

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

Electronics & Communication Engineering VII-Semester

EC- 701 VLSI Design

Course Objective:

- To understand the fabrication process of CMOS technology.
- To teach fundamentals of VLSI circuit design and implementation using circuit simulators and layout editors.
- To study various problems due to VLSI technology advancement.
- To study digital circuits using various logic methods and their limitations.
- To highlight the circuit design issues in the context of VLSI technology.

Course Contents:

UNIT I

Practical Consideration and Technology in VLSI Design

Introduction, Size and complexity of Integrated Circuits, The Microelectronics Field, IC Production Process, Processing Steps, Packaging and Testing, MOS Processes, NMOS Process, CMOS Process, Bipolar Technology, Hybrid Technology, Design Rules and Process Parameters.

UNIT II

Device Modeling

Dc Models, Small Signal Models, MOS Models, MOSFET Models in High Frequency and small signal, Short channel devices, Sub threshold Operations, Modeling Noise Sources in MOSFET's, Diode Models, Bipolar Models, Passive component Models.

UNIT III

Circuit Simulation

Introduction, Circuit Simulation Using Spice, MOSFET Model, Level 1 Large signal model, Level 2 Large Signal Model, High Frequency Model, Noise Model of MOSFET, Large signal Diode Current, High Frequency BJT Model, BJT Noise Model, temperature Dependence of BJT.

UNIT IV

Structured Digital Circuits and Systems

Random Logic and Structured Logic Forms, Register Storage Circuits, Quasi Static Register Cells, AStatic Register Cell, Micro coded Controllers, Microprocessor Design, Systolic Arrays, Bit-Serial Processing Elements, Algotronix.

UNIT V

CMOS Processing Technology

Basic CMOS Technology, A Basic n-well CMOS Process, Twin Tub Processes, CMOS Process Enhancement, Interconnects and Circuit Elements, Layout Design Rules, Latch up, Physical Origin, Latchup Triggering, Latch up Prevention, Internal Latch up Prevention Techniques.

Course Outcome:Upon successful completion of this course, the student will be able to:

- Demonstrate a clear understanding of CMOS fabrication flow and technology scaling.
- Design MOSFET based logic circuit
- Draw layout of a given logic circuit
- Demonstrate an understanding of working principle of operation of different types of memories
- Demonstrate an understanding of working principles of clocking, power reduction and Distribution

References:

1. Geiger, Allen andStrader: VLSI Design Techniques for Analog and Digital Circuits, TMH.
2. Sorab Gandhi: VLSI Fabrication Principles, Wiley India.
3. Weste and Eshraghian:Principles of CMOS VLSI design, Addison-Wesley
4. Weste, Harris and Banerjee: CMOS VLSI Design, Pearson-Education.
5. Pucknell and Eshraghian: Basic VLSI Design, PHI Learning.
6. Botkar: Integrated Circuits, Khanna Publishers.
7. Sze:VLSI Technology, TMH.

StreamTechNotes

**Subject Name:** VLSI Design**Subject Notes****UNIT-I****Practical Consideration and Technology in VLSI Design**

Introduction, Size and complexity of Integrated Circuits, The Microelectronics Field, IC Production Process, Processing Steps, Packaging and Testing, MOS Processes, NMOS Process, CMOS Process, Bipolar Technology, Hybrid Technology, Design Rules and Process Parameters.

Introduction

An integrated circuit or monolithic integrated circuit (also referred to as an IC, a chip, or a microchip) is a set of electronic circuits on one small plate of semiconductor material, normally silicon. This can be made much smaller than a discrete circuit made from independent components. ICs can be made very compact, having up to several billion transistors and other electronic components in an area the size of a fingernail. The width of each conducting line in a circuit can be made smaller and smaller as the technology advances; in 2008 it dropped below 100 nanometers and in 2017 about 7 nanometer.

There are two main advantages of ICs over discrete circuits: cost and performance. Cost is low because the chips, with all their components, are printed as a unit by photolithography rather than being constructed one transistor at a time. Furthermore, much less material is used to construct a packaged IC die than to construct a discrete circuit. Performance is high because the components switch quickly and consume little power (compared to their discrete counterparts) as a result of the small size and close proximity of the components.

Terminologies associated with integrated circuit design are:

- 1) Integrated circuit (IC) :** IC is a combination of interconnected circuit elements associated on or within a substrate.
- 2) Substrate:** Supporting material upon or within an IC is fabricated.
- 3) Hybrid IC:** IC consists of a combination of two or more IC or an IC with some discrete elements.
- 4) Monolithic IC:** An IC whose elements are formed in place upon or within a semiconductor substrate with atleast one of the elements formed within the substrate.
- 5) Wafer or Slice:** It is a physical unit used in processing. Typically wafer is circular; production wafers have a diameter of **4,5 or 6 in.**
- 6) Chip or die or bar:** It is one of the repeated ICs on a wafer. Production wafer may contains 20 or 30 ICs, or several hundreds or thousands depending upon the complexity and size of circuit fabricated.
- 7) Test Plug:** It is a special chip repeated only for few times on each wafer. Used to monitor the process parameters of the technology. After processing, the verification of process is verified by measuring, at the wafer probe level, the characteristics of devices or circuits on the test plug. If the key parameters at the test plug level are not matched , the wafer is discarded. Test plug is also known as process control bar (PCB) or process control monitor (PCM).
- 8) Test cell:** Special chip repeated only few times on each wafer. The circuit designers include test cell to monitor the performance of subcircuits or subcomponents.

Size and complexity of Integrated Circuits

IC are classified in terms of device count used in the design of the circuit and in terms of the minimum feature size (such as minimum gate length or minimum polysilicon width or minimum metal width) or in terms of the pitch (minimum of the sum of the minimum width of a feature and minimum spacing between similar features). The pitch is often nearly twice the minimum feature size. Figure 01 shows the 3-dimentional view of FET.

Classification of IC by device count:

Nomenclature		Active device count
SSI	Small scale integration.	1-100
MSI	Medium scale integration.	100-1000.
LSI	Large scale integration.	1000-100000
VLSI	Very large scale integration.	$10^5 - 10^6$

Classification based on feature size:

Year	Minimum Feature Size in microns (μ)
1970	7 to 10
1980	5
Mid-1980	2 to 1.25
1990	0.75 to 0.25

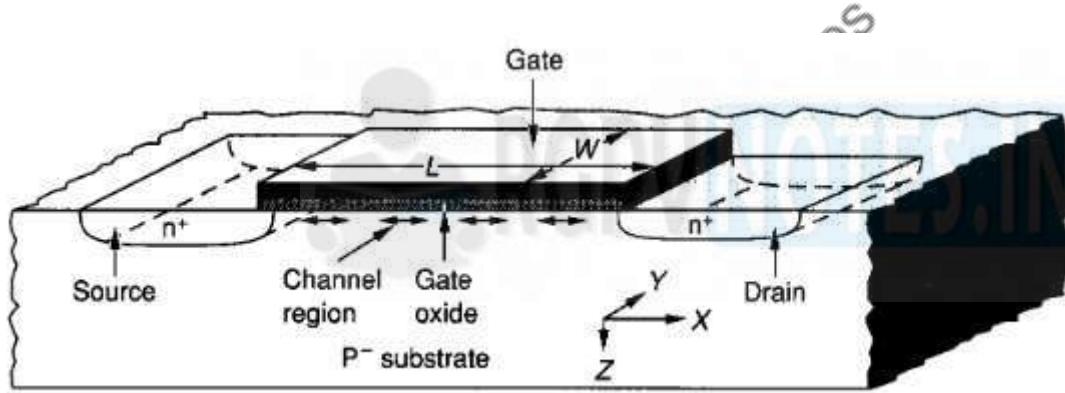


Figure:01: 3-dimensional view of a FET

For Example: In 5μ process, the minimum value of Width (W) and Length (L) of a gate would be 5μ and the area required for the gate of FET is $25\mu^2$.

Impact of shrinking of feature size:

The number of devices (transistors) that can be fabricated on a given piece of silicon can be determined. For 4-inch wafer used in a 5μ process can accommodate-

$$N_{5\mu} = \frac{\pi(2 \text{ in})^2}{25 \mu^2} \frac{2.54 \times 10^{-2} (\mu)}{\text{in}} = 3.24 \times 10^8 \text{ Transistors}$$

Now when feature size shrinks – the number of transistors for 4-inch wafer used in a 0.5μ process can accommodate-

$$N_{0.5\mu} = \frac{\pi(2 \text{ in})^2}{0.25 \mu^2} \frac{2.54 \times 10^{-2} (\mu)}{\text{in}} = 3.24 \times 10^{10} \text{ Transistors}$$

This is the impact of shrinking the feature size from 5μ process to 0.5μ process.

Due to the shrinking of feature size, there is the significance of 100-fold increase in the device count as well as the speed of the circuit increases linearly with the feature size reduction.

Limitations associated with shrinking the feature size:

- a) Wear and tear in matching characteristics.
- b) Increased cost of equipments required for processing the wafers.
- c) Advanced software design aids required.
- d) Increased impact of interconnection delays.
- e) Increased power dissipation and processing complications associated with heat cycling limitations during fabrication.

Wafer Size: The number of devices that could potentially be placed on a wafer is strongly dependent upon the wafer size. For example, in a 5μ process, the 1 cm^2 chip can accommodate the gates of about 4 million $5\mu \times 5\mu$ transistors.

Advantages of using smaller size die or chip are:

- a) Fabrication of more chips per wafer.
- b) Reduction in effective cost per chip.
- C) Percentage of good chips increases.
- D) Rectangular chips are fabricated on round wafers, the amount of wafer wasted around the periphery is reduced.

Major factors which place limits on decreasing device dimensions:

- a) **Gate oxide thickness:** If gate oxides thickness becomes thinner about 50 Angstrom, by decreasing dimensions, quantum mechanical tunneling occurs, thus placing a practical bound on oxide thickness.
- b) **Electric field strength:** High electric field strengths are also taken into concern. Voltages upto 5v are placed across 1000 Angstrom silicon dioxide insulating layer. Hence, the electric field of magnitude is

$$E_{1000 \text{ A}^\circ} = \frac{5v}{1000 \text{ A}^\circ} = \frac{5v}{1000 \times 10^{-10} \text{ m}} = 500 \text{ kV/cm}$$

This field is large but less than the break down voltage of silicon dioxide ie. (5 -10 MV/cm)

If the same voltage is applied to 100 A° thickness of oxide layer, the electric field would be very near to the breakdown field for the oxide. So the only option is to decrease the voltage applied across the oxide layer. But this is not an option because the voltage decrease noise effect becomes more significant and thus increasing the chance of errors in the circuit.

Question :

As an OP-AMP required an area $100 \text{ mil} \times 100 \text{ mil}$ and a microprocessor required an area $1 \text{ cm} \times 1 \text{ cm}$. How many of each type of chip can be fabricated on a 5 inch wafer?

Solution

$$N_{\text{opamp}} = \pi R_{\text{opamp}}^2 / A^2 = \pi (2.5 \text{ in})^2 / (0.1 \text{ in})^2 = 1963$$

$$N_{\mu\text{p}} = \pi R_{\mu\text{p}}^2 / A^2 = \pi (2.5 \text{ in})^2 / (1 \text{ cm})^2 = \pi (2.5 \text{ in})^2 / (.39 \text{ in})^2 = 126$$

If the yield for the OP-AMP is 98% and that for the microprocessor is 30% compare the average no. of good chips per wafer of each device that can be anticipated.

Solution

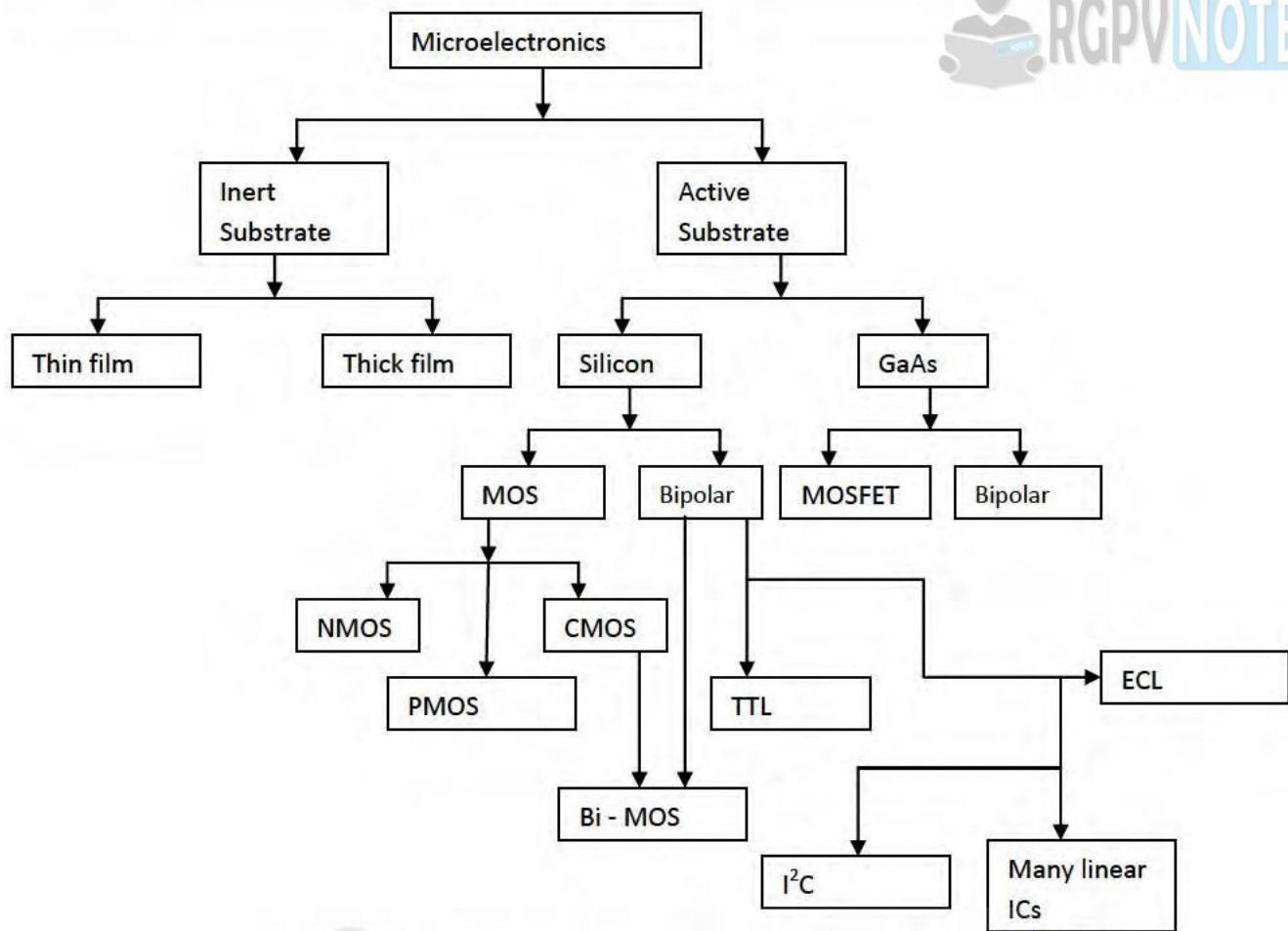
$$N_{\text{opamp effective}} = (0.98) \times (1963) = 1923$$

$$N_{\mu\text{p effective}} = (0.3) \times (126) = 37$$

The Microelectronics Field

Microelectronics is a subfield of electronics. As the name suggests, microelectronics relates to the study and manufacture (or micro fabrication) of very small electronic designs and components. These devices are typically made from semiconductor materials. Many components of normal electronic design are available in a microelectronic equivalent. These include transistors, capacitors, inductors, resistors, diodes and insulators and conductors can all be found in microelectronic devices. Unique wiring techniques such as wire bonding are also often used in microelectronics because of the unusually small size of the components, leads and pads. This technique requires specialized equipment and is expensive.

Block diagram of types of major processes used in IC fabrication is as follows:



1st division in major process type occurs between active and inert substrate. High volume ICs utilizes active substrates whereas low volume ICs uses inert substrates. Inert substrates are also used in hybrid ICs.

Inert substrate utilizes two types of processes- thick and thin film processes. These processes are used in producing resistors with attractive temperature characteristics. This is difficult to achieve with standard active substrate processes. Active substrate is a silicon or doped silicon or gallium arsenide (GaAs). Two types of silicon processes have evolved:

(a) BJT process use BJT as basic active device. This process offers potential for operation at very high frequencies and offers some performance advantages as large transconductances, which are benefits in most linear application. Power dissipation in BJT ICs are quite high. Examples: TTL, ECL etc.

(b) MOS process use MOSFET (or IGFET) as basic active device. MOS process is divided into 3-categories: NMOS, PMOS and CMOS. Basic devices in MOS processes are p-channel and n-channel MOSFETs. PMOS refers to the MOS process that uses only p-channel FETs. NMOS refers to the MOS process that uses only n-channel FETs. PMOS are rarely used than NMOS because the mobility of p-type material is poorer than that of n-type material. CMOS process refers to the MOS process that simultaneously provides both n and p channel devices.

In digital applications, MOS devices offer very low static power consumption. In analog applications, circuit complexity is reduced in CMOS process rather than using NMOS or PMOS process. Applications of CMOS process are memories, interfacing, microprocessors, basic logic functions and a host of linear and mixed linear and digital applications.

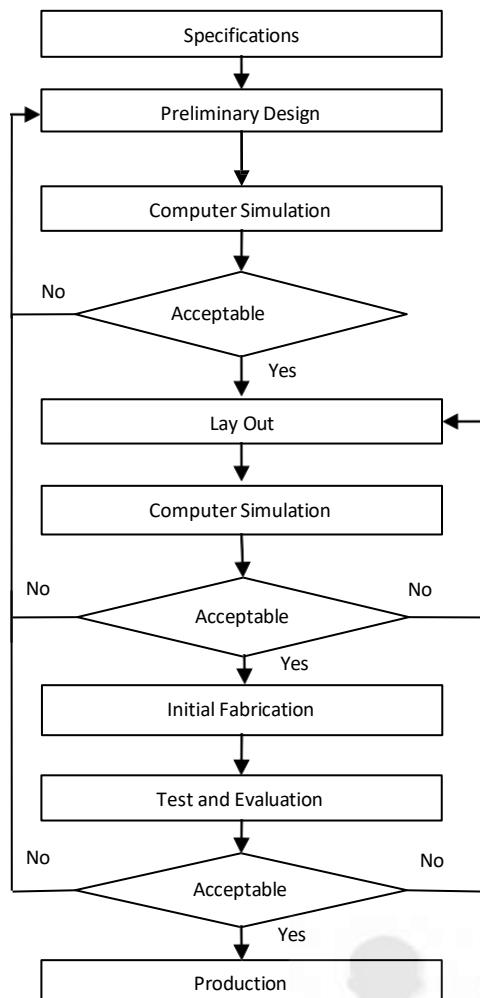
Bi-MOS is a process combining bipolar and MOS devices in a single process. This is complex and expensive process.

Some processes also include thin film components with MOS and or bipolar devices, but the expense associated with adding the thin film layer is acceptable only in specific applications.

IC Design Process:

Two approaches to IC design:

- Bottom up approach: Design starts at the transistor level or gate level and design subcircuits which are then interconnected to realize the required functionality.
- Top-down approach: Designer decomposes the system level specifications into groups and subgroups of simpler task. The lowest level task are implemented in silicon and tested.



Starting point is a set of design specifications.

: Preliminary designs are based upon simple models of devices or subcircuits. Simple models are typically at the behavioral or logical level for digital circuits and component or device level for analog circuits.

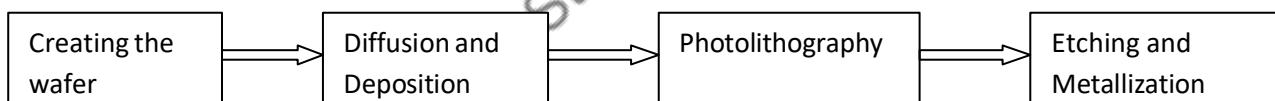
: Computer simulation is used to verify the performance of preliminary design.

