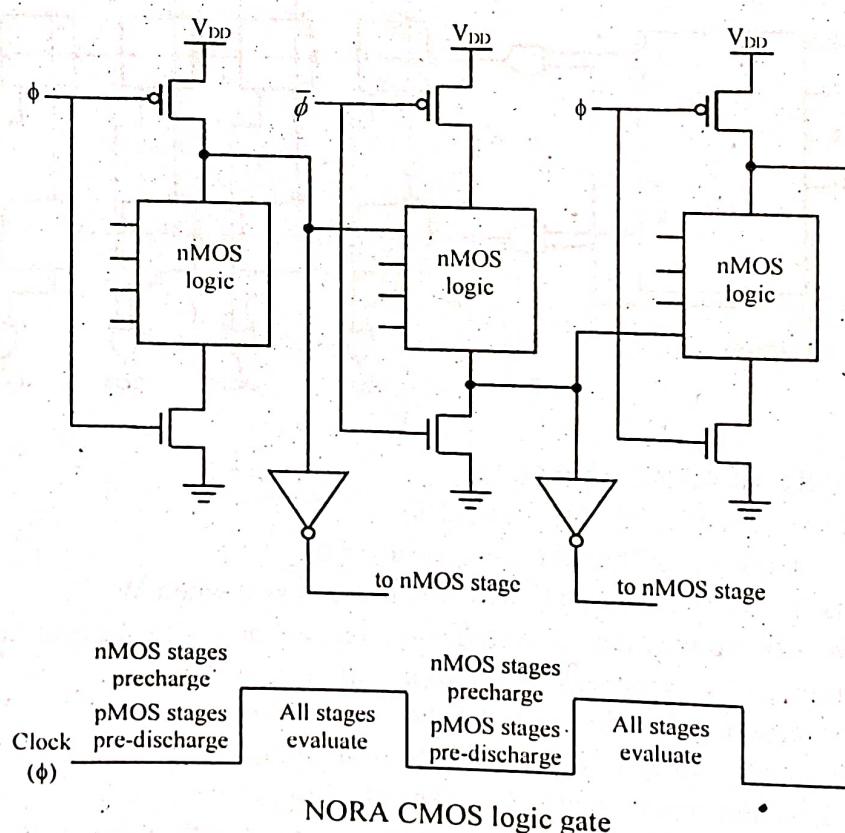
In this type of dynamic CMOS logic, alternate nMOS and pMOS logic stages are used. Here the pre-charge and evaluate timing of nMOS logic stages is accomplished by the clock signal  $\phi_1$ , whereas the pMOS logic stages are controlled by the inverted clock signal  $\bar{\phi}$ .

When the clock signal is low, the output nodes of nMOS logic blocks are pre-charged to  $V_{DD}$  through the pMOS pre-charged transistors whereas the output nodes of pMOS logic blocks are pre-charged to  $0^V$  through the nMOS discharge transistor, driven by  $\phi$ . When the clock signal makes low to high transition, all cascaded nMOS and pMOS logic states evaluate one after the other similar to domino CMOS logic.

The advantages of this logic are:

1. Static CMOS inverter is not required for cascading of stages
2. Compatible with CMOS domino logic.
3. If allows pipelined system architecture,

NORA CMOS logic gates suffer from charge sharing and leakage.



d) Propagation delay time ( $\tau_{PLH}$ ) for CMOS inverter:

The input and output voltage waveforms of a typical inverter circuit are shown in fig.(a)