: Once the preliminary design is accepted, the actual layout takes place. A good overall floor plan is obtained. The floor plan contains all major busing and cell (subcircuits) placement information as well as I/O pad designations.

: Additional computer simulation on subcircuits after layout is undertaken. These simulations will provide the parasitic effects associated with the layout. Parasitics tend to degrade the performance and causes delays.

: Following an acceptable computer simulation of the entire circuits, the circuit is committed to fabrication. In evaluation process of complicated designs, subcircuits and/or test structures are often fabricated early in the design process to provide modeling information and to verify functionality of subcircuits. A single error in the circuit design, simulation or layout makes circuit totally non-functional. Based upon the expected evaluation, either the circuit is released to production or the appropriate steps of the design process is re-entered.

IC Production Process, Processing Steps



Wafer preparation:-

A wafer is a thin slice of semiconducting material, such as a silicon crystal, upon which microcircuits are constructed by doping (for example, diffusion or ion implantation, etching and deposition of various materials. Wafers are cut out of silicon ingots (a single crystal silicon) from which wafers are cut using diamond saws.

Oxidation:-

Process where growth of an oxide layer on the surface of substrate. Oxidation process produces silicon dioxide. It serves as a good insulator between the substrate and whatever is placed upon it. Oxidation of silicon is achieved by heating silicon wafers in an oxidizing atmosphere. As an alternative to oxidation, the SiO_2 layer can be applied by CVD. Other types of doped deposited oxides such as phosphosilicate glass (PSG) are also used as insulating layers on top of polysilicon in fabricating ICs. This doping helps reduce sharp boundaries introduced during etching of polysilicon.

Masking:-

IC masks are high-contrast photographic positives or negatives, used to prevent light from striking a photosensitized wafer during a photolithographic process. The masks are made of glass covered with a thin film of opaque material. Electron beam method is used to generate the actual patterns directly onto the final masks. This method produces the best quality masks and is used for small geometries. Another method used to generate masks is laser beam pattern generator.

Photolithography:-

The regions of dopants, polysilicon, metal and contacts are defined using masks. Places covered by the mask, ion implantation not occur or the dielectric or metal layers might be left intact. When mask is left or absent, implantation can occur or dielectric or metal could be etched away.

The patterning is achieved by a process called photolithography. Figure 02 shows the photo masking with negative resist . Photolithography steps are:

Ist step : By use of photoresist materials on the wafer, we have to define the area of interest where we want material to be present or absent.

2nd step: This photoresist wafer is subjected to selective illumination through photomask.

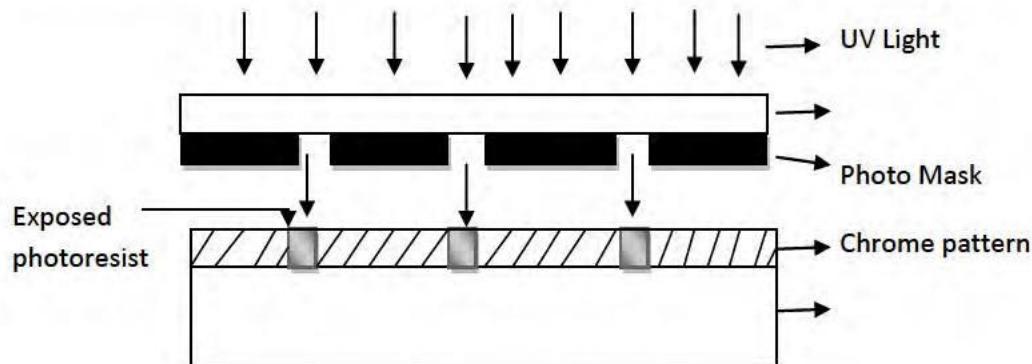


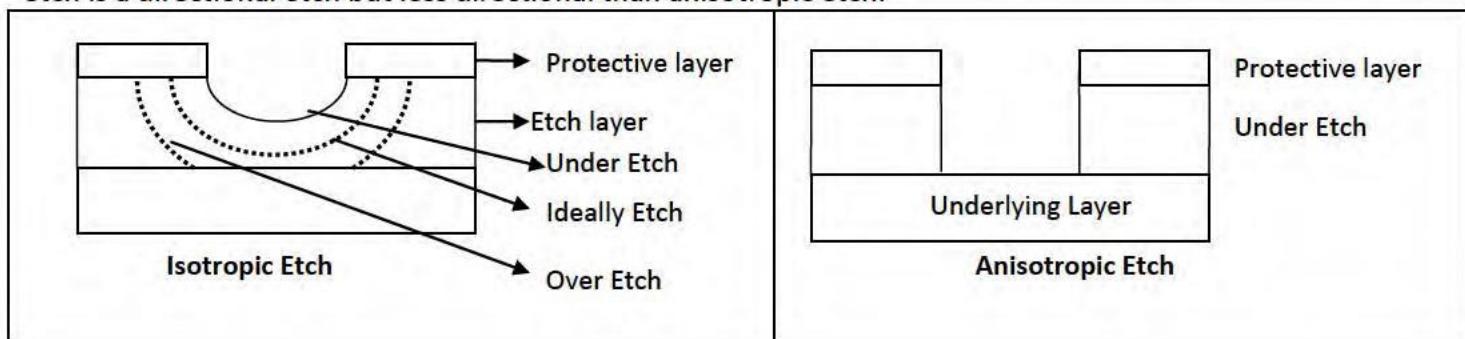
Figure:02: Photomasking with a negative resist (Photolithography process)

Photo mask is constructed with chromium covered quartz glass. UV light is used to expose the photoresist. Photo mask has chrome where light should be blocked. Positive photoresist is initially insoluble and when exposed to UV light becomes soluble. Reverse of negative photoresist.

Deposition:- Films of various materials must be applied on the wafer during processing. Films that are deposited include insulators, resistive films, conductive films, dielectrics, n and p type materials Available technologies consist of physical vapor deposition (PVD), chemical vapor deposition (CVD), electrochemical deposition (ECD) and molecular beam epitaxial (MBE) and more recently, atomic layer deposition (ALD) among others.

Removal or Etching Process:-Etching is removing of unwanted materials from the surface of the substrate. The physical characteristics of the surface are changed by etching. The chemical used for etching reacts with the unprotected areas on the wafer .

2-types of etching: Wet etching called chemical etching uses liquids etching agents. Dry etching also called ion etching. This technique includes sputter etching, ion beam etching and plasma etching. Dry etching is widely used. Figure03, Characteristics of etches: (a) Isotropic etch is non directional etch. Problem with isotropic etch is that effective opening and the undercut will be larger than desired. (b) Anisotropic etch is a directional etch. Has an abrupt edge. Problem for applying subsequent layers uniformly across the edge. (c) Preferential etch is a directional etch but less directional than anisotropic etch.



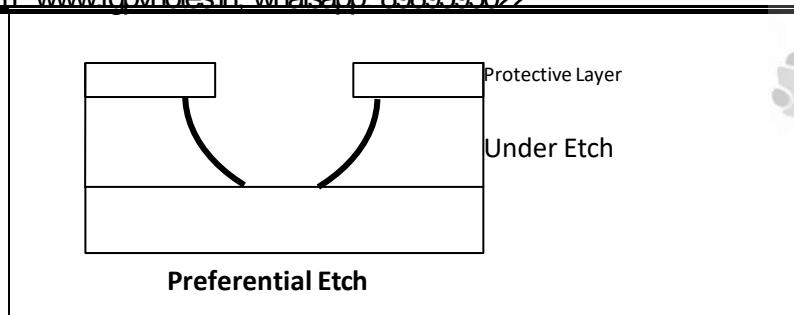


Figure:03: Characteristics of etches

Diffusion:

Diffusion refers to the controlled forced migration of impurities into the substrate. This plays a major role in the performance of the IC and is affected by temperature and time during the processing. Methods of diffusion: A solid deposition layer or a gaseous layer above the surface can be used as a source of impurities. Impurities can be bombarded to the substrate so that they actually become lodged inside the substrate very near the surface. This is ion implantation, very accurate control of impurity concentrations but causes damage to the crystal surface.

Impurities typically diffuse both vertically and laterally from the surface at comparable rates.

Contacts and Metallization:

Contact cuts are made to source, drain and gate according to the contact mask. These holes are etched. Aluminium is used for wires but tungsten can be used as a plug to fill contact holes. Metallization is a process of building wires to connect the devices. Metal films are useful for interconnections that carry large currents. Nonmetallic films (polysilicon) are widely used for conductors and interconnections when current flow is small. Polysilicon is good conductor when heavily doped and good resistor when lightly doped. Polysilicon is often used for gates of MOSFETs and electrodes for capacitors. Polysilicon can be deposited over silicon dioxide.

Packaging and Testing

After processing, the integrated circuits are tested and packaged. The first step in the testing process generally involves a process verification to make sure that the process parameters are within the tolerance acceptable for the product. To facilitate this verification, test plug containing special test structures specially designed for this purpose are included on the wafer at several locations in place of the regular circuits themselves. A wafer prober is used to make mechanical contact with the test plugs so that the electrical measurements can be made. Assuming the process parameters are within tolerance, the individual dies are automatically probed and electrically tested. After probing, the wafer is scribed with wafer saw, and dies are separated. Following separation the individual dies are die attached or die bonded. These bonding wires typically (01 mil range) made up of either gold or aluminum are used. After wire bonds are complete, the packages are formed and a final electrical test is completed.

Proper packaging technology is critical to the success of the chip development. Package issues have to be taken into consideration in early stages of chip development. Considerable effort on a worldwide basis is focused on the packaging problem.

SEMICONDUCTOR PROCESSES:

Three basic processes used for the fabrication of monolithic ICs containing active devices, these are NMOS, CMOS and Bipolar processes. The NMOS and CMOS processes are quite similar in that both have the FET as the basic active device. In NMOS, n-channel MOSFETs are available as the active devices whereas in CMOS both n & p channel devices are available. In bipolar process, the basic active device is BJT. Higher speeds are currently available with the bipolar process than for the NMOS and CMOS processes. The hybrid process combines thin and/or thick film passive components that are on one or more separate substrates with active devices from a separate substrate onto a common carrier. This makes hybrid ICs quite expensive. For applications that require precise and temperature stable passive components, the hybrid process often offers a practical solution.

MOS PROCESSES:

The principle of operation of the MOS transistor at DC and low frequency provide a insight into the MOS process itself.

Operation of MOSFET:

Operation of n-channel enhancement MOSFET and typical output characteristics are shown in figure 04 and 05 .

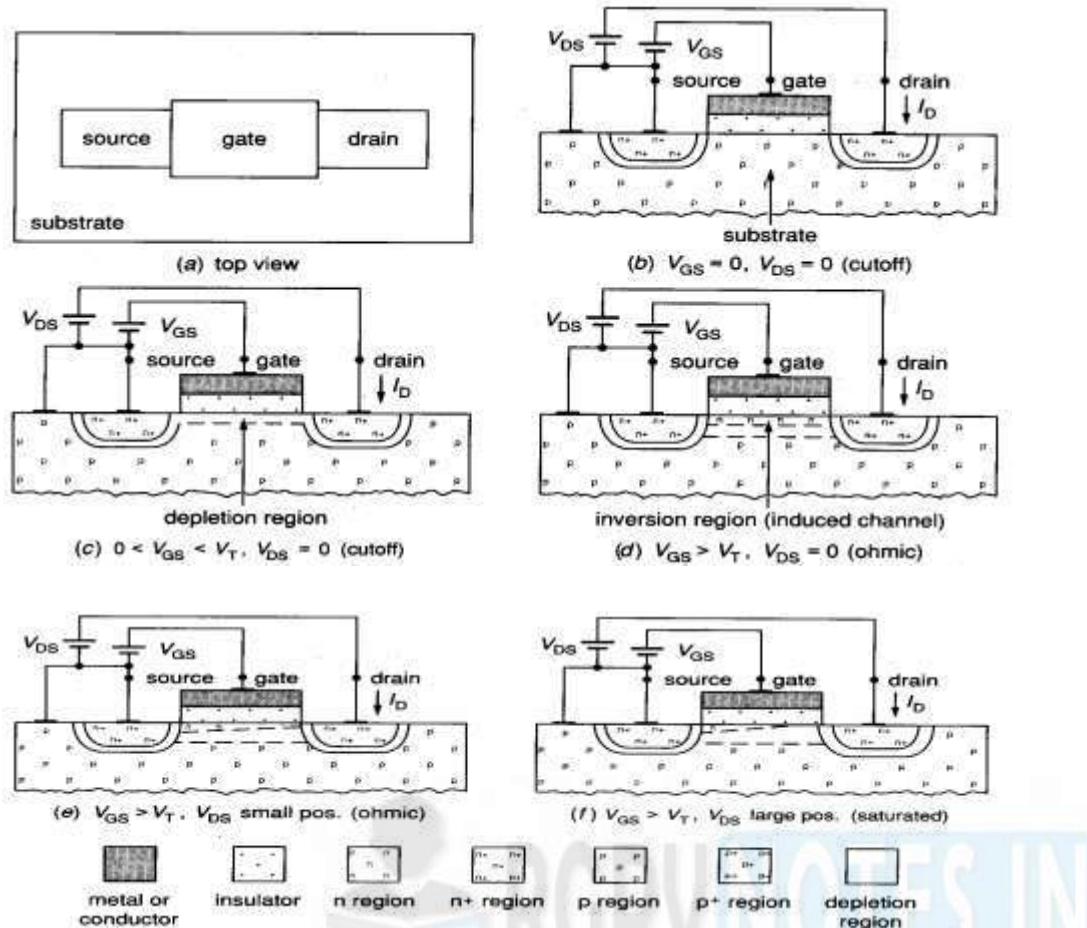


Figure:04: Operation of n-channel MOSFET

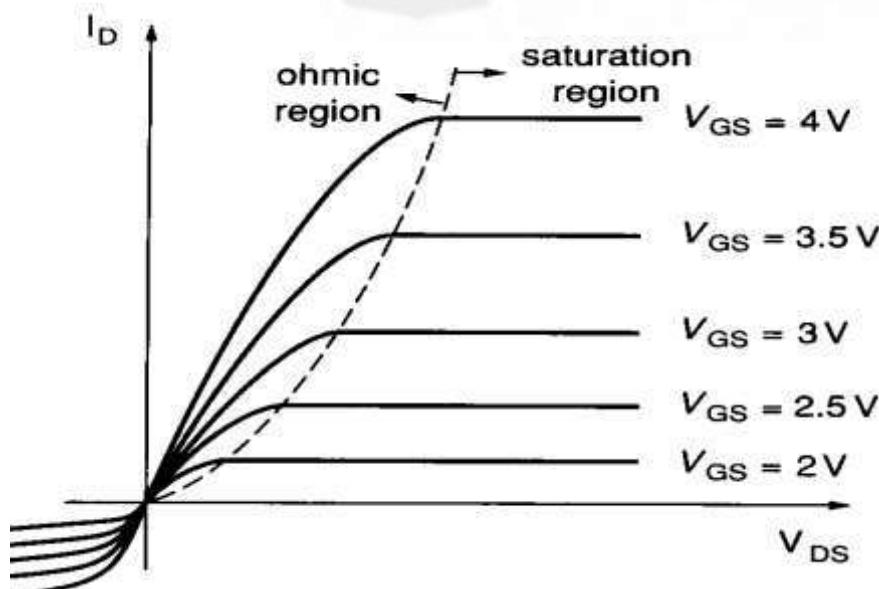


Figure:05: Typical output characteristics

When gate-source voltage is greater than the threshold voltage, the MOS transistor is said to be operating in the ohmic region prior to the pinching of the channel and in the saturation region when the channel is pinched off. If the gate-source voltage is less than the threshold voltage, almost no drain or source current will flow even when a bias is applied to the drain and source contacts. This is the case of cutoff. The relationship between I_D , V_{DS} and V_{GS} for a MOSFET, the output characteristics is shown in figure 05.

The value of threshold voltage is determined by the concentration of the p-type impurities in the substrate. If some n-type impurities are added to the region under the gate near the surface of the substrate, the threshold voltage will decrease. An n-channel device with positive threshold voltage is termed as an

enhancement MOSFET and those with a negative threshold voltage are termed as depletion MOSFETs. MOS devices formed in a p-substrate (or tub) and thus having n-type drain and source diffusions and n-type channel are termed n-channel transistors.

NMOS PROCESS:

A generic double-polysilicon self aligned silicon gate NMOS enhancement/depletion process follows. Polysilicon (a good conductor) gates are used instead of metal. FETs with polysilicon gates are also called MOS transistors or MOSFETs.

The devices that are available in the process are: 1) n-channel enhancement and depletion MOSFETs. 2) Capacitors and resistors

The method of physical constructing each of these components in this process and interconnecting them is shown in figure 06.

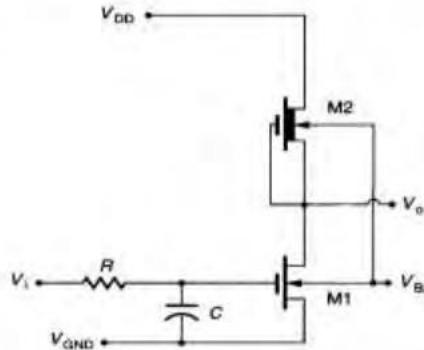


Figure:06: A simple NMOS circuit

For NMOS process, the starting point is a polished p-type silicon disc (500μ). A SiO_2 layer of 1000 Angstroms thick is first added to the entire wafer using Oxidation process. On top of this, a layer of Si_3N_4 (about 1500 Angstrom thick) is applied by the CVD process.

Now application of photoresist, MASK#01 (moat or n^+ diffusion mask) is used to pattern the surface. Mask defines in photoresist the drain, source and channel regions of all transistors as well as other regions where n^+ implants are desired. After exposure, development removes the photo resist layer in the areas that are not to be moat. A top and cross sectional view is shown in figure 07.

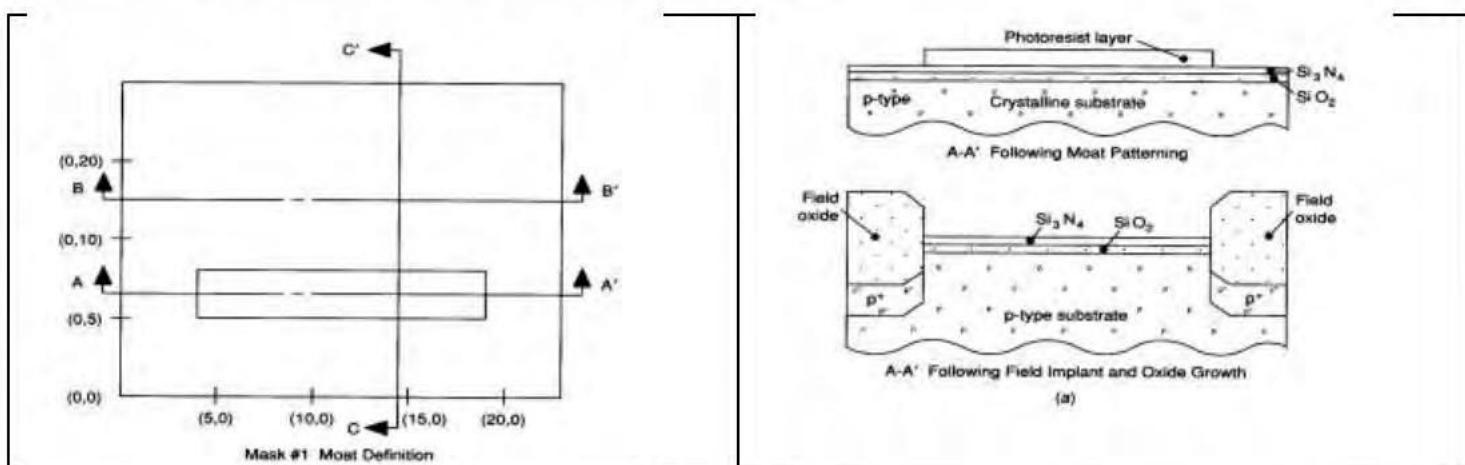


Figure 07: MASK#01

The Si_3N_4 is then etched from the areas not protected by the photoresist. A high energy implant of p-type impurities is then applied to entire wafer. The remaining photoresist protects the moat region from this implant. This heavy implant is used to raise the threshold voltage in the anti moat region (often called field) and to provide electrical isolation between adjacent devices. After this field and drive in diffusion, the remaining photoresist is stripped.

A thick layer of SiO_2 about 10,000 Angstroms is grown by the oxidation process over the wafer. This layer is formed in the field but not on the Si_3N_4 . This thick field oxide layer is termed a local oxidation layer. Following removal of Si_3N_4 and thin layer of SiO_2 under the Si_3N_4 is stripped and another SiO_2 layer is grown. Now a n-type implant over the entire wafer can be applied (optional) to set the threshold voltage of the enhancement devices.

Now a heavier selective implant is required in the regions that are used to serve as the channels of depletion transistors. To achieve this a second layer of photoresist is applied to the entire wafer, and the second mask, MASK#02, (implant mask) is used to pattern the photoresist so that only the channel regions of depletion transistors are unprotected.

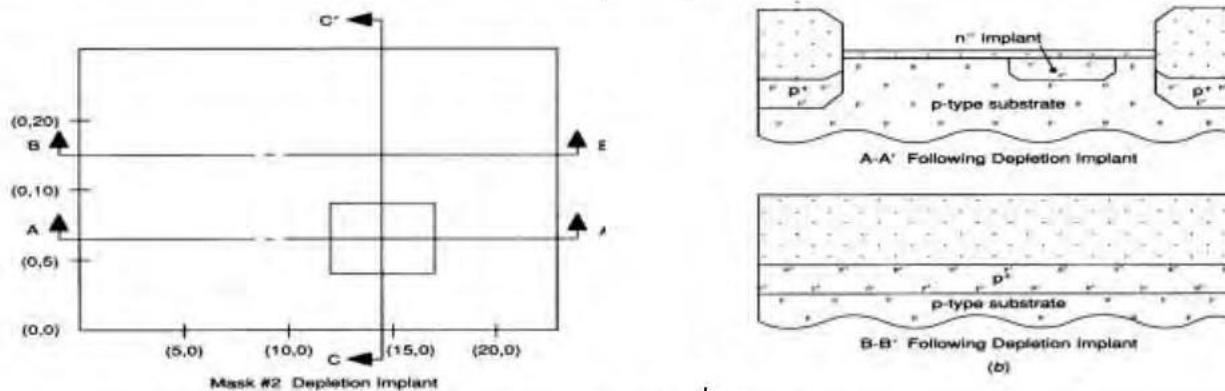


Figure 08: MASK#A & MASK#2

Another n-type implant is used to make the exposed regions n-type with the remaining photoresist serving as an implant mask. After stripping the photoresist, the wafer is as shown in figure 08.

Direct contact between the polysilicon layer and moat are termed as buried contact. To make buried contact, the thin gate oxide must be patterned and etched to remove the insulating SiO_2 layer and create paths through which the following polysilicon layer can contact the moat. MASK#A (figure 08) is used to pattern the buried contact paths.

After stripping of thin oxides from the previous stage, a uniform thin layer of SiO_2 is grown on the surface of wafer. Stripping and regrowing of oxides provides better control of the critical gate oxide thickness. A layer of polysilicon (POLY-I) about 2000 Angstroms thick is then deposited on the surface of the entire wafer. This is covered with photo resist, patterned with MASK#3 (figure 09), and etched to remove unwanted POLY-I. The POLY-I layer is used as gates for both enhancement and depletion transistors, as a plate for capacitors, as a conductors and for resistors. Note that POLY-I layer is over gate oxide in cross section AA' and above field oxide in cross section BB'.

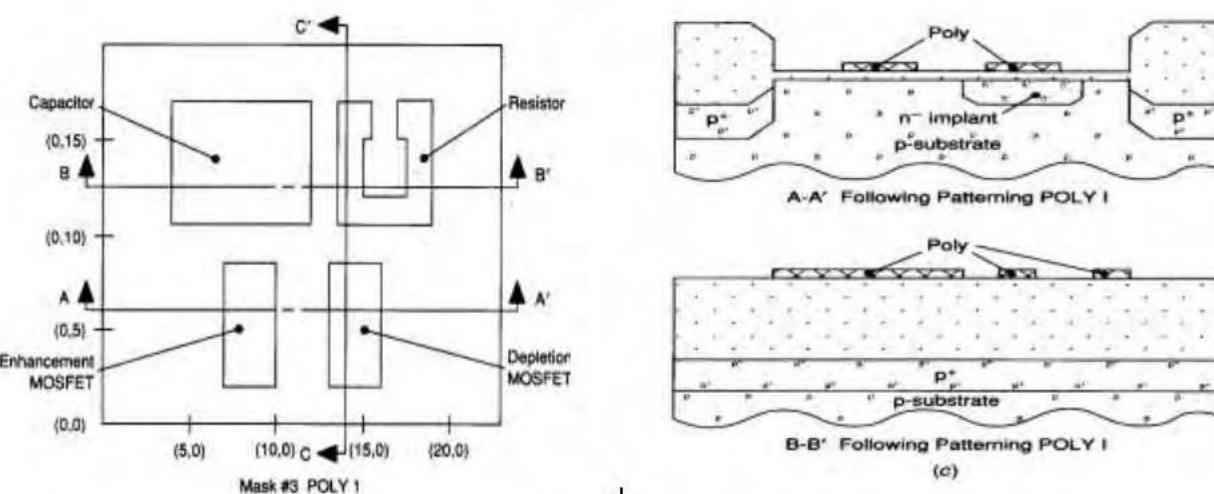


Figure 09: MASK#03

The remaining uncovered thin layer of SiO_2 are stripped and another thin layer of SiO_2 is again grown. This servers as dielectric for POLY-I & POLY-II capacitors, and an insulators for POLY-I & POLY-II crossovers, and as gate oxide for transistors that use the POLY-II layer as the gate. A second layer of POLY-II is then deposited followed by photoresist and patterned with MASK#B (Figure 10).

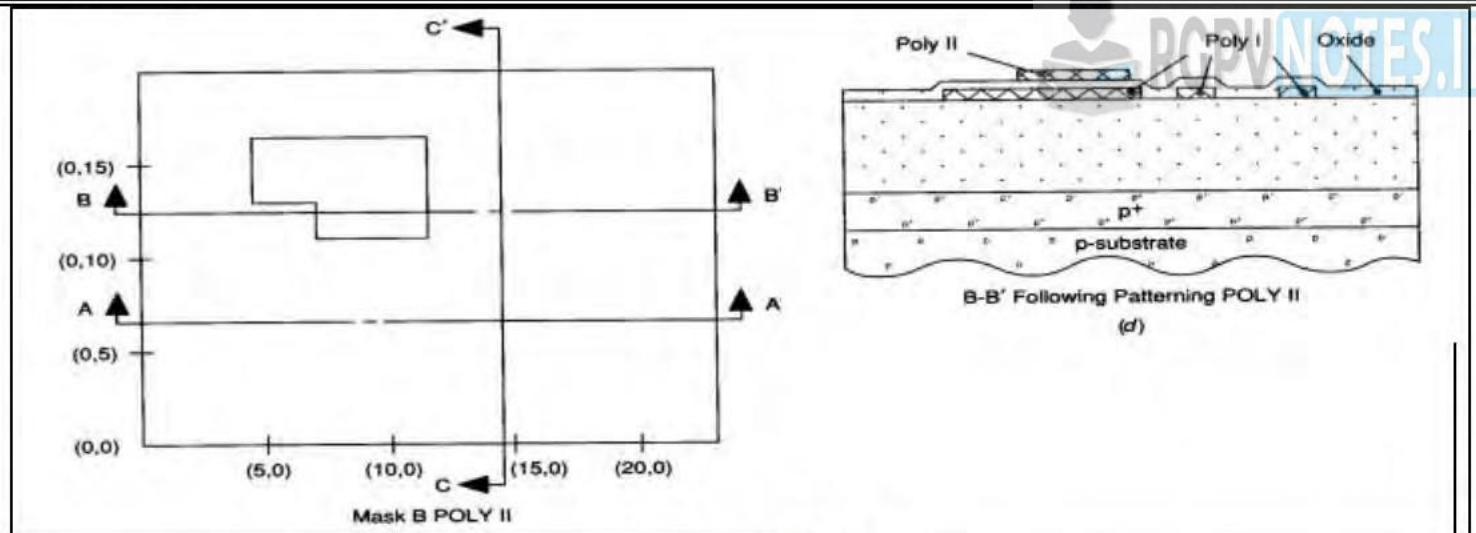


Figure 10: MASK # B

Another insulating layer is deposited over the wafer which serves as an insulator between uppermost polysilicon layer and the subsequent metal layer. The entire wafer is covered with photoresist and MASK#4 (figure 11) is used to pattern the contact openings for the purpose of obtaining the electrical contacts. After photoresist is develop, an etch that attacks the insulating layer but does not affect the polysilicon makes the required openings.

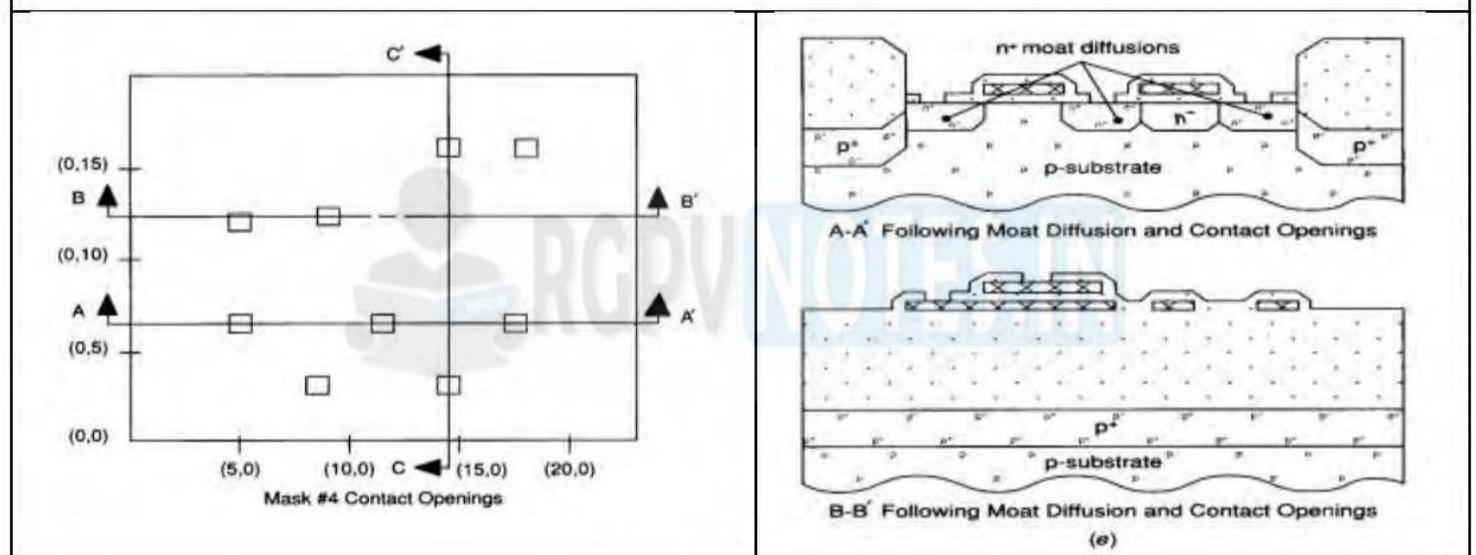


Figure 11: MASK#04

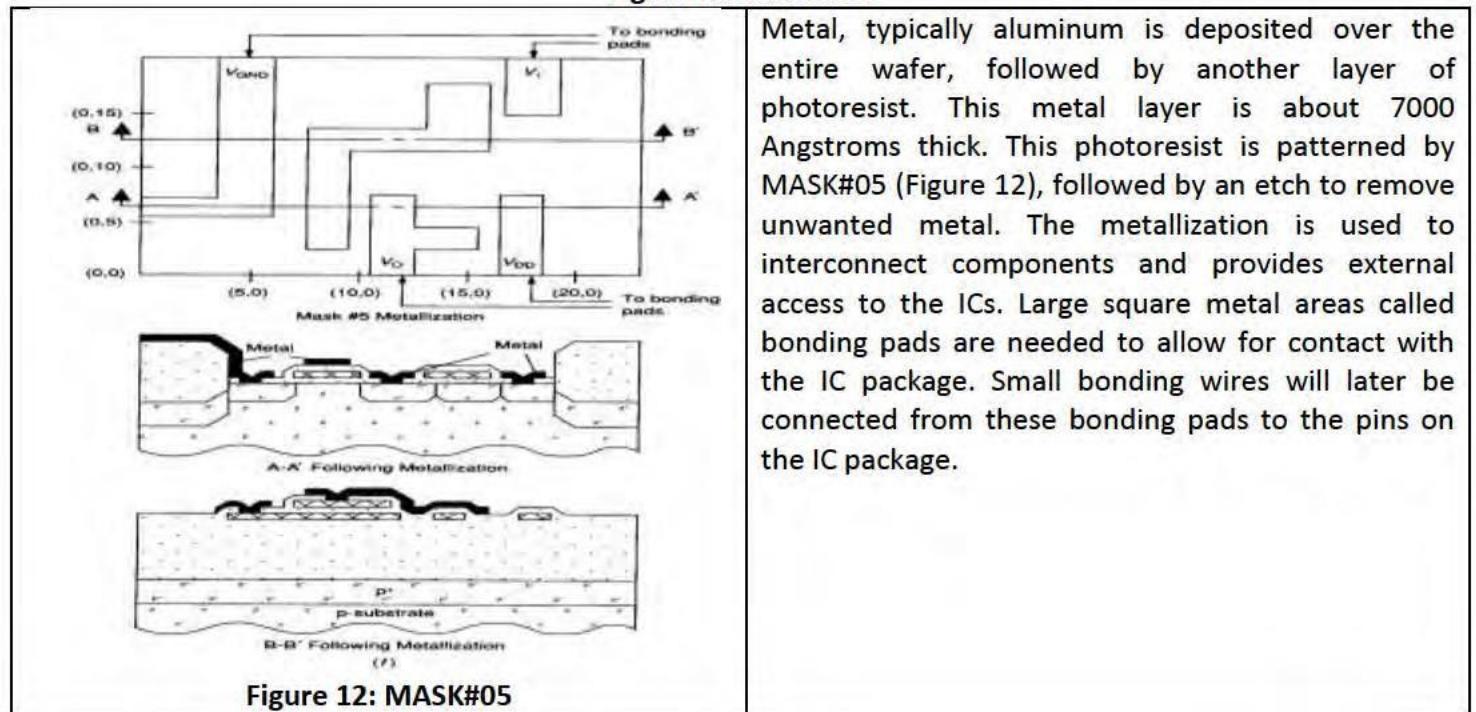


Figure 12: MASK#05

Metal, typically aluminum is deposited over the entire wafer, followed by another layer of photoresist. This metal layer is about 7000 Angstroms thick. This photoresist is patterned by MASK#05 (Figure 12), followed by an etch to remove unwanted metal. The metallization is used to interconnect components and provides external access to the ICs. Large square metal areas called bonding pads are needed to allow for contact with the IC package. Small bonding wires will later be connected from these bonding pads to the pins on the IC package.



CMOS PROCESS:

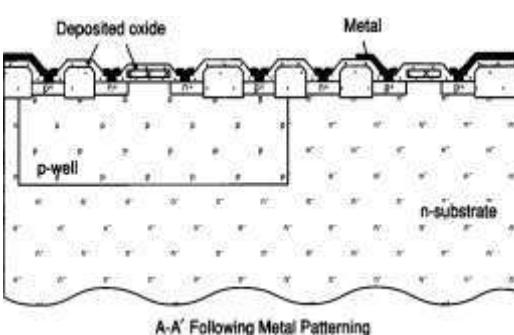
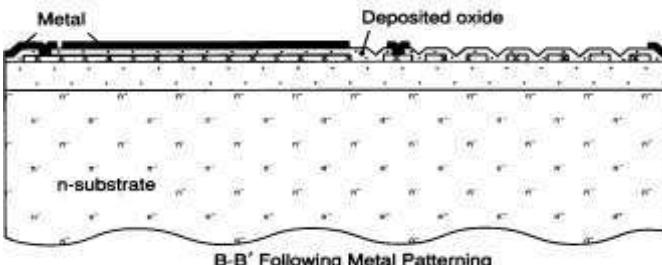
A generic single polysilicon silicon gate, p well, n-substrate CMOS process follows. The devices available in the CMOS process under consideration are as follows: 1) n-channel & p-channel MOSFETs 2) Capacitors, resistors & diodes 3) npn & pnp bipolar transistors.

Diodes and bipolar transistors are often considered as parasitic components and are not used in circuit design.

CMOS Process steps	Process description for CMOS process
Clean Wafer	
Grown Thin Oxide	
Apply Photoresist	
Pattern P-Well (MASK#01)	
Develop photoresist	
Deposit and diffuse p-type impurities	
Strip photoresist	
Strip thin oxide	
Grown thin oxide	
Apply thin layer of Si_3N_4	
Apply photoresist	
Pattern Si_3N_4 . Active area define. (MASK#02)	
Develop photoresist	
Etch Si_3N_4	
Strip photoresist	
Grow field oxide	
Strip Si_3N_4	
Strip thin oxide	
Grow gate oxide	
Polysilicon deposition (POLY01)	
Apply photoresist	
Pattern polysilicon (MASK#03)	
Develop photoresist	
Etch polysilicon	
Strip photoresist	
Apply photoresist	
Pattern P-channel drains and sources and P^+ guard rings (P-well ohmic contacts) (MASK#04)	
Develop photoresist	
P^+ implant	
Strip photoresist	
Apply photoresist	
Pattern N-channel drains and sources and N^+ guard rings (Top ohmic contacts to substrate) (MASK#05)	
Develop Photoresist	



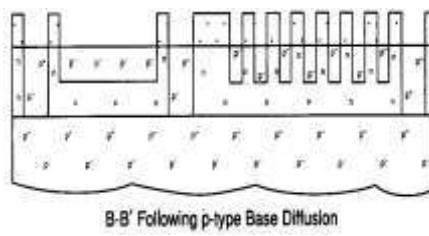
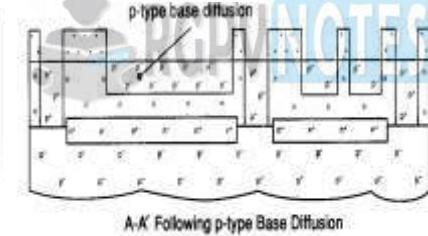
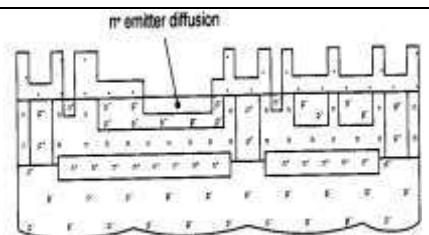
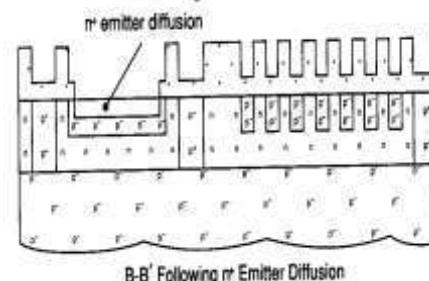
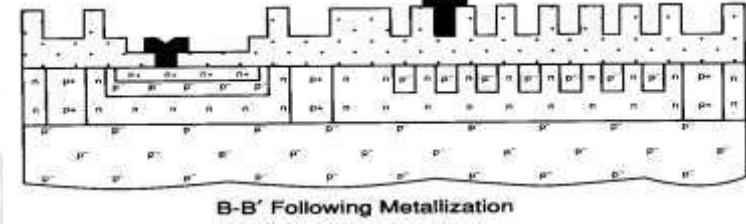
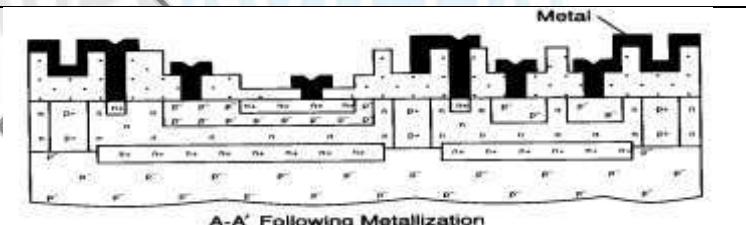
n ⁺ implant
Strip photoresist
Strip Thin Oxide
Grow oxide
Apply photoresist
Pattern contact openings (MASK#06)
Develop photoresist
Etch oxide
Strip photoresist
Apply metal
Apply photoresist
Pattern metal (MASK#07)
Develop photoresist
Etch metal
Strip photoresist
Apply photoresist
Apply Passivation to provide electrical contact.
Pattern pad openings (MASK#08)
Develop photoresist
Etch passivation
Strip Photoresist
Assemble, package and Test.