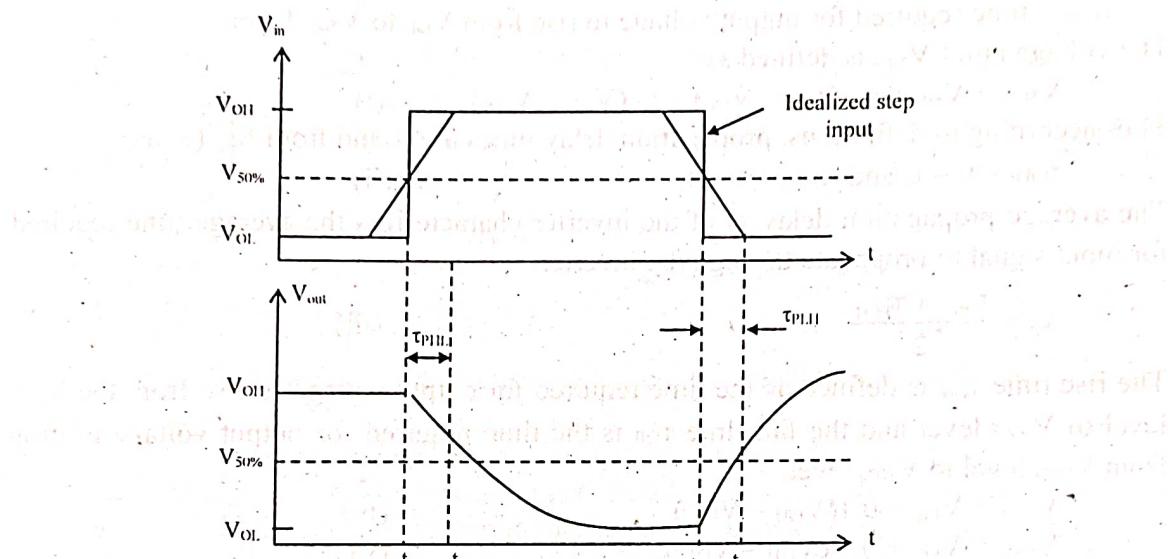


Fig: (a) Input and output voltage waveform of an inverter

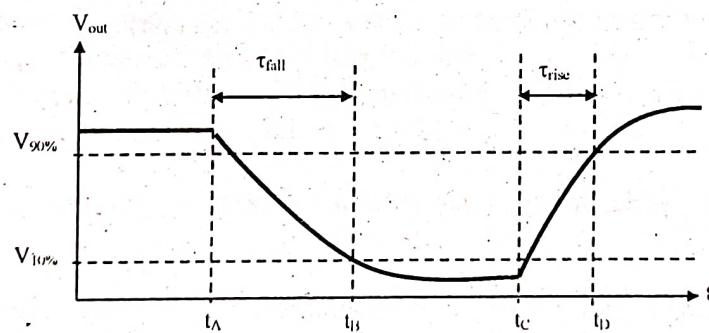


Fig: (b) output voltage rise and fall time

Here,  $\tau_{PHL}$  = Input to output signal delay during high to low transition.

$\tau_{PLH}$  = Input to output signal delay during low to high transition.

By definition:

$\tau_{PHL}$  = Time delay between  $V_{50\%}$  transition of the rising input voltage and  $50\%$  transition of the falling output voltage.

and

$\tau_{PLH}$  = Time delay between  $50\%$  transition of the falling input voltage and  $50\%$  transition of the rising output voltage.

To simplify the analysis we may consider the input pulse, an ideal one with zero rise and fall time.

## POPULAR PUBLICATIONS

So, we may consider:

$\tau_{PHL}$  = time required for output voltage to fall from  $V_{OH}$  to  $V_{50\%}$  level.

$\tau_{PLH}$  = time required for output voltage to rise from  $V_{OL}$  to  $V_{50\%}$  level.

The voltage point  $V_{50\%}$  is defined as:

$$V_{50\%} = V_{OL} + \frac{1}{2}(V_{OH} - V_{OL}) = \frac{1}{2}(V_{OL} + V_{OH}) \quad \dots (i)$$

Thus according to definitions, propagation delay times are found from fig. (a) as

$$\tau_{PHL} = t_1 - t_0 \text{ and } \tau_{PLH} = t_3 - t_2 \quad \dots (ii)$$

The average propagation delay  $\tau_p$  of the inverter characterizes the average time required for input signal to propagate through the inverter.

$$\tau_p = \frac{\tau_{PHL} + \tau_{PLH}}{2} \quad \dots (iii)$$

The rise time  $\tau_{rise}$  is defined as the time required for output voltage to rise from the  $V_{10\%}$  level to  $V_{90\%}$  level and the fall time  $\tau_{fall}$  is the time required for output voltage to drop from  $V_{90\%}$  level to  $V_{10\%}$  level.

$$V_{10\%} = V_{OL} + 0.1(V_{OH} - V_{OL}) \quad \dots (iv)$$

$$V_{90\%} = V_{OL} + 0.9(V_{OH} - V_{OL}) \quad \dots (v)$$

According to fig (b), output rise and fall times are found:

$$\tau_{fall} = t_B - t_A \quad \dots (vi)$$

$$\tau_{rise} = t_D - t_C \quad \dots (vii)$$

# PHYSICAL DESIGN AUTOMATION

## Short Answer Type Questions

**1. What Is The Most Challenging Task You Handled? What Is The Most Challenging Job In P&r Flow? [MODEL QUESTION]**

**Answer:**

- o It may be power planning- because you found more IR drop
- o It may be low power target-because you had more dynamic and leakage power
- o It may be macro placement-because it had more connection with standard cells or macros
- o It may be CTS-because you needed to handle multiple clocks and clock domain crossings
- o It may be timing-because sizing cells in ECO flow is not meeting timing
- o It may be library preparation-because you found some inconsistency in libraries.
- o It may be DRC-because you faced thousands of violations