BIPOLAR TECHNOLOGY :-

The basic active devices in the bipolar process are the npn and pnp transistors. These devices are constructed with 3-layers of n or p type semiconductor materials. These layers are fabricated either vertically or horizontally. The components available in the bipolar process are: 1) npn & pnp bipolar transistors. 2) Resistors, capacitors, diodes & zener diodes 3) JFETs- not available in all bipolar process. Figure 13 shows the Vertical and lateral transistors in the bipolar process.

BIPOLAR Process steps	Process description for BIPOLAR process
Clean Wafer (P-type)	
Grown thin oxide	
Apply photoresist	
Pattern n⁺ buried layer MASK#01	
Develop photoresist	
Deposition and diffusion of n ⁺ buried layer	
Strip photoresist	
Strip oxide	
Grow epitaxial layer (n-type)	
Grow oxide	
Apply photoresist	
Pattern P⁺ isolation regions MASK#02	
Develop photoresist	
Etch oxide	
Deposition and diffusion of P ⁺ isolation	
Strip photoresist	
Grow oxide	
Apply photoresist	

Pattern base regions MASK#03		
Develop photoresist		
Etch oxide		
Deposition and diffusion of p-type base		
Strip photoresist		
Grow oxide		
Apply photoresist		
Pattern n-type emitter regions MASK#04		
Develop photoresist		
Etch oxide		
n ⁺ deposition and diffusion		
Strip photoresist		
Grow oxide		
Apply photoresist		
Pattern contact opening MASK#05		
Develop photoresist		
Etch oxide		
Strip photoresist		
Apply metal		
Apply photoresist		
Pattern metal MASK#06		
Develop photoresist		
Etch metal		
Strip photoresist		
Apply passivation		
Apply photoresist		
Pattern pad openings MASK#07		
Develop photoresist		
Etch passivation		
Strip photoresist		
Assemble, package and test		

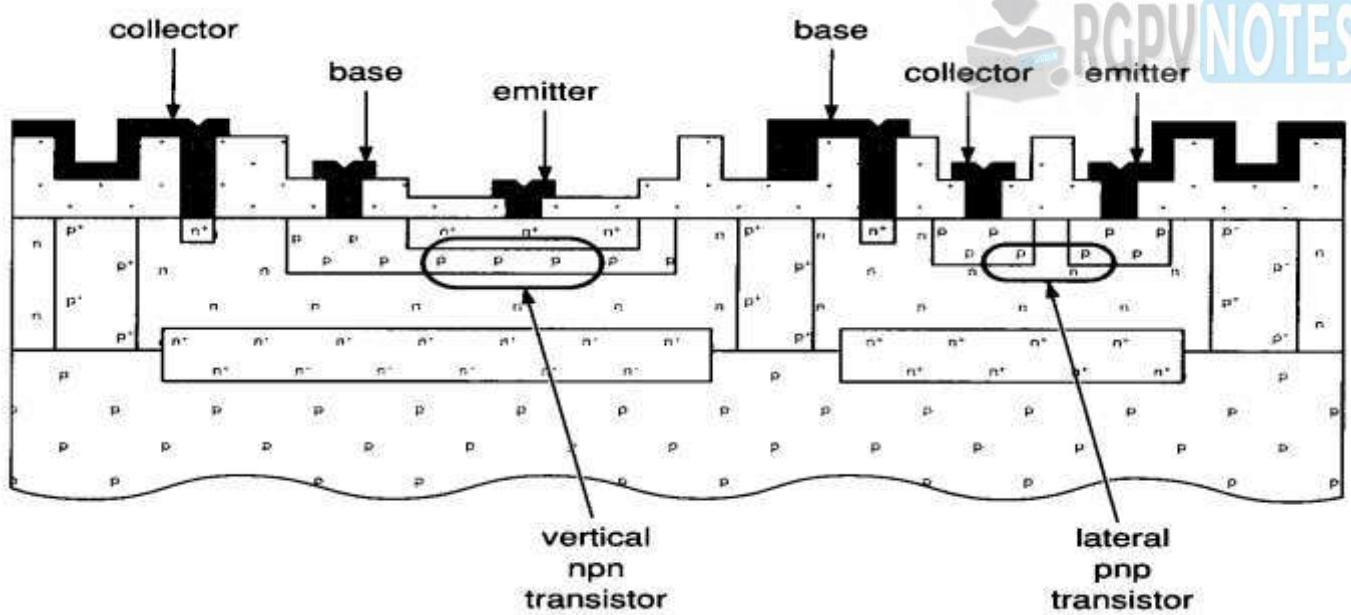


Figure 13: Vertical and lateral transistors in the bipolar process

Hybrid Technology

An integrated circuit design which involve attaching of two or more IC dies, along with some discrete components in some cases, in a single package is called Hybrid IC. Hybrid IC are typically more costlier than monolithic IC. Hybrid circuit containing discrete components occupy less area as compared with conventional discrete component approach. These ICs major applications is in analog signal processing applications such as high resolution A/D and D/A converters and precision active filters.

Disadvantage in standard MOS and bipolar processes is that, this processes have limited on the development of monolithic ICs for precision continuous analog signal processing applications in terms of tolerances, temperature dependence, and area induced components value of resistors and capacitors.

To offer reasonable tradeoffs between area required and components value, thick and thin film passive components are used which have reasonable tolerances, easily trimmable, have acceptable temperature coefficients that can be tailored for tracking.

THICK FILM CIRCUITS: Thick film technology requires a considerable area compared to monolithic ICs. This technology can be used for high power applications and can be applied at relatively high frequencies upto 01 GHz. Resistor and capacitor components are available in thick film process along with conducting interconnections.

Layers of different material are screened onto the insulating substrate. These materials are used for resistors and conductors as well as dielectric for capacitors. The number of resistive layers varies but practical restrict to three. Typical thickness of these layers (called pastes or inks) is about 20μ . Thick film process is advantageous for resistor fabrication. Although capacitors are often included, the electrical characteristics of thick film capacitors are not outstanding and capacitance density is quite low. Minimum conductor width in thick film process is about 250μ , and minimum resistors widths are about 1250μ .

Screening, involves forcing the paste through small holes in a tightly stretched piece of fabric called a screen, typically made up of steel. Screen with a grid spacing ranging from 100 to 300 filament/ inch are typical. Holes in the screen are plugged by a mask. A squeegee is used to forced the ink through the unrestricted area. Following screening, each layer is fired to harden it. Inks are available with sheet resistances that satisfy the equations: $1\Omega/\square < R < 10M\Omega/\square$. This large latitude of ink characteristics allows for a wide range of resistor values. Adjusting the length is a convenient means of establishing values. Long thick film resistors are avoided because they develop hot spots and are difficult to trim. Hot spots are caused at the regions where the resistive layer is little thinner or narrower than surrounding regions. This causes increased in resistance in this small regions which increases power dissipation which causes heating, which typically further increases resistance and power dissipation. Heating causes deterioration of the film layer at this point. Deterioration causes device failure. Short, wide resistors are also be avoided because of the inability to accurately specify the size of the resistor. Thumb rule for the allowable width (W) / Length (L) ratio for a rectangular resistor is

$1/10 < W/L < 3$. Good practice to make the resistors large if the area is available to minimize the edge roughness effects, increased power dissipation capability, and make trimmer easier.

Thick film resistors can be trimmed by with a laser or by abrasion. These trims are accurate.

A capacitor is constructed by screening a conductive layer, followed by a dielectric, followed by another conductor. The dielectric is applied in two coats to minimize pinholes, which would short the capacitor plates together. Thick film capacitor is a parallel plate capacitor and capacitance is given by

$$C = \frac{\epsilon_0 \epsilon_r A}{t} \quad \text{where } \epsilon_r \text{ is relative dielectric constant, } \epsilon_0 = 8.854 \text{ pF/m, } A \text{ is area of capacitor plate \& } t \text{ is the dielectric thickness.}$$

Thick film capacitors can be trimmed by abrasively removing the part of the upper plate along with some dielectric.

THIN FILM CIRCUITS: The components available in thin film are resistors and capacitors. For conductors film thickness are 100 to 500 angstrom, 100 to 2000 angstrom for resistors and 3000 angstrom region thickness for dielectric of capacitors. The sheet resistance range for thin film resistors is $50\Omega/\square$ to $250\Omega/\square$. The minimum feature sizes for the thin film components are comparable to those of MOS and bipolar process. Hot spots are not major problem with thin film resistors because films are more uniform and requires smaller current flow. Thin film circuits are expensive to design because of need of sophisticated equipments required for photolithographic process, film depositions and etching. Used in telecommunication circuits at low frequencies but also find applications at high frequencies upto 30GHz. Thin film can be accurately trimmed by a laser.

Design Rules and Process Parameters:

Design rules are well documented specifications consists of minimum width of features (resistors, conductors etc), minimum spacing allowable between adjacent features, overlap requirements and other measurements compatible with the given process. Factors such as mask alignment, mask non-linearities, wafer wrapping, oxide growth profile, lateral etch, optical resolution and their relationship with performance and yields are considered when specifying the design rules for the process. The rules were derived under the assumption that large circuits with many devices sized at the minimum allowable levels must have good performance and high yields. Consider the design rules and process parameters as a set of constraints within which the circuit designer must work.

CMOS design rules:

Two sets of design dimensions are specified. First corresponds to the 3μ CMOS process provided by MOSIS. Second is in terms of the scaling parameter, λ , which characterizes the feature size of the process. The feature size (minimum poly width, active width, and metal width) is 2λ .

Design rules 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.

Lambda rules, which specify the layout constraints in terms of a single parameter 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. It must be emphasized, however, that most of the submicron CMOS process design rules do not lend themselves to straightforward linear scaling. The use of lambda-based design rules must therefore be handled with caution in sub-micron geometries. In the following, we present a sample set of the lambda-based layout design rules devised for the MOSIS CMOS process and illustrate the implications of these rules on a section a simple layout which includes two transistors.



Design rules for a typical p-well CMOS process

Sr. No.	Rule number	Description	Dimensions	
			Microns	Scalable
1	p-well Layer			
	1.1	Width	5	4λ
	1.2	Spacing to well at different potential	15	10λ
	1.3	Spacing to well at same potential	9	6λ
2	Active (Diffusion) Layer			
	2.1	Width	4	2λ
	2.2	Spacing to active	4	2λ
	2.3	P+ active in n-subs to p-well edge	8	6λ
	2.4	n+ active in n-subs to p-well edge	7	5λ
	2.5	n+ active in p-well to p-well edge	4	2λ
	2.6	p+ active in p-well to p-well edge	1	λ
3	Poly			
	3.1	Width	3	2λ
	3.2	Spacing	3	2λ
	3.3	Field poly to active	2	λ
	3.3	Poly overlap of active	3	2λ
	3.4	Active overlap of poly	4	2λ
4	P+ Select			
	4.1	Overlap of active	2	λ
	4.2	Space to n+ active	2	λ
	4.3	Overlap of channel	3.5	2λ
	4.4	Space to channel	3.5	2λ
	4.5	Space to P+ select	3	2λ
	4.6	width	3	2λ



Device Modeling:

DC Models, Small Signal Models, MOS Models, MOSFET Models in High Frequency and small signal, Short channel devices, Sub threshold Operations, Modeling Noise Sources in MOSFET's, Diode Models, Bipolar Models, Passive component Models

2.1 Introduction to device modeling :

Device models used for developing design equations, hand analysis and initial computer simulations. DC model are useful for biasing and large signal analysis whereas AC models are useful for small signal steady state analysis.

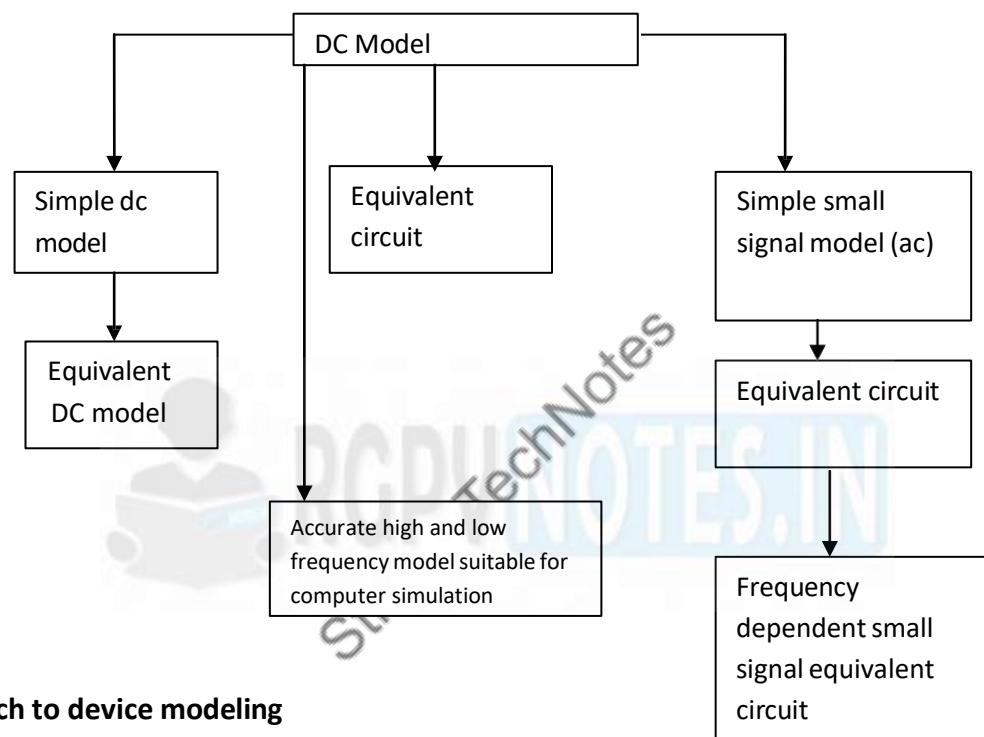


Fig.2.1 Approach to device modeling

Starting point for modeling for BJT and MOSFET will be a DC model as shown in fig.2.1. From this DC model, a linear small signal model and equivalent simplified ac and dc circuits will be derived. The models will be expanded to provide a better agreement between the theoretical and experimental results for use in computer simulation.

2.2 DC Models:

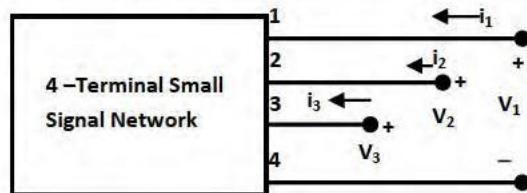
The DC model of a device is a mathematical or numerical relationship that relates the actual terminal voltage and current of the device at DC and low frequency. The dc model should be valid over a large range of terminal voltage and current. It is valid at frequency where the difference between the actual and dc solution is deemed negligible for the problem under investigation.

2.3 Small Signal Models:

Small-signal modeling is a normal analysis technique in electronics circuit which is used to show the performance of electronic circuit contains nonlinear devices with linear equations. A small-signal model is an AC equivalent circuit in which the nonlinear circuit elements are replaced by linear elements whose values are given by the first-order (linear) estimate of their characteristic curve near the bias point. It is relevant to electronic circuits in which the AC signals, the time-varying currents and voltages in the circuit, have a small magnitude compared to the DC bias currents and voltages.

A small signal model takes a circuit and based on an operating point (bias) and liberalize all the components. Nothing changes because the assumption is that the signal is so small that the operating point (gain, capacitance, etc.) doesn't change.

For the linear small signal four-terminal network shown in Figure 2.2 A, one terminal (terminal 4) is selected as a reference. With this reference and the assumption of linearity, it follows by definition that the y-parameter (admittance parameter) relate the terminal voltages and currents by the expressions



$$\begin{aligned} i_1 &= Y_{11}v_1 + Y_{12}v_2 + Y_{13}v_3 \\ i_2 &= Y_{21}v_1 + Y_{22}v_2 + Y_{23}v_3 \\ i_3 &= Y_{31}v_1 + Y_{32}v_2 + Y_{33}v_3 \\ \text{Where } Y_{kj} &= i_k / v_j \quad \text{limit } v_m = 0, m \in \{1, 2, 3\}, \\ m \neq j \end{aligned}$$

Fig.2.2A: Linear small Signal 4-terminal network

Since the small signal voltage and current variables are the time-varying part of the corresponding total terminal voltage and current variables in the parent network relative to the Q-point figure 2.2.B, then it follows that the Y-parameters can be obtained from the large signal variables and thus from the dc model by the expression $Y_{kj} = \partial i_k / \partial V_j$ limit $v_m = \text{quiescent value}$, $m \in \{1, 2, 3\}$.

A small signal equivalent circuit of the multiport network is often found useful for circuit analysis and design. It is easy to verify that the circuit shown in Figure 2.2.C has the same i-v characteristics as the four terminal network of Figure 2.2 A, which is characterized by and hence is electrically indistinguishable.

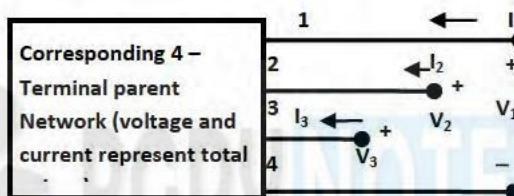


Fig.2.2.B: Parent 4-terminal network

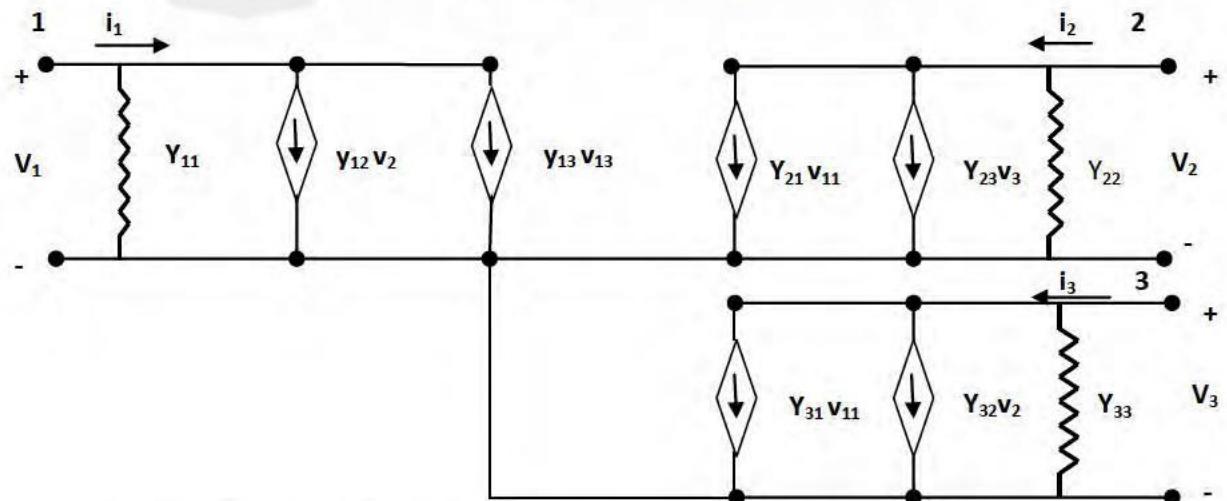
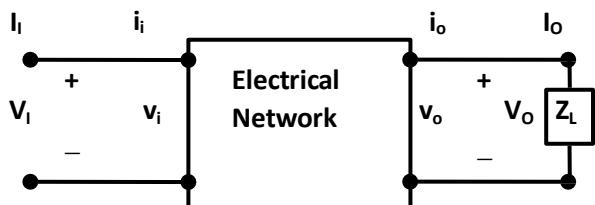


Fig. 2.2.(C) Equivalent circuit for small signal network

Use of DC model in circuit analysis:

Designers are interested in controlling the relationship between two or more electrical port variables. A two port network as shown in fig. 2.3 having both large and small signal port variables. Relationship includes- large Signal transfer characteristics, such as V_o Vs V_i or V_i Vs I_i and small signal transfer characteristics such as, voltage gain $A_v = v_o / v_i$ or input impedance, $Z_{in} = v_i / i_i$.



The dc transfer characteristics of a circuit are obtained by using the dc device models to characterize all devices in the network. To develop the dc performance of simpler networks, it is necessary to make simplifying assumptions on the device models. Without these assumptions mathematical expressions relating the electrical parameters of interest are often complicated that they obscure even the basic functionality of the circuit. The simplified dc analysis is often useful for biasing purposes and can provide useful insight into the small signal operation of the circuit. The low frequency small signal characteristics of a circuit are obtained in two ways- 1) replacing all devices and elements with a linear small signal equivalent circuit in which parameters in the small signal models are a function of the Q-point. Independent dc voltage sources are short circuits and independent dc current sources are open circuits.

2) is based upon dc transfer characteristics. If x_i and x_o are small signal input and output variables and X_i and X_o are corresponding large signal variables, then these variable are related by expression $(x_o / x_i) = (\partial X_o / \partial X_i)$ where the partial derivative of the large signal is evaluated at the Q-point.

The result obtained from direct small signal analysis will be identical to those obtained by differentiation of large signal variables provided that the small signal models of the devices are obtained from the same dc model used to obtain the dc transfer characteristics.

Advantages offered by obtaining the small signal transfer characteristics directly from a small signal analysis rather than from differentiation of the dc variables are-

- 1) Since small signal equivalent circuit is linear, the analysis of the small signal equivalent circuit is typically less involved than the dc analysis.
- 2) More accuracy is practically attainable with the direct small signal analysis.
- 3) Frequency response of a network is expressed in terms of the small signal electrical parameters.

2.4 MOS Models:-

The MOS transistor is a majority-carrier device in which the current in a conducting channel between the source and drain is controlled by a voltage applied to the gate. In an nMOS transistor, the majority carriers are electrons; in a pMOS transistor, the majority carriers are holes. The behavior of MOS transistors can be understood by first examining an isolated MOS structure with a gate and body but no source or drain. A simple MOS structure. The top layer of the structure is a good conductor called the gate. Early transistors used metal gates. Transistor gates soon changed to use Polysilicon, i.e., silicon formed from many small crystals, although metal gates are making resurgence at 65 nm. The middle layer is a very thin insulating film of SiO₂ called the gate oxide. The bottom layer is the doped silicon body. The figure shows a p-type body in which the carriers are holes. The body is grounded and a voltage is applied to the gate. The gate oxide is a good insulator, so almost zero current flows from the gate to the body.

In Figure 2.4 (a) a negative voltage is applied to the gate, so there is negative charge on the gate. The mobile positively charged holes are attracted to the region beneath the gate. This is called the accumulation mode. In Figure 2.4 (b), a small positive voltage is applied to the gate, resulting in some positive charge on the gate. The holes in the body are repelled from the region directly beneath the gate, resulting in a depletion region forming below the gate. In Figure 2.4 (c), a higher positive potential exceeding a critical threshold voltage V_t is applied, attracting more positive charge to the gate. The holes are repelled further and some free electrons in the body are attracted to the region beneath the gate. This conductive layer of electrons in the p-type body is called the inversion layer. The threshold voltage depends on the number of dopants in the body and the thickness t_{ox} of the oxide. It is usually positive, as shown in this example, but can be engineered to be negative.

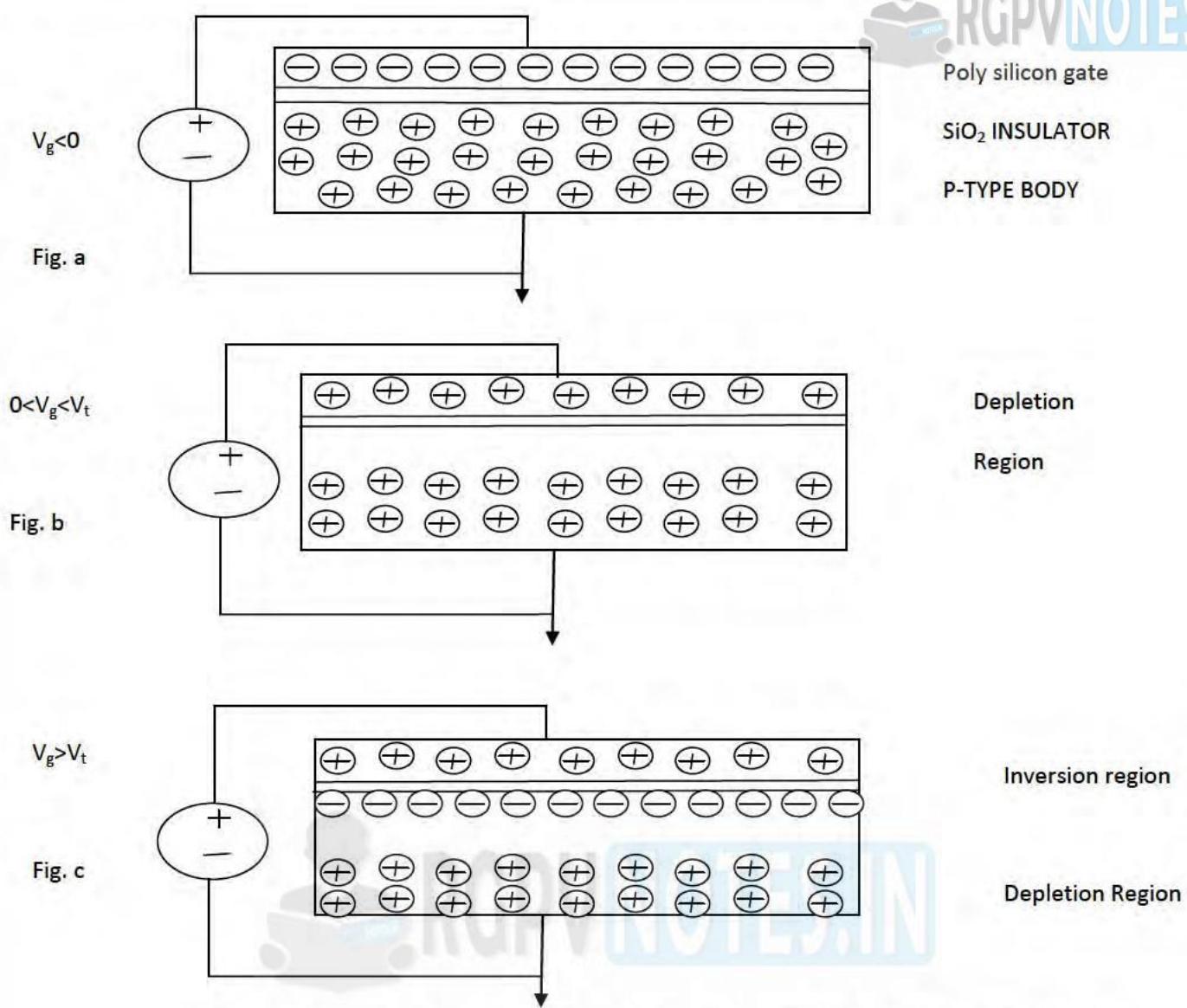
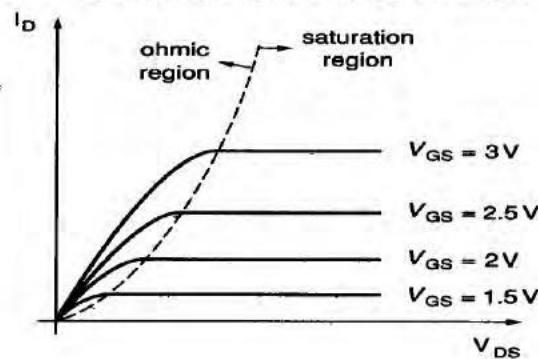


Fig.2.4 :MOS structure demonstrating (a) Accumulation, (b) Depletion, and (c) Inversion

DC MOSFET Model:-

It is desired to obtain a mathematical model of the MOSFET that will accurately predict the experimental characteristics over a wide range of geometrical and process parameters. The starting point will be the dc model which is introduced by SAH (first-order model) relating the I-V characteristics for the nMOS transistor.



In cutoff region ($V_{GS} < V_t$), there is no channel and almost zero current flows from drain to source.

In other regions, the gate attracts carriers to form a channel. The electrons drift from source to drain at a rate proportional to the electric field between these regions.

Thus we can compute currents if the amount of charge in the channel and the rate at which it moves are known.

$$\text{Thus charge in the channel } Q_{\text{channel}} = C_g (V_{gc} - V_t) \quad (1)$$

Where C_g is the capacitance of gate to channel & $(V_{gc} - V_t)$ is the amount of voltage attracting charge to the channel beyond the minimum required to invert from p to n.



Coefficient λ is quite small for long channel devices and increases considerably for short channel devices.

The MOSFET model defined by EQ-02 & 04 are termed as Shichman-Hodges model.

Threshold voltage effects:

So far, we have treated the threshold voltage as a constant. However, V_t increases with the source voltage, decreases with the body voltage, decreases with the drain voltage, and increases with channel length.

Considered a transistor to be a three-terminal device with gate, source, and drain. However, the body is an implicit fourth terminal. When a voltage V_{sb} is applied between the source and body, it increases the amount of charge required to invert the channel, hence, it increases the threshold voltage.

The threshold voltage can be modeled as

$$V_T = V_{TO} + \gamma(\sqrt{\phi_s - V_{BS}} - \sqrt{\phi_s}) \quad \text{--- 05}$$

Where, V_{BS} is the bulk to source voltage; V_{TO} , γ and ϕ_s are process parameters.

V_{TO} is threshold voltage for $V_{BS} = 0$; γ is bulk threshold parameter (0.4 to 1 $V^{1/2}$); ϕ_s is inversion surface potential.

For n-channel transistors the devices are termed as enhancement if $V_{TO} > 0$ and depletion if $V_{TO} < 0$.

At this point, the designer can control V_t with V_{bs} and thus have the second means of controlling the characteristics of MOS transistor. (First means is by controlling W/L ratio)

This V_{bs} dependence, which is controlled through bulk voltage V_{bb} , offers very little flexibility to designers for two reasons: 1. V_t is weakly depend upon V_{bs} & 2. Bulk of all transistors are common in nMOS process, the bulk voltage for entire IC is at single value.

Equivalent circuit for DC operation of the MOSFET is shown in figure

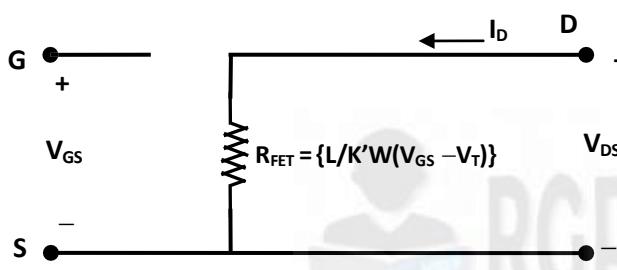


Figure 2.5(a): Ohmic region (small V_{DS})

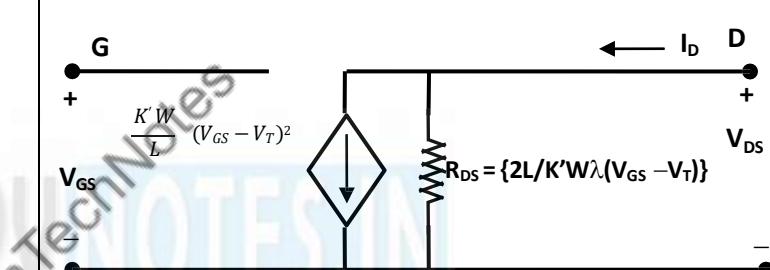


Figure 2.5 (b): Saturation region

First circuit in figure 2.5(a), is valid for small V_{DS} in ohmic region. Model corresponds to this circuit is obtained from expression for I_D in ohmic region.

By assuming $V_{DS}/2$ small and negligible as compared with $V_{GS} - V_t$. With this assumption a relationship between I_D and V_{DS} arises and results in an equivalent FET resistance as:

$$R_{FET} \cong \frac{L}{WK'(V_{GS} - V_T)} \quad \text{--- 6}$$

MOSFET is often used in the ohmic region as a voltage variable resistor (VVR) by using V_{GS} as the controlling voltage.

The equivalent circuit in 2.5 (b) is valid in the saturation region. As V_{GS} increases towards $V_{GS} - V_t$, the relationship between I_D and V_{DS} becomes non-linear. The effect of R_{DS} in the equivalent circuits are negligible in most DC applications.

Problem: Using NMOS process parameters determine the output voltage for the circuit shown- if $R = 15 \text{ k}\Omega$.

$$V_{TO} = 0.75 \text{ V}$$

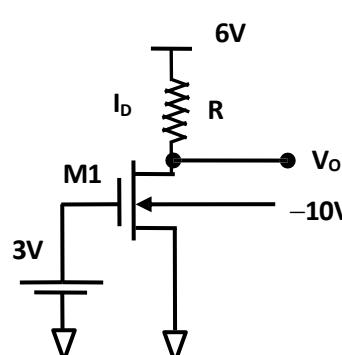
$$W = 30 \mu \text{m}$$

$$L = 3 \mu \text{m}$$

$$\gamma = 0.4 \text{ V}^{1/2}$$

$$\phi = 0.6 \text{ V}$$

$$K' = 24 \mu \text{A/V}^2$$



Answer: (a) Initially the region of operation of M1 is determined. So, obtain V_T .

$$V_T = V_{TO} + \gamma(\sqrt{\phi - V_{BS}} - \sqrt{\phi}); V_T = 0.75 + 0.4 V^{1/2}(\sqrt{0.6} - \sqrt{0.6}); V_T = 1.74 \text{ V.}$$

Since, $V_{GS} = 3 \text{ V}$,

it can be concluded that M1 is in either ohmic or saturation region. So, assuming M1 is in the saturation region.

Then neglecting λ effect-

$$I_D = \frac{K'W}{2L} (V_{GS} - V_T)^2 = 24 \mu\text{A}/V^2 \cdot \frac{30\mu}{3\mu} \cdot \frac{1}{2} (3V - 1.74V)^2 = 190.5 \mu\text{A}.$$

$$\text{Thus, } V_O = 6V - I_D R = 3.14 \text{ V}$$

Since, $V_{DS} = V_O$, $V_{DS} > V_{GS} - V_T$. Thus the initial assumption as the region of operation was valid and $V_O = 3.14$ volt is correct.

(b) Maximum value of R (R_{MAX}) for saturation region operation will be that value which makes $V_{DS} = V_{GS} - V_T$

$$\text{So, } 6V - 190.5 \mu\text{A} R_{MAX} = 24.9 \text{ k}\Omega.$$

Problem. Obtain the small signal model of the two port device shown in figure 01, for $V_1 > 0$ and $V_2 > 0$. Assume a 2-port device is characterized by the equations, $I_1 = 0$ and $I_2 = hV_1^2(1 + \lambda V_2)$

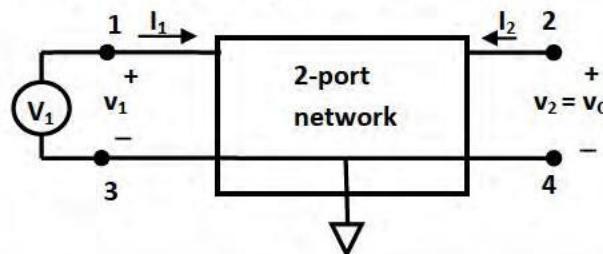


Fig.01

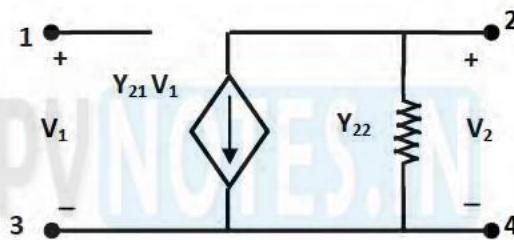
Answer: By definition, the y -parameter relates the terminal voltages and currents by the expressions as

$$y_{21} = \partial I_2 / \partial V_1 = 2hV_1(1 + \lambda V_0)$$

$$y_{22} = \partial I_2 / \partial V_2 = \lambda h V_1^2$$

$$y_{11} = y_{12} = 0$$

Thus small signal device model is



2.5 Small Signal MOSFET model:

Small signal MOSFET model can be obtained directly from DC model. Since there are 3-regions of operation identified in the DC model, hence there are different small signal models for the MOSFET corresponding to each of these regions.

In cutoff region, $I_D \approx 0$, resulting in a minor small signal model. Since MOSFET is not biased for small signal operation in ohmic region. Most small signal applications employ the MOSFET biased in the saturation region. In saturation region-

$$I_D = \frac{K'W}{2L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS})$$

$$\text{And, } I_G = I_B = 0.$$

For notational convenience, g , d & b will be used to designate the gate, drain and bulk nodes of the MOSFET.

Parameter, $Y_{dg} = g_m$ is a dominant parameter in the model.

$$Y_{dg} = g_m = \partial I_D / \partial V_{GS} = \frac{K'W}{L} [(V_{GS} - V_T)] (1 + \lambda V_{DS})$$

Upon relating $V_{GS} - V_T$ to the quiescent drain current and noting that the λ effects are typically negligible in this model,

$$\text{it follows that } g_m \cong \frac{2K'W}{L} |I_{DQ}|$$

$$g_{mb} = \eta g_m \quad \text{where, } \eta = \gamma / 2\sqrt{(\phi - V_{BSQ})}$$

$$g_{ds} = \lambda |I_{DQ}|$$

$$r_{ds} = 1/g_{ds}$$

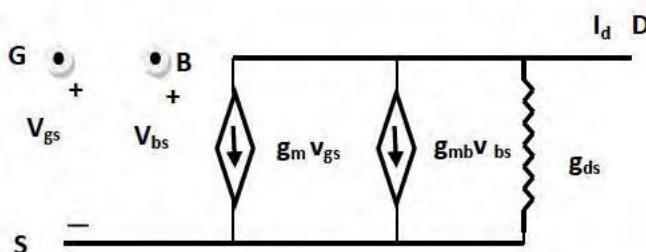


Figure 2.6: Small Signal Model equivalent circuit of MOSFET

From the figure 2.6. this model consists of two dependent current generators $g_m v_{gs}$ and $g_{mb} v_{bs}$ are proportional to the time varying v_{gs} and v_{bs} . Proportionality constant is the transconductance g_m . Arrow pointing from drain to source indicate a phase change of 180 degree between input and output voltages.

In low frequency model , gate current is zero, input resistance (gate-to-source resistance) is very large.

Noted that, if v_{bs} is not equal to zero, any ac signal that appears on the bulk will also modulate the drain current. Since the current $g_m v_{gs}$ dominates the drain current, it should be apparent that the small signal MOSFET is inherently a good transconductance amplifier.

2.6 MOSFET Models in High Frequency :

At high frequencies both the dc and small signal models of the MOSFET introduced in the previous sections are generally considered inadequate. These limitations are to a large extent attributable to the unavoidable parasitic capacitances inherent in existing MOS structures. These parasitic capacitances can be divided into two groups. The first group is composed of those parasitic capacitors formed by sandwiching an insulating dielectric of fixed geometric dimensions between two conductive regions. The capacitance of these types of devices remains essentially constant for local changes in the voltage applied to the plates of the capacitor. Assuming the area of the normally projected intersection of the capacitor plates is A and that the distance between the plates is constant with thickness d, then this capacitance is given by the expression

$$C = \epsilon A / d$$

Where ϵ the permittivity of the dielectric material separating the plates.