**2. How Can You Avoid Cross Talk? [MODEL QUESTION]**

**Answer:**

- o Double spacing => more spacing => less capacitance => less cross talk
- o Multiple Bias => less resistance => less RC delay
- o Shielding => constant cross coupling capacitance => known value of crosstalk
- o Buffer insertion => boost the victim strength

**3. How Shielding Avoids Crosstalk Problem? What Exactly Happens There? [MODEL QUESTION]**

**Answer:**

- o High frequency noise (or glitch) is coupled to VSS (or VDD) since shielded layers are connected to either VDD or VSS.
- o Coupling capacitance remains constant with VDD or VSS.

**4. How Spacing Helps In Reducing Crosstalk Noise? [MODEL QUESTION]**

**Answer:**

width is more => more spacing between two conductors => cross coupling capacitance is less => less cross talk

**5. Why Double Spacing And Multiple Vias Are Used Related To Clock? [MODEL QUESTION]**

**Answer:**

- o Why clock-- because it is the one signal which changes its state regularly and more compared to any other signal. If any other signal switches fast then also we can use double space.
- o Double spacing => width is more => capacitance is less => less cross talk

## POPULAR PUBLICATIONS

- Multiple Bias => resistance in parallel => less resistance => less RC delay

### **6. How Do You Validate Your Floorplan And What Analysis You Do During Floorplan? [MODEL QUESTION]**

**Answer:**

- Overlapping of macros.
- Global route congestion => in order to finalize Min. Channel spacing.
- Allowable IR drop.
- Physical information of the design (report\_design\_physical)

### **7. Did You Get Antenna Problem In Your Project For All The Metal Layers? How Did You Fix Them? [MODEL QUESTION]**

**Answer:**

Metal Jumper and Antenna diode are two methods to resolve Antenna violations. But Metal Jumper is preferred approach as it does not need change to the Netlist and placement. This methodology works for antenna violations on all metal layers except for the top most layer. In this methodology, we will switch the small portion of routing to higher level metal close to the location of failing gate. This will make sure that accumulated charges on metal layer does not affect the gate as gate will not be connected to the charge carrying metal route until higher level metal is manufactured.

For example, lets say antenna violation is in M2. This means that M2 has enough area to accumulate large charge that induces high electron voltage to destroy the gate. To solve this problem, we cut a portion of M2 close to failing gate and move the routing to M3. This makes sure that when M2 is being manufactured, it does not get connected to gate. Connection happens only when M3 gets manufactured which is much later in time. By then charges on Metal M2 would have leaked away.

When metal jumper is not possible to implement (probably due to routing congestion or violation happening in top most layer) we try to fix it by inserting antenna diode closed to gate failing antenna. Antenna diode provide electrical path for safe conduction of accumulated charges to the substrate. Antenna diode is a reversed biased diode but acts like resistor during manufactured process (CMP) due to high temperature environment.

### **8. How Do You Reduce Power Dissipation Using High V<sub>t</sub> And Low V<sub>t</sub> On Your Design? [MODEL QUESTION]**

**Answer:**

1. Use HVT cells for timing paths having +ve slacks.
2. Use LVT cells for timing paths having -ve slacks.

HVT cells have a larger delay but less leakage. +ve slack in a design is not useful as having only some paths working faster will not help overall design. We are good if the slack is 0. In such cases give up the slack by using HVT cells but gain on power dissipation.

LVT cells are very fast but very leaky. Limit the use of LVT cells to only those paths that have difficulty in closing time.

**9. What are the physical cells?**

[MODEL QUESTION]

**Answer:**

**End Cap cells:**

1. These cells prevent the cell damage during fabrication.
2. Used for row connectivity and specifying row ending.
3. To avoid drain and source short.
4. These are used to address boundary N-Well issues for DRC cleanup.