C_d called the capacitance density.

$$C = C_d A \quad \text{where } C_d \text{ is a process parameter and } A \text{ is a design parameter.}$$

The second group is composed of the capacitors formed by the separation of charge associated with a pn junction. The depletion region associated with the semiconductor junction serves as the dielectric. These junction capacitors are quite voltage dependent. They are typically expressed in terms of the process parameter C_{j0} , which denotes the junction capacitance density at zero volts bias. The capacitance of these devices can be approximated by

$$C = C_{j0} \cdot A / (1 - V_F / \phi_B)^n$$

Where A is the junction area, V_F is forward DC bias voltage , ϕ_B barrier potential typically 0.7V range, n is constant, depending upon the type of junction such as step graded junction, $n = 1/2$ whereas linear graded junction, $n=1/3$.

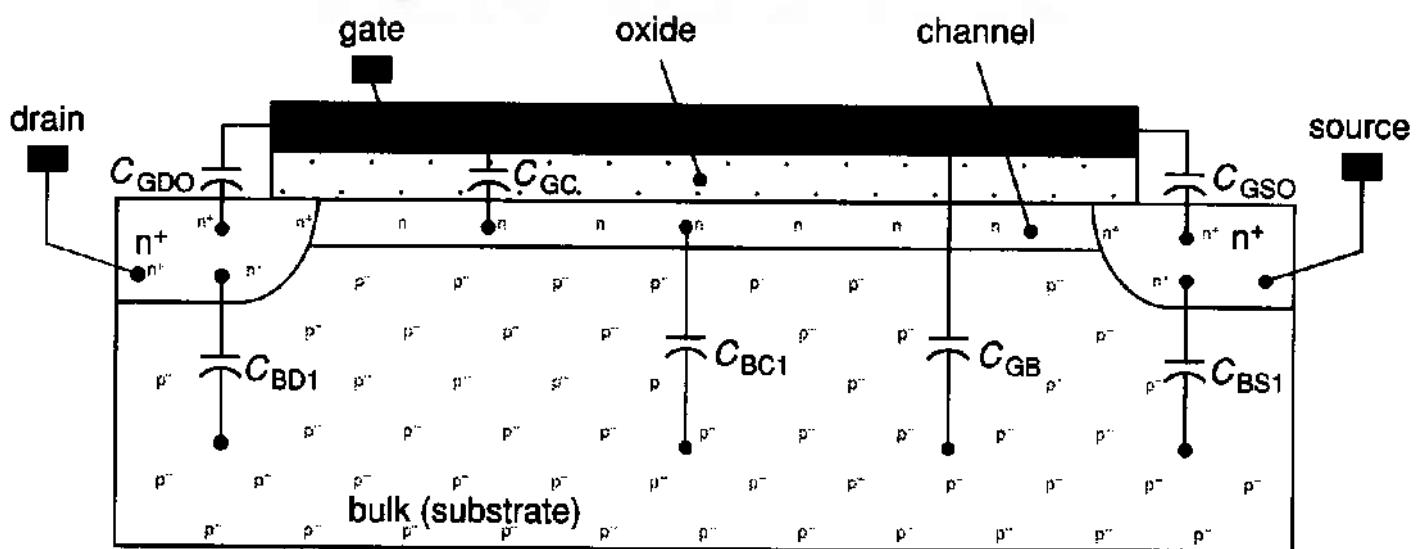


Fig.2.7 : Parasitic capacitors in MOS transistor for n channel device

The parasitic capacitors that dominate the high-frequency behavior of MOS transistors are shown in figure 2.7. The capacitors C_{BD1} , C_{BC1} C_{BS1} are all voltage-dependent junction capacitors. The remaining capacitors are parallel plate capacitors governed by $C = C_d/A$. Some of the parasitic capacitors are actually distributed devices. In addition, some of the capacitor values are operating region dependent. To simplify analysis, a model of the MOSFET in each of the three operating regions containing only lumped nodal capacitors will be presented.

- C_{ox} represents the capacitance density of the gate/oxide/channel capacitor.
- $C_{ox} = (\epsilon_0 \cdot \epsilon_{SiO_2})/t_{ox}$
- Where ϵ_{SiO_2} is the relative permittivity of the silicon dioxide dielectric, which is assumed to be of thickness t_{ox} .



- C_{GDO} and C_{GSO} are gate-drain and gate-source overlap capacitors. Their existence is due to unwanted diffusion under the gate (impurities to create the drain-source regions).
- This diffusion causes a small decrease in effective length of transistors as well as contributing to the parasitic capacitance.
- If L_D represents the distance of the lateral moat diffusion under the gate, then both these parasitic capacitors are nearly rectangular with length W and width L_D , resulting in a capacitance of $C_{GDO} = C_{GSO} = C_{ox} \cdot W \cdot L_D$
- C_{GC} represents the gate-to-channel capacitance. In cutoff region, channel is not formed. So. $C_{GC} = 0$.
- In ohmic region, channel is quite uniform under the entire gate region. So, total gate-to-channel capacitance under ohmic region is given by $C_{GC} = C_{ox} \cdot W \cdot L$
- In saturation region, typically $C_{GC} = 2/3 C_{ox} \cdot W \cdot L$ is modeled as a lumped element between gate and source and remaining $1/3 C_{ox} \cdot W \cdot L$ is neglected.
- C_{GB} represents the gate-to-bulk capacitance. In cutoff region, no inversion, no depletion region, so bulk region extends to the bottom side of the gate oxide layer. So. $C_{GB} = C_{ox} \cdot W \cdot L$
- This capacitance is voltage dependent and decreases with bias with a depletion region present in channel. In saturation and ohmic regions, $C_{GB} = 0$ because of existence of inversion layer.
- C_{BC1} represents the bulk-to-channel junction capacitance. In cutoff region, no inversion, no depletion region, causing vanishing of C_{BC1} . In saturation and ohmic regions, C_{BC1} exist, with value approximated by $C = C_{j0} A / (1 - V_F/\Phi_B)^n$
- C_{BS1} represents C_{BD1} are capacitances of bulk-to-source and bulk-to-drain junction capacitance.
- For short channel transistors (less than 3-5 μ in length), the lateral capacitance associated with the edge of junctions becomes significant. These are called sidewall capacitances. Another source of parasitic capacitance in both small and large devices is the capacitance associated with layout-dependent interconnections.

Comments about each of the parasitic capacitors follow.

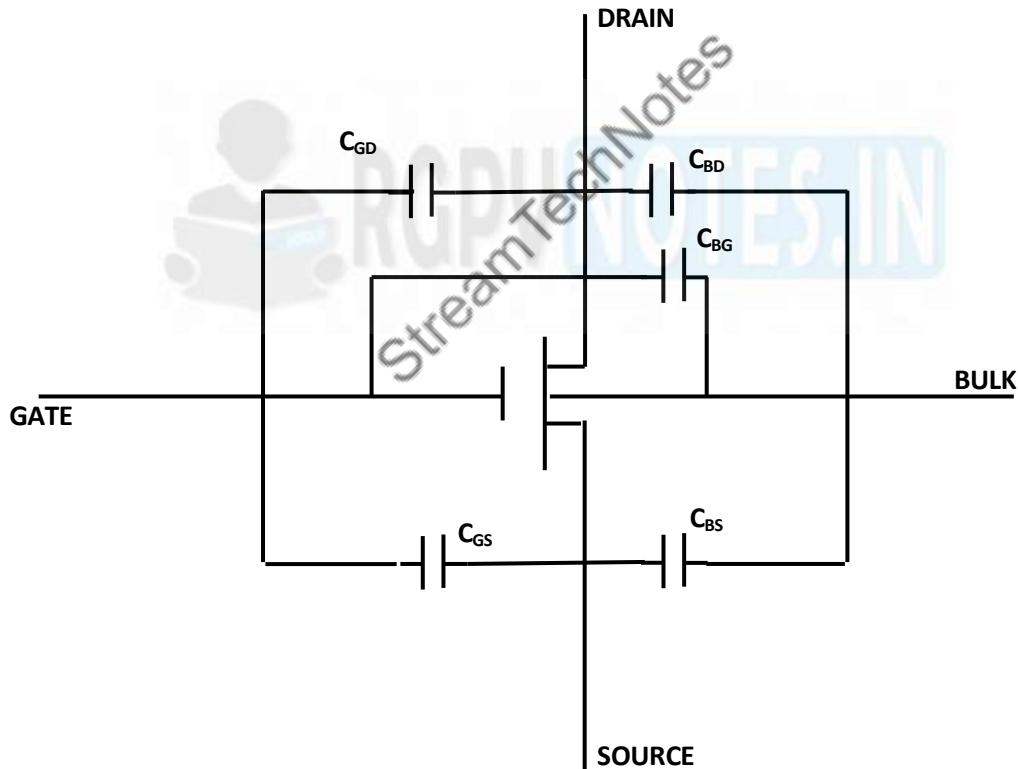


Fig 2.8 : Parasitic Capacitor lumped network

In the above figure show that parasitic capacitance between the source and drain it can be seen from figure that no physical mechanism exists to create such a capacitor.

	Region		
	Cutoff	Ohmic	Saturation
C_{GD}	$C_{ox} WL_D$	$C_{ox} WL_D + (1/2) WL_{Cox}$	$C_{ox} WL_D$
C_{GS}	$C_{ox} WL_D$	$C_{ox} WL_D + (1/2) WL_{Cox}$	$C_{ox} WL_D + (2/3) WL_{Cox}$
C_{BG}	$C_{ox} W L$	0	0
C_{BD}	C_{BD1}	$C_{BD1} + (C_{BC1})/2$	C_{BD1}
C_{BS}	C_{BS1}	$C_{BD1} + (C_{BC1})/2$	$C_{BD1} + (C_{BC1})/(2/3)$

Table 2.1 : Parasitic capacitor value according to operating region

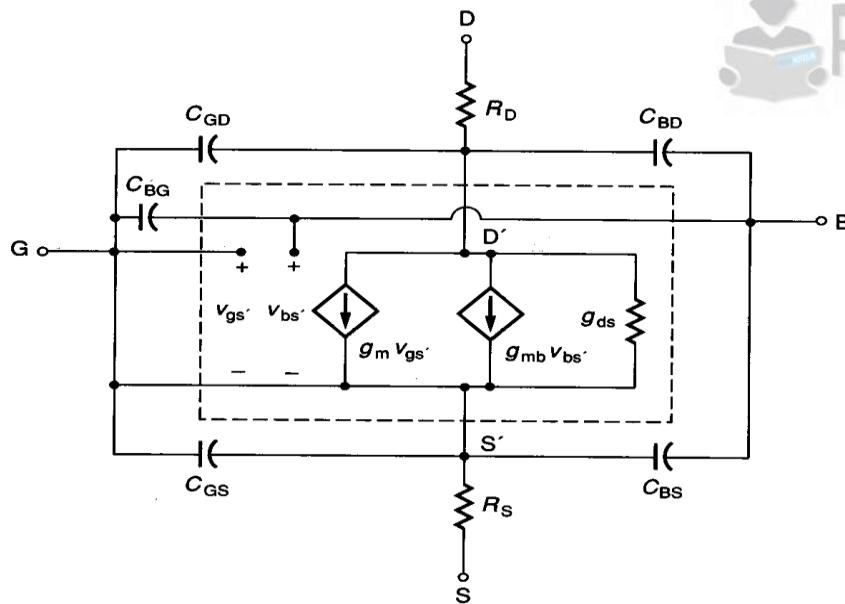


Fig : Parasitic Capacitor small signal equivalent circuit

From the circuit, resistors RD and RS are ohmic resistance in drain and source regions.

For large physical drain and source regions, these resistors become quite large.

For minimum size drain and source regions, these resistances will be in the 50 ohms range.

2.7 Short channel devices :

Transistors with channel length less than 3 to 5 μ are termed as short channel devices. Ratio between lateral and vertical dimensions is reduced. Channel region changes from a uniform thin right-rectangular region to an irregular structure.

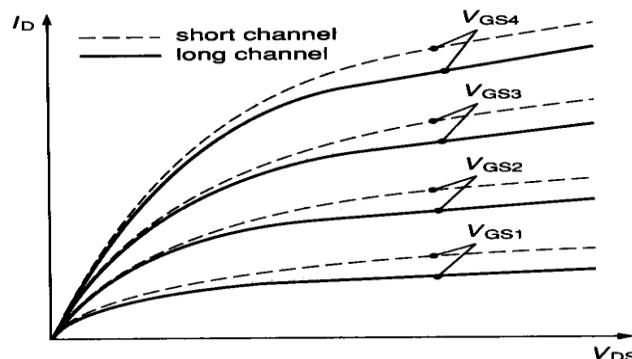
Advantages of short channel transistors over a larger transistors:

Reduced area.

Improvement in speed.

Major disadvantage is the output impedance characteristics deterioration; matching of characteristics of short channel devices are difficult to achieve.

Comparison of short and long channel devices output characteristics:



Short channel effects on MOSFET performance. Solid curve for 5 μ device, dashed curve for 1 μ device.

Short channel MOSFET models:

$$I_G = 0$$

$$I_D = \begin{cases} 0 & V_{GS} < V_T \text{ (cutoff)} \\ \frac{K'W_{eff}}{L_{eff}} \left(V_{GS} - V_T - \frac{V_{DS}}{2} \right) V_{DS} \left[1 + \lambda \left(\frac{1 + \theta L_{eff}}{\theta L_{eff}} \right) V_{DS} \right] & V_{GS} > V_T, 0 < V_{DS} < V_{GS} - V_T \text{ (ohmic)} \\ \frac{K'W_{eff}}{2L_{eff}} (V_{GS} - V_T)^2 \left[1 + \lambda \left(\frac{1 + \theta L_{eff}}{\theta L_{eff}} \right) V_{DS} \right] & V_{GS} > V_T, V_{DS} > V_{GS} - V_T \text{ (saturation)} \end{cases}$$

Parameter Θ has units length⁻¹. Typical value is $0.2 \text{ } 1/\mu$. Parameter W_{eff} and L_{eff} denote the effective channel width and length.

Short channel devices are widely used in digital-logic, where improvement in speed and reduction in circuit area are achievable inspite of performance limitations of short channel MOSFETS.

2.8 Subthreshold operation of a MOSFET:

In DC MOSFET model, $I_D = 0$ when $V_{GS} < V_T$ whereas I_D is non zero when $V_{GS} > V_T$. But in physical devices, experimentally, a small drain current flows when $V_{GS} < V_T$. If $V_{GS} > V_T$, the devices are said to be operated in strong inversion whereas if $V_{GS} < V_T$, the device is said to be operated in weak inversion, or equivalently in subthreshold region. At room temperature, the transition between strong and weak inversion occurs around $V_{GS} = V_T + 100\text{mV}$. The expression $V_{GS} = V_T + 2nV_t$, where n is constant between 1 and 2, the term V_t is equivalent to kT/q where, k is boltzmann's constant, T is absolute temperature in degree Kelvin and q is the charge of an electron. At room temperature (300°K), $V_T = 26\text{mV}$.

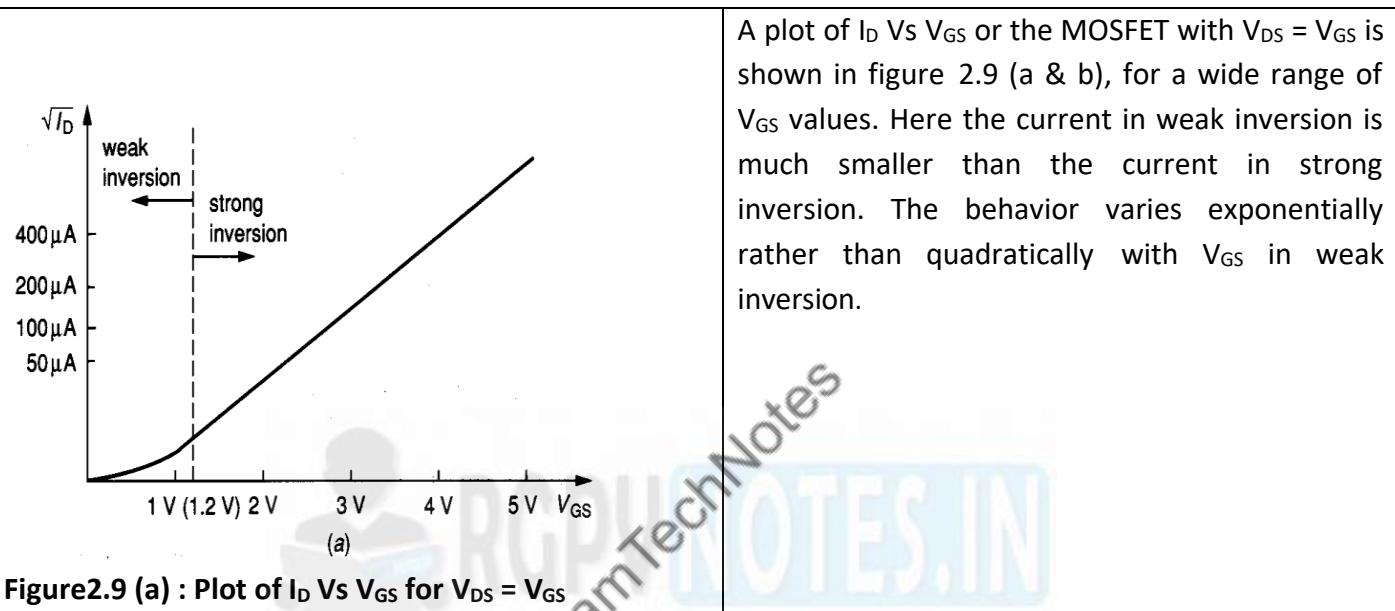


Figure 2.9 (a) : Plot of I_D Vs V_{GS} for $V_{DS} = V_{GS}$

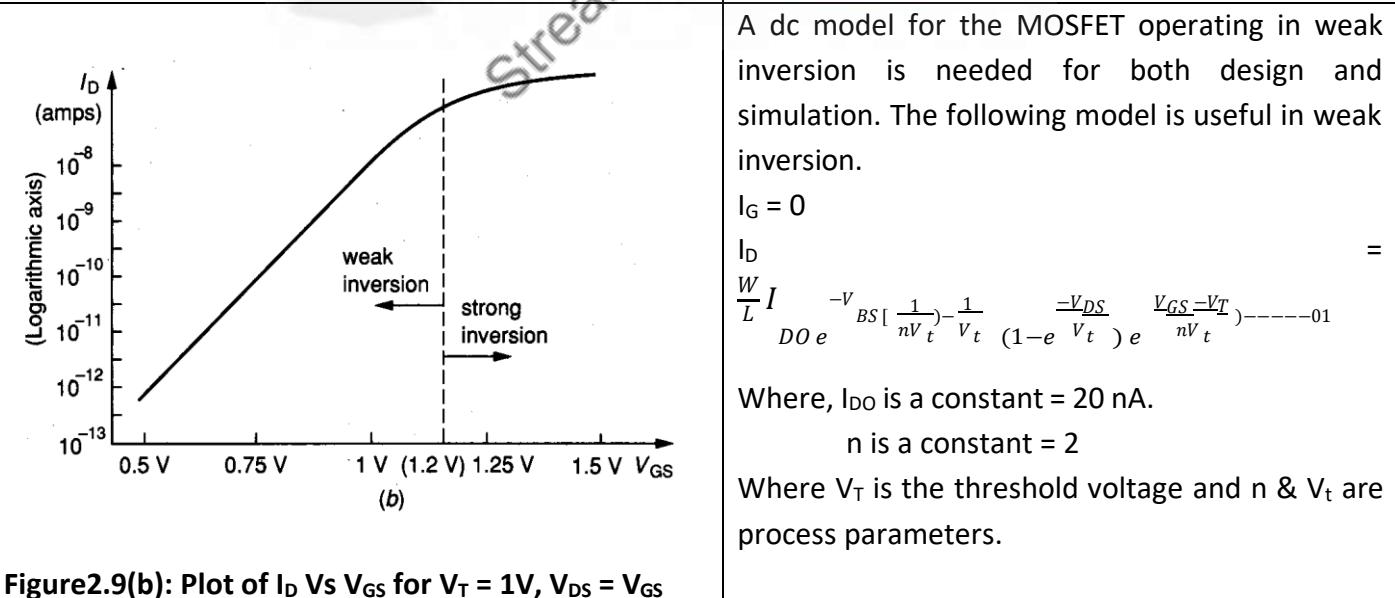
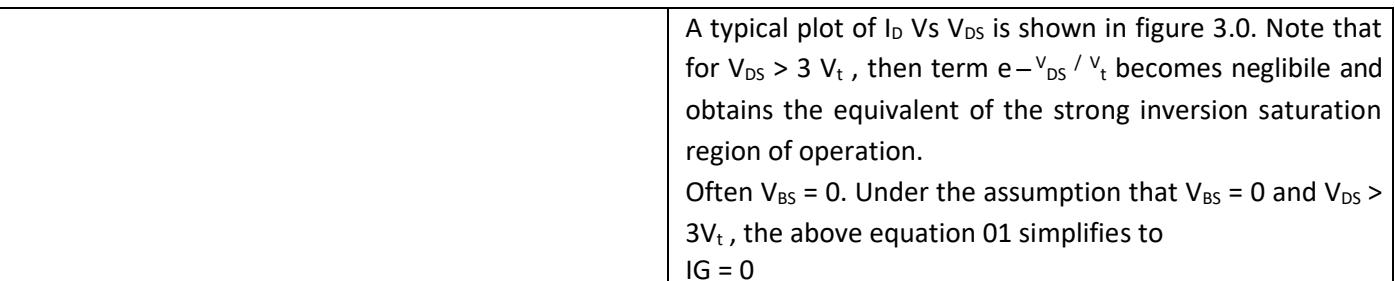


Figure 2.9(b): Plot of I_D Vs V_{GS} for $V_T = 1\text{V}$, $V_{DS} = V_{GS}$



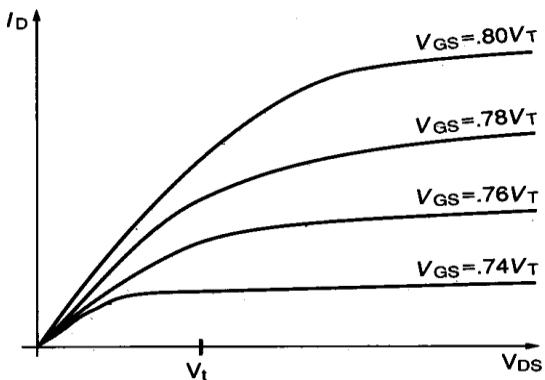


Figure 3.0: Typical I_D Vs V_{DS} characteristics for MOSFET operating in sub threshold

$$I_D = \frac{W}{L} I_{DO} e^{\frac{V_{GS}-V_T}{nV_t}}$$

The parameter I_{DO} is related to the transconductance parameter and is given by expression

$$I_{DO} = [K' 2(nV_t)^2] / e^2$$

Practical Limitations of device operated in weak inversion region:

- 1) Poor frequency response: The maximum current required to charge and discharge the geometrically determined parasitic capacitances, is very less causing a deterioration of frequency response.
- 2) The drain and source substrate currents associated with reverse-biased moat-substrate junctions are not negligible as compared to subthreshold drain currents.
- 3) Linearity is quite poor for $V_{DS} < 3V_T$, making linear designs quite more challenging.
- 4) The deterioration of matching characteristics of MOS transistors with decreasing drain currents further complicates the linear design.

Because of these practical limitations, it is often desirable to operate near the transition region between strong and weak inversion region, where benefits of reduced power dissipation is derived in weak inversion region and other limitations are not too problematic.

Diffusion current dominates weak inversion operation and drift current dominates strong inversion operation, and both current interacts in the transition region and thus complicate the modeling problem.

Weak inversion models used in SPICE simulations have major limitations.

Applications, do exist, where requirement of current is extremely small.

These includes, biomedical application such as pacemakers and other implantable devices that must operate for several years with small non-rechargeable batteries.

2.9 Noise Sources in MOSFET's:

Two mechanisms are primary contributors of presence of noise in MOSFETs.

- 1) Thermal noise (white noise)
- 2) Flicker noise

Thermal noise associated with the carriers in the channel.

Flicker noise associated with the trapping and releasing of electrons in Si-SiO₂ interface region.

Both these noise sources contribute to the total drain current and can be modeled as a current source between drain and source in either large or small signal device model.

Thermal Noise:

Thermal noise current is characterized by its spectral density:

$$S_{IW} = \begin{cases} \frac{4kT}{R_{FET}} & \text{ohmic region} \\ \frac{8kTg_m}{3} & \text{saturation} \end{cases}$$

Where T is temperature in degree kelvin. k is Boltzmann's constant. R_{FET} is equivalent FET resistance

g_m is small signal transconductance at operating point. Coefficient 8kT/3 = 1.1 × 10⁻²⁰ V.A.sec

In saturation region, thermal noise current is equal to that in a resistor (channel source resistance) of value 3/2g_m.

In ohmic regions, thermal noise current is identical to that of a resistor of value R_{FET}.

Flicker or 1/f Noise :

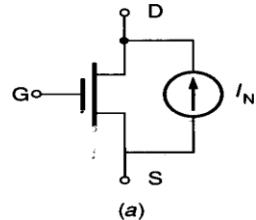
Flicker noise current in both the ohmic and saturation regions is characterized by its spectral density:

$$S_{If} = \frac{2K_f K' I_{DQ}}{C_{ox} L^2 f}$$

Where K_f flicker noise coefficient = $3 \times 10^{-24} \text{ V}^2 \cdot \text{F}$. I_{DQ} is quiescent current. f is the frequency.

K' , COX and L are MOSFET model parameters

Hence the spectral current density of noise current source , IN in fig. is given as: $S_N = S_{IW} + S_{If}$



It is known that RMS noise current source in the frequency band (f_1, f_2) can be obtained from the spectral density and is given by:

$$I_{NB} = \sqrt{\int_{f_1}^{f_2} S_N df}$$

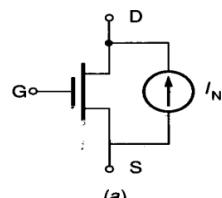
If we define the RMS white noise and flicker noise currents by the square root of the integrals of the spectral densities

$$I_{WB} = \sqrt{\int_{f_1}^{f_2} S_{IW} df} \quad I_{FB} = \sqrt{\int_{f_1}^{f_2} S_{If} df}$$

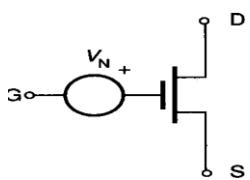
Then these currents adds in the RMS by sense

$$I_{NB} = \sqrt{I_{WB}^2 + I_{FB}^2}$$

The current I_{NB} is the contribution to the I_N in frequency band (f_1, f_2). If fig. a



(a) is replaced by its small signal model, then an equivalent small signal model that characterizes the noise performance can be obtained by replacing the noise current source from drain to source with a noise voltage source in series with the gate, V_N , shown in fig.b.



(b) This V_N , is termed as input-referred noise source. The spectral density of input referred noise source Is given by expression $S_{VN} = S_{IN} / g_m^2$

Correspondingly, the input referred noise source relates to the noise current source is given by expression

$$V_N = I_N / g_m$$

Where g_m is the small signal transconductance gain of MOSFET.

Noise model in fig. a, is valid for small signal or large signal model of the MOSFET whereas fig.b model is valid for only small signal model because this model was developed under the assumption of small signal operation of MOSFET. Device sizes and bias currents affect both the noise currents. At higher frequencies, thermal noise effects dominates whereas the flicker noise, because of its 1/f type spectral density, dominates at lower frequencies.

2.10 Diode Models:

DC Diode Model:

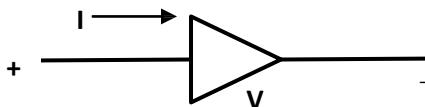


Fig. 2.7: Variable convention for a diode

The diode is characterized by the low frequencies by the equation
 $I = I_s (e^{V/nV_t} - 1)$ where parameters I_s , n , and V_t characterize the device. Where I_s is saturation current, n is emission coefficient.

and parameter $Vt = kT / q$ where k is boltzmann's constant, T is the absolute temperature in degree Kelvin and q is the charge of an electron. At room temperature (300°K), Vt is 26 mV. The emission coefficient takes on a value of about 1 for silicon. The saturation current can be expressed as $I_s = J_s A$, where J_s is a process parameter equal to the saturation current density and A is the cross sectional area of the junction.

The diode equation simplifies for $V \gg nV_t$. Under forward biased greater than 0.25 V, no accuracy is lost in modeling the diode by $I = I_s \cdot e^{V/nV_t}$. Likewise, for $V < -0.25$ V, the diode is accurately modeled by $I = -I_s$.

The piecewise-linear model shown in fig.2.8, is often used for biasing and/or large signal applications. The piecewise-linear model is mathematically given by the equations

$V = I r_\gamma + V_\gamma$ for $I > 0$ and $I = 0$ for $V < V_\gamma$. Where r_γ is an ON resistance of the diode and V_γ is diode cut-in voltage.



Fig. 2.8: Piecewise-linear diode model

2.11 BJT MODEL (DC BJT MODEL):

The starting point will be the dc model introduced by Ebers and Moll in 1954.

In this model the relation between the terminal variable is given by

$$I_C = I_S (e^{(V_{BE}/V_t)} - 1) - I_S / \alpha_R ((e^{(V_{BC}/V_t)} - 1))$$

$$I_E = -I_S / \alpha_f ((e^{(V_{BE}/V_t)} - 1) + I_S (e^{(V_{BC}/V_t)} - 1))$$

Where I_S , α_f , α_R and V_t four parameter characterize by device.

I_S = transport saturation

α_f = large signal forward current gain of common base configuration

α_R = large signal Reverse current gain of common base configuration

$V_t = kT/q$ where, k is boltzmann's constant, T is absolute temperature in degree Kelvin and q is the charge of an electron. At room temperature (300°K), $V_t = 26$ mV.

The two Ebers - Moll equation with KVL and KCL ie. $I_B = -I_C - I_E$; $V_{CE} = V_{BE} - V_{BC}$ applied to transistor itself provide four independent equations relating the six terminal variables, I_C , I_B , I_E , V_{CE} , V_{BC} , and V_{BE} . The BJT is characterized by these four equations. It should be noted that there is complete functional symmetry in the device model with respect to the collector and emitter terminals. This is comforting since the NPN transistor can be thought of as a sandwich of a p-type region between two n-channel layers.

The parameters α_f and α_R are determined by impurity concentrations and junction depths, and as such are process parameters. The parameter Vt is a function of two physical constants and temperature and thus cannot be considered a design parameter. The parameter Vt should not be confused with the threshold voltage for a MOSFET, which uses the symbol V_T . Because of the temperature dependence, V_t can be considered an environmental parameter. The transport saturation current can be expressed as

$$I_S = J_S A \quad \text{where } A \text{ is the area of the emitter and } J_S \text{, the transport saturation current density.}$$

An equivalent circuit for the bipolar transistor based upon the Ebers-Moll model is sometimes useful. The circuit of figure below serves as such an equivalent circuit where the currents in the two diodes satisfy the standard diode equations

$$I_F = I_S / \alpha_F (e^{(V_{BE}/V_T)} - 1); I_R = I_S / \alpha_F (e^{(V_{BC}/V_T)} - 1); J_S = 6 \times 10^{-10} \mu\text{A}/\text{mil}^2$$

$$\alpha_F = 0.99; \alpha_R = 0.3; V_{AF} = 200\text{V}; V_{AR} = 200\text{V}$$

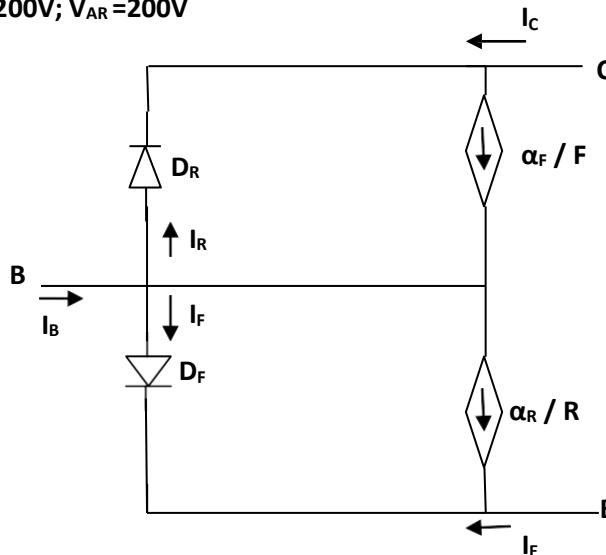


Figure: Equivalent circuit of BJT based upon Ebers-Moll Model

2.12 PASSIVE COMPONENT MODELS:

Discrete passive components are quite easy to model. Resistors and capacitors can generally be modeled by ideal resistors and capacitors respectively. The major limitations are manufacturing tolerances and temperature deviations - both of which can be reduced to acceptable levels in most applications through judicious component selection/specification.

Monolithic resistors and capacitors are far from ideal. They are typically both temperature and voltage dependent. The practical range of values is seriously limited by area constraints. Large resistor or capacitor values are impractical. Process deviations preclude accurate control of absolute component values. Parasitic effects are often quite significant. Relative accuracy between passive components is, however, often quite good.

The following figure of merits used to characterize passive components.

Resistors

1. Sheet resistance
2. Resistance density
3. Temperature coefficient of resistance
4. Voltage coefficient of resistance
5. Absolute accuracy
6. Relative (ratio) accuracy

Capacitors

1. Capacitance density
2. Temperature coefficient of capacitance
3. Voltage coefficient of capacitance
4. Absolute accuracy
5. Relative (ratio) accuracy

The resistance density is generally a function of the process parameters and layout design rules.

The capacitance density is a process parameter.

The absolute accuracy is a measure of how accurately the actual resistor and capacitor values can be controlled during processing.

The relative (ratio) accuracy is a measure of how closely two resistors or capacitors can be matched.

The ratio accuracy is affected by component placement on a die, device geometry, the physical size of the components, and the nominal relative values of the components themselves.

Monolithic Capacitors:

Any structure in which a voltage induced separation of charge occurs can serve as a capacitor.

In a MOS process, the most common capacitors are formed by sandwiching a thin oxide layer between two conductive polysilicon layers. These capacitors are nearly independent of applied voltage and can be modeled as ideal capacitors. The major limitation is the large parasitic capacitor that is always formed between the lower plate and the substrate.

When the luxury of the double polysilicon layers is not present in MOS processes, metal-poly, metal-diffusion (lower plate formed by a diffused region in the substrate), or poly-diffusion capacitors with an SiO_2 dielectric are

used. These capacitors typically have a lower capacitance density and/or increased voltage dependence and/or a less conductive lower plate than the double poly capacitors.

In the bipolar process, the most desirable common capacitors are metal diffusion capacitors with an SiO_2 dielectric. The heavily doped emitter diffusion is used for the lower diffusion plate. The characteristics are quite good, but an additional mask step is required for forming the dielectric region.

Alternatives include the voltage-dependent junction capacitances formed by either the B-C or B-E junction. The B-E junction offers reasonable capacitance density at the expense of a limited reverse breakdown voltage (typically 5 to 7 V). The B-C junction reverse breakdown voltage is quite high (typically in the 30 V range), but the capacitance density is quite low. All junction capacitors are limited by requirements that the junctions remain reverse biased (actually, not forward biased by more than a few tenths of a volt).

Monolithic Resistors:

Monolithic resistors are passive devices and others contain active devices.

Major tradeoffs must be made between linearity, area, biasing complexity, and temperature characteristics in monolithic resistors.

In standard MOS processes, the most ideal resistors are merely strips of polysilicon. Diffusion strips are also used for resistors but exhibit an undesirable nonlinear relationship between voltage and current. Ion implants offer some advantages over depositions for the introduction of impurities to control absolute resistance values in diffused resistors. Thin film resistors with excellent characteristics are added in some specialized processes. For each of these types of resistors, a serpentine pattern is often used to improve packing density. The major limitations of these resistors are the low resistance densities, which limit the total resistance to quite small values; the high deviations in resistance due to process variations; and large temperature coefficients.

In bipolar processes epitaxial strips or diffusion strips are commonly used for resistors. These devices are quite linear. The base diffusion is often used because of its reasonably high sheet resistance.

To prevent forward biasing of the "base-collector" junction, a contact is needed to the epitaxial layer. This will be typically connected to the most positive power supply voltage used for the circuit.

It can be argued that the resistance of the base-diffused resistor could be increased if the depth of the p-base diffusion could be decreased. The depth of this diffusion, however, is generally determined to optimize performance of the BJTs themselves.

An alternative is to place an n⁺ emitter diffusion in the p base region. This masking step already exists and will result in a significant increase of the sheet resistance of the underlying p diffusion. Such a device, which is termed a pinch resistor

Contact must be made to both the n⁺ emitter diffusion and the n- epitaxial region. These regions are typically both connected to the most positive power supply voltage used for the circuit. Although the resistance increases significantly due to this pinching, the variance in emitter and base diffusion depths due to process variations makes the tolerances of pinch resistors quite wide. They also exhibit an increased voltage dependent nonlinearity and are limited in voltage range to circumvent break-down of the reverse-biased base-emitter junction. Other types of pinch resistors (e.g., epitaxial pinch) can also be made.

Several active resistors often offer considerable reductions in area requirements compared to passive resistors at the expense of increased nonlinearity and/or reduced signal swing and/or complicated biasing requirements.



Subject Notes
UNIT-III

Circuit Simulation :

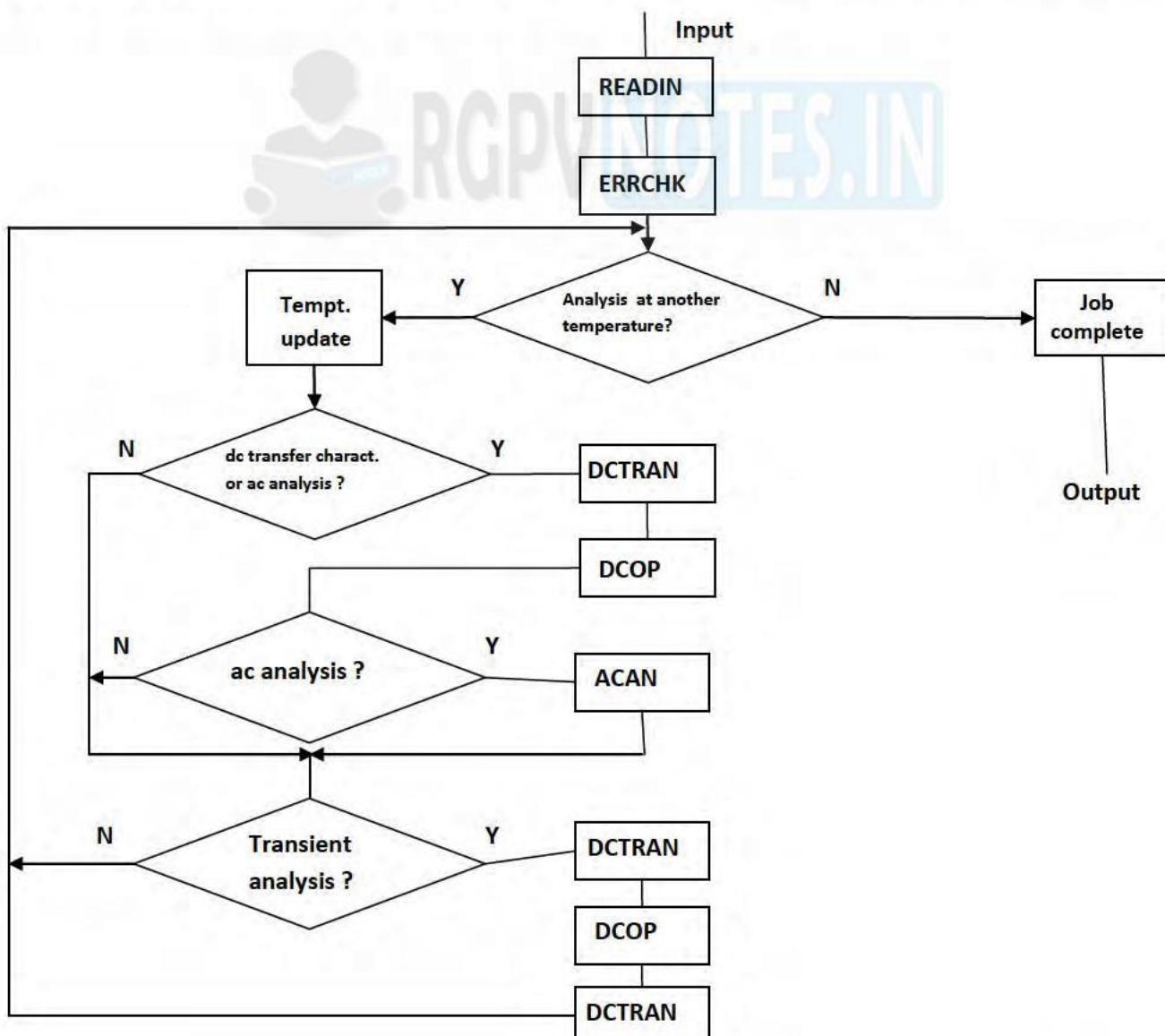
Introduction, Circuit Simulation Using Spice, MOSFET Model, Level 1 Large signal model, Level 2 Large Signal Model, High Frequency Model, Noise Model of MOSFET, Large signal Diode Current, High Frequency BJT Model, BJT Noise Model, and Temperature Dependence of BJT.