**Well Tap cells:**

1. These are used to connect VDD and GND to substrate and N-Well respectively because it results in lesser drift to prevent latch-up.
2. If we keep well taps according to the specified distances, N-Well potential leads to proper electrical functioning.
3. To limit the resistance between power and ground connections to wells of the substrate.

**De-cap Cells:**

1. They are temporary capacitors which are added in the design between power and ground rails to counter the functional failure due to dynamic IR drop.
2. To avoid the flop which is far from the power source going into metastable state.

**10. What are all the fixing methods for setup and hold violations?**

[MODEL QUESTION]

**Answer:**

**A. Setup:**

1. Upsizing the cells
2. Replace buffer with two inverters
3. HVT to LVT
4. If the net delay is more than break the net and insert the buffer
5. Pin swapping
6. Pulling the launch and pushing the capture
7. Cloning

**B. Hold:**

1. Inserting the buffers
2. Downsizing the cells
3. LVT to HVT
4. Pushing the launch and pulling the capture

**11. What is the main objective of floor planning?**

[MODEL QUESTION]

**Answer:**

**Floor planning Goals and Objectives**

The input to a floor planning tool is a hierarchical netlist that describes the interconnection of the blocks (RAM, ROM, ALU, cache controller, and so on); the logic cells (NAND, NOR, D flip-flop, and so on) within the blocks; and the logic cell connectors (the terms terminals, pins, or ports mean the same thing as connectors). The

## POPULAR PUBLICATIONS

netlist is a logical description of the ASIC; the floorplan is a physical description of an ASIC. Floor planning is thus a mapping between the logical description (the netlist) and the physical description (the floorplan).

The goals of floor planning are to:

- arrange the blocks on a chip,
- decide the location of the I/O pads,
- decide the location and number of the power pads,
- decide the type of power distribution, and
- decide the location and type of clock distribution.

The objectives of floor planning are to minimize the chip area and minimize delay. Measuring area is straightforward, but measuring delay is more difficult and we shall explore this next.

### **Measurement of Delay in Floor planning**

Throughout the ASIC design process we need to predict the performance of the final layout. In floor planning we wish to predict the interconnect delay before we complete any routing. Imagine trying to predict how long it takes to get from Russia to China without knowing where in Russia we are or where our destination is in China. Actually it is worse, because in floor planning we may move Russia or China.

To predict delay we need to know the **parasitics** associated with interconnect: the **interconnect capacitance** (**wiring capacitance** or **routing capacitance**) as well as the interconnect resistance. At the floor planning stage we know only the **fanout** (FO) of a net (the number of gates driven by a net) and the size of the block that the net belongs to. We cannot predict the resistance of the various pieces of the interconnect path since we do not yet know the shape of the interconnect for a net. However, we can estimate the total length of the interconnect and thus estimate the total capacitance. We estimate interconnect length by collecting statistics from previously routed chips and analyzing the results. From these statistics we create tables that predict the interconnect capacitance as a function of net fanout and block size. A floor planning tool can then use these **predicted-capacitance tables** (also known as **interconnect-load tables** or **wire-load tables**). Fig. 1 shows how we derive and use wire-load tables and illustrates the following facts:

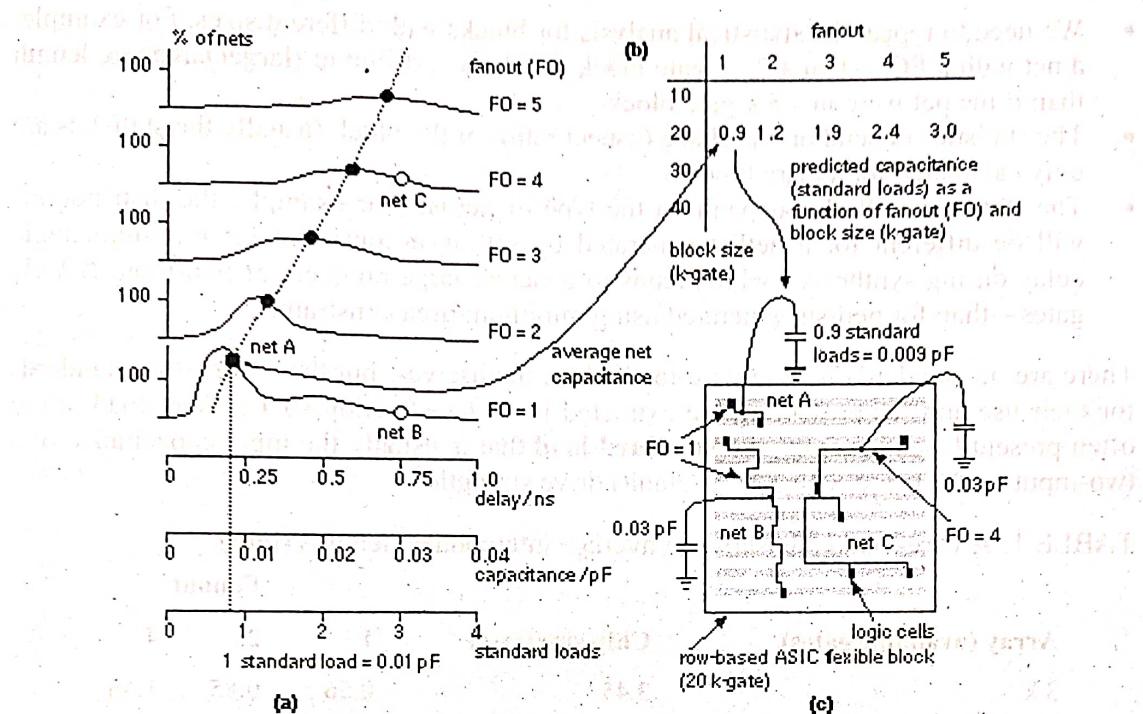


Fig. 1: Predicted capacitance. (a) Interconnect lengths as a function of fanout (FO) and circuit-block size. (b) Wire-load table. There is only one capacitance value for each fanout (typically the average value). (c) The wire-load table predicts the capacitance and delay of a net (with a considerable error). Net A and net B both have a fanout of 1, both have the same predicted net delay, but net B in fact has a much greater delay than net A in the actual layout (of course we shall not know what the actual layout is until much later in the design process).

- Typically between 60 and 70 percent of nets have a  $FO = 1$ .
- The distribution for a  $FO = 1$  has a very long tail, stretching to interconnects that run from corner to corner of the chip.
- The distribution for a  $FO = 1$  often has two peaks, corresponding to a distribution for close neighbors in subgroups within a block; superimposed on a distribution corresponding to routing between subgroups.
- We often see a twin-peaked distribution at the chip level also, corresponding to separate distributions for **interblock routing** (inside blocks) and **intrablock routing** (between blocks).
- The distributions for  $FO > 1$  are more symmetrical and flatter than for  $FO = 1$ .
- The wire-load tables can only contain one number, for example the average net capacitance, for any one distribution. Many tools take a worst-case approach and use the 80- or 90-percentile point instead of the average. Thus a tool may use a predicted capacitance for which we know 90 percent of the nets will have less than the estimated capacitance.

## POPULAR PUBLICATIONS

- We need to repeat the statistical analysis for blocks with different sizes. For example, a net with a FO = 1 in a 25 k-gate block will have a different (larger) average length than if the net were in a 5 k-gate block.
- The statistics depend on the shape (aspect ratio) of the block (usually the statistics are only calculated for square blocks).
- The statistics will also depend on the type of netlist. For example, the distributions will be different for a netlist generated by setting a constraint for minimum logic delay during synthesis—which tends to generate large numbers of two-input NAND gates—than for netlists generated using minimum-area constraints.

There are no standards for the wire-load tables themselves, but there are some standards for their use and for presenting the extracted loads (see Section 16.4). Wire-load tables often present loads in terms of a **standard load** that is usually the input capacitance of a two-input NAND gate with a 1X (default) drive strength.

TABLE 1: A wire-load table showing average interconnect lengths (mm).

Array (available gates)	Chip size (mm)	Fanout		
		1	2	4
3 k	3.45	0.56	0.85	1.46
11 k	5.11	0.84	1.34	2.25
105 k	12.50	1.75	2.70	4.92

Table 1 shows the estimated metal interconnect lengths, as a function of die size and fanout, for a series of three-level metal gate arrays. In this case the interconnect capacitance is about  $2 \text{ pFcm}^{-1}$ , a typical figure.

Fig. 2 shows that, because we do not decrease chip size as we scale down feature size, the worst-case interconnect delay increases. One way to measure the worst-case delay uses an interconnect that completely crosses the chip, a **coast-to-coast** interconnect. In certain cases the worst-case delay of a 0.25 m m process may be worse than a 0.35 m m process, for example.

Fig. 2: Worst-case interconnect delay. As we scale circuits, but avoid scaling the chip size, the worst-case interconnect delay increases.