Introduction:

Computer simulation of a circuit entails using a computer to predict or simulate the performance of circuit or systems. The circuit which is simulated may include from few component to several component. Many different types of computer programs are used for simulations of IC, depending on the type of analysis required and the size of circuit involved.

Circuit Simulation Using Spice:

SPICE utilizes a modified nodal analysis approach. It can be used for non-linear-dc, non-linear transient, and linear-ac analysis problems. It also includes modules for more specialized analysis such as noise analysis, temperature analysis, etc. The inputs can be constant or time varying. For non linear dc and transient analysis, the program includes the nonlinear effects of all devices specified by the user. For the ac analysis, the dc operating point is internally determined. Block diagram showing the fundamental operation of SPICE.



As shown in the block diagram, the subroutine READIN reads the SPICE input file. ERRCHK verifies that the input syntax is correct. Temperature effects are handled in SPICE by repeating the entire analysis for each temperature. The subroutines DCTRAN, DCOP, and CAN perform the bulk of the manipulations. DCTRAN, DCOP are used for DC transfer characteristics, DC operating point calculations, and transient analysis, CAN perform the small signal AC analysis. SPICE is the inclusion of respectable models for the basic devices, specially, the diode, BJT, JFET and MOSFET. The user can create generic models for the active devices that correspond to, and are consistent with, the process parameters and design rules of the process used for specific design. SPICE version has different MOSFET models designated as levels. Device models used by SPICE (Simulation Program for Integrated Circuit Engineering) simulators can be divided into three classes:

1. First Generation Models (Level 1, Level 2, Level 3 models)
2. Second Generation Models (BISM, BSIM2, HSPICE Level 28)
3. Third Generation Models (Level 7, Level 48, BSIM3, etc.)

To run SPICE, the user must first create a file, referred to as spice deck, which contains the complete description of the circuit (including device models) along with a description of the excitation and the type of analysis desired. This file is accessed when the SPICE program runs.

Brief introduction to SPICE:

- Simulation Program with Integrated Circuit Emphasis (SPICE)
- Developed at UC Berkeley
- Industry-standard general-purpose circuit simulator
- Numerical simulation
- Modified nodal analysis
- External nodes define boundary conditions
- Circuit elements represented by device models
- Text input, text or graphical output

Circuit Elements in SPICE

- Resistors
- Capacitors
- Inductors and coupled inductors
- Independent sources (V, I)
- Dependent sources
- Transmission lines
- Switches
- Uniform distributed RC lines
- Active devices
- Diodes, BJTs, JFETs, MOSFETs, MESFETs

SPICE File

- A SPICE file is made of a series of statements.
- First line is a title statement.
- Last line is an end statement.
- The order of other statement is arbitrary.
- Element statements describe the circuit.
- Control statements describe model parameters and execute analyses.

MOSFET Model, Level 1 Large signal model, Level 2 Large Signal Model :

MOSFET Model:

SPICE provides a wide variety of MOS transistor models with various trade-offs between complexity and accuracy. Level 1 and Level 3 models were historically important, but they are no longer adequate to accurately model very small modern transistors. BSIM models are more accurate and are presently the most widely used. Some companies use their own proprietary models.

SPICE Version 2G has three different MOSFET models, designated as Level 1, Level 2, and Level 3. Version 3 of SPICE also contains a fourth MOSFET model, the BSIM model. The BSIM model is a process oriented model which places a major emphasis on short channel devices. The discussion here will be restricted to the Level 1, Level 2, and Level 3 models.

The Level 1 model is termed the Shichman-Hodges model. It closely follows the work of Sah 13. The Level 1 model is the simplest model and is useful for verifying that no errors occurred in the hand calculations. In some applications the Level 1 model may be adequate for computer simulations.

The Level 2 model differs from the Level 1 model both in its method of calculating the effective channel length and the transition between the saturation and ohmic regions. A time consuming polynomial routine is required for the Level 2 model to determine the transition point between the linear and saturation regions. The Level 2 model offers improvements in performance which are particularly significant for short channel devices.

The Level 3 model is termed a semi-empirical model. Several empirical parameters are introduced in the Level 3 model. These parameters may offer improvements in fit of the model. The Level 3 model also offers a reduction in time required to calculate the transition point between the linear and saturation regions of operation.

The Level 1, Level 2, and Level 3 device models can be found respectively in subroutines MOSEQ1, MOSEQ2, and MOSEQ3 of the SPICE source code. Subroutine MODCHK is used for some of the hierarchical parameter definitions. The terminal voltages of the MOSFET are passed to these subroutines, and the nodal currents are returned along with the small signal model parameters at the operating point. In each interval of time in a transient analysis, nodal currents are comprised of two parts. The first is the dc current obtainable from the large signal device model.

The second is the charge current associated with the parasitic capacitances in the devices. This latter current plays a major role in high-frequency transient analyses.

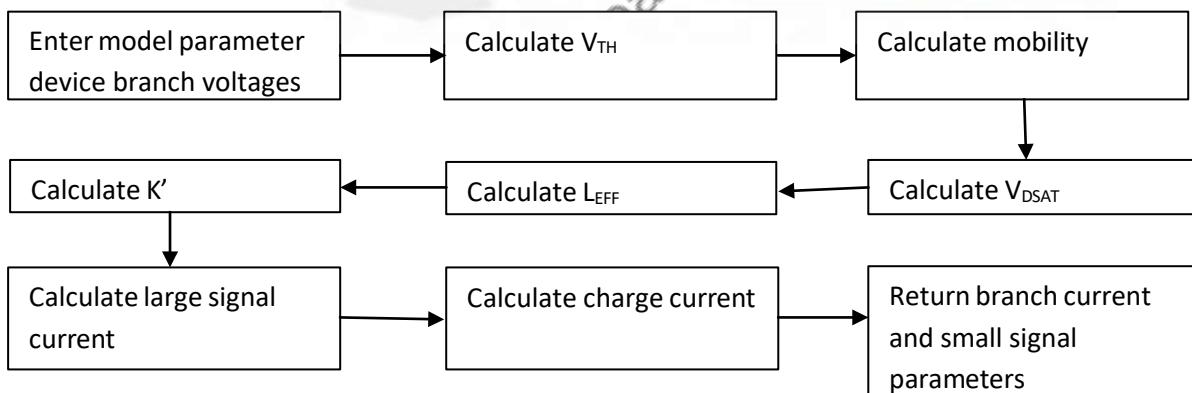


Fig. 3.1 : Simplified flow chart for spice subroutines MOSEQ 1, MOSEQ2, MOSEQ3

A simplified flow chart for spice subroutines MOSEQ 1, MOSEQ2, MOSEQ3 is shown in figure above . The large signal currents in quadrant-1 of $I_d - V_{ds}$ are calculated from the expression:

$$I_G = I_B = 0$$

$$I_D = I_{\text{CUTOFF}} \quad V_{GS} < V_{TH}$$

$$I_D = I_{\text{OHMIC}} \quad V_{GS} > V_{TH}; \quad V_{DS} < V_{DSAT}$$

$$I_D = I_{\text{SAT}} \quad V_{GS} > V_{TH}; \quad V_{DS} > V_{DSAT}$$

Level 1 large signal model:

The regions of operation are given by:

$$I_G = I_B = 0$$

$$I_{\text{CUTOFF}} \quad V_{GS} < V_{TH}$$



$$I_D = \begin{cases} I_{\text{OHMIC}} & V_{GS} > V_{TH} \quad V_{DS} < V_{DSAT} \quad (\text{Where } V_{DSAT} = V_{GS} - V_{TH}) \\ I_{\text{SAT}} & V_{GS} > V_{DS} > V_{DSAT} \end{cases}$$

And different drain currents are :

$$I_{\text{CUTOFF}} = 0$$

$$I_{\text{OHMIC}} = K'(W/L_{\text{EFF}})([V_{GS} - V_{TH}] - V_{DS}/2)V_{DS}$$

$$I_{\text{SAT}} = K'/2 (W/L_{\text{EFF}})[V_{GS} - V_{TH}]^2$$

where V_{DSAT} is given by $V_{DSAT} = V_{GS} - V_{TH}$

The parameters V_{TH} and L_{EFF} represent the threshold voltage and effective length of the device and are given by:

$$V_{TH} = V_{TO} + \gamma * v(\phi - V_{BS}) - v(\phi)$$

$$L_{\text{EFF}} = L_{\text{ADJ}} / (1 + \lambda V_{DS})$$

V_{TO} is zero bias threshold voltage; λ is channel length modulation and SPICE input parameter;

γ is the bulk threshold parameter, ϕ is the surface potential;

The parameter K' is given as:

$$K' = \mu_0 C_{ox}$$

Where μ_0 is nominal channel mobility; C_{ox} gate oxide capacitance and T_{ox} oxide thickness

$$C_{ox} = \epsilon_{ox} / T_{ox} \text{ where, } \epsilon_{ox} \text{ dielectric constant of SiO}_2$$

L_{adj} represent the adjusted length which is drawn length reduced the lateral diffusion on the drain and source, LD:

$$L_{\text{adj}} = L - 2 LD$$

The parameter in V_{TH} if not input, are calculated from

$$\gamma = (\sqrt{2q\epsilon_{Si}N_{SUB}}) / C_{ox}; \quad \phi = (2kT/q) \ln (N_{SUB} / n_i)$$

$$V_{TO} = V_{FB} + \gamma \sqrt{\phi} + \phi$$

where, ϵ_{Si} is the dielectric constant of silicon, N_{SUB} is the substrate doping, q is the charge of an electron, n_i is the intrinsic carrier concentration of silicon. N_{SUB} is a spice input parameter and ϵ_{Si} & n_i are physical constants.

The flatband voltage, V_{FB} is given by

$$V_{FB} = \phi_{ms} - (qN_{ss} / C_{ox})$$

Where V_{FB} is flat band voltage; input parameter, N_{ss} , is the effective surface density and ϕ_{ms} is the semiconductor work function difference and is calculated as: $\phi_{ms} = W_{FN} - (\phi/2)$ where the parameter W_{FN} is an internal function of physical constants characteristics of the materials involved, TGP which specifies the type of materials used to construct the device and temperature.

Level 2 Large Signal Model:

$$\begin{aligned} I_{\text{CUTOFF}} &= 0 & NFS=0 \\ &= I_{\text{weak inversion}} & NFS \neq 0 \end{aligned}$$

$$I_{\text{OHMIC}} = \frac{K'W}{L_{eff}} [V_{GS}V_{TH} - \eta \frac{V_{DS}}{2}] + I_{BS0}$$

$$I_{\text{SAT}} = \frac{K'W}{L_{eff}} [V_{GS} - V_{TH}]^2 (2 - \eta) + I_{BS0}$$

Where the cutoff transition region is determined by:

$$V_{TH} = V_{TO} + \gamma [\sqrt{\phi - V_{BS}} - \sqrt{\phi} - \gamma \alpha \sqrt{\phi - V_{BS}} + (\phi - V_{BS}) \frac{\pi}{4} \frac{\epsilon_{Si}}{C_{ox}L_{adj}}] \quad (\text{DELTA})$$

Parameter V_{MAX} is used to characterize the saturation/ohmic transition, it denotes the maximum drift velocity of the carriers in the channel. If parameter V_{MAX} is not input, the saturation or ohmic transition region is determined by:

$$V_{DSAT} = \frac{V_{GS} - V_{TH}}{\eta} + \frac{1}{2} \left[\frac{\gamma}{\eta} \right]^2 \cdot \left\{ 1 - [1 + 4 \left(\frac{\gamma}{\eta} (1 - \alpha) \right)]^{-2} \left(\frac{V_{GS} - V_{TH}}{\eta} + \phi - V_{BS} \right)^{1/2} \right\}$$

The parameters η , V_{TH} , I_{BS0} and I_{BS0} are given as follows:

$$\eta = 1 + \frac{\pi}{4} \frac{\epsilon_{Si}}{C_{ox}L_{adj}} \quad (\text{DELTA})$$

$$V_{TH} = V_{TO} - \gamma \sqrt{\phi} + (\phi - V_{BS}) \frac{\pi}{4} \frac{\epsilon_{Si}}{C_{ox}L_{adj}} \quad (\text{DELTA})$$

$$I_{BS0} = - \frac{2 K' W \gamma s}{3 L_{eff}} \left[(\phi + V_{DS} - V_{BS})^{\frac{3}{2}} - (\phi - V_{BS})^{\frac{3}{2}} \right]$$

$$I_{BS} = \frac{K'_2 W}{L_{eff}} \left[(V_{GS} - V_{TH}) (1 - \eta) - \frac{\eta}{2} (V_{DSAT} - (V_{GS} - V_{TH})) \right] (V_{DSAT} - (V_{GS} - V_{TH})) - \frac{2}{3} \frac{\gamma_s}{\eta} [(V_{DSAT} - V_{BS} + \varphi)^{\frac{3}{2}} - (f - V_{BS})^{\frac{3}{2}}]$$

Parameter K'_2 is obtained by

$$K'_2 = K' \frac{\mu_s}{\mu_0} \text{ where } \mu_s \text{ is effective surface mobility.}$$

The parameter DELTA is a SPICE input parameter and is used to characterize width reduction where effective width, W_{eff} is given by $W_{eff} = W (2 - \eta)$

Parameter V_{TO} , L_{adj} and φ are determined from Level 1 model

Parameter γ_s is given by $\gamma_s = \gamma(1 - \alpha)$

Parameter α is given by

$$\alpha = \frac{X_J}{L_{adj}} \sqrt{1 + \frac{2W_s}{X_J} + \sqrt{1 + \frac{2W_D}{X_J}}} - 2$$

$$\text{where } W_s = X_D \sqrt{\varphi - V_{BS}}$$

$$W_D = X_D \sqrt{\varphi - V_{BS} + V_{DS}}$$

Where X_J is SPICE input parameter representing the metallurgical junction depth.

Parameter $L_{eff} = L_{adj}(1 - \lambda V_{DS})$.

If not specified as an input parameter, λ is calculated from the expression

$$\lambda = \frac{X_D}{L_{adj} V_{DS}} \sqrt{\frac{X_D V_{max}}{2\mu_s}^2 + (V_{DS} - V_{DSAT})} - \frac{X_D V_{max}}{2\mu_s} \text{ and parameter } X_D = \frac{\sqrt{2\varepsilon_{si}}}{qN_{SUB}}$$

$$\begin{aligned} \text{Effective surface mobility is given by } \mu_s &= \mu_0 \frac{\frac{UCRIT \cdot \varepsilon_{si}}{c_{ox} (V_{GS} - V_{TH})}}{UEXP} \text{ For Level 2} \\ &= \mu_0 \frac{\frac{UCRIT \cdot \varepsilon_{si}}{c_{ox} (V_{GS} - V_{TH} - UTRA \cdot V_{DS})}}{UEXP} \text{ For Level 3} \end{aligned}$$

Parameters UCRIT, UTRA, UEXP are SPICE input parameters used characterized mobility degradation.

High Frequency MOSFET Model:

The high frequency MOSFET model is obtained from the dc model by adding the identifiable parasitic capacitances to dc model. Parasitic device capacitors in SPICE are essentially modeled. Those capacitors that are voltage independent are modeled by:

$C = \varepsilon A/t$; Where ε is dielectric constant and A is area of intersection of two plate and t is dielectric thickness.

The voltage independent capacitances are comprised of those that are operation region dependent and are comprised of those overlap capacitors that appear on the periphery of MOSFET.

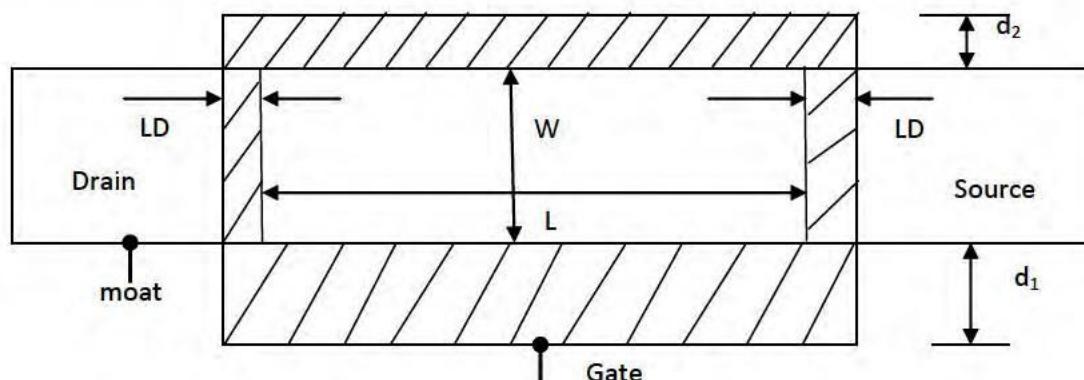


Fig. 3.2 : Top view of operation region independent or overlap (shaded) capacitor ,in MOS transistors



The gate source overlap capacitance $C_{GSOL} = \epsilon L D.W / t$ OR $C_{GSOL} = C_{GSO} W$

where, $C_{GSO} = \epsilon L D/t$ represents the overlap capacitance per unit width of the gate. $W.LD$ is effective area of the overlap component of the gate –source capacitance where W is the channel width and LD is lateral diffusion of source. C_{GSO} serves as an input parameter in SPICE to model overlap capacitance.

If C_{GSO} is not entered in SPICE, then $C_{GSOL} = Cox. (L D).W$

The gate-drain overlap capacitance is modeled in the same manner.

From above figure 3.2, gate-bulk overlap capacitance is linearly proportional to L and dependent upon d_1 and d_2 . The dielectric thickness of the gate bulk overlap is essentially equal to the gate oxide thickness on the sides adjacent to the channel and equal to the height of the field oxide on the sides from the channel. It is voltage independent and proportional to L , it is modeled by

$C_{GBOL} = C_{GBO} . L$ where, C_{GBO} represent the bulk capacitance per unit length of the device. C_{GBO} serves as an input parameter in SPICE to characterize the gate-bulk overlap.

Those capacitor that are voltage dependent are the pn junction capacitors. They occur between bulk and source, bulk and drain, and bulk and channel. They are modeled as:

$$C = C_j A / (1 - V_F / \phi_B)^n \text{ for } V_F < FC. \phi_B$$

C_j is zero bias junction capacitance density; FC is called the forward biased capacitance coefficient ($FC \approx 0.5$), ϕ_B is the barrier potential, n is constant that characterize the junction type, A is the junction cross sectional area and V_F is the forward biased on the junction (usually negative for MOS devices).

A cross section of diffused region showing parasitic junction capacitors is shown in figure 3.3.

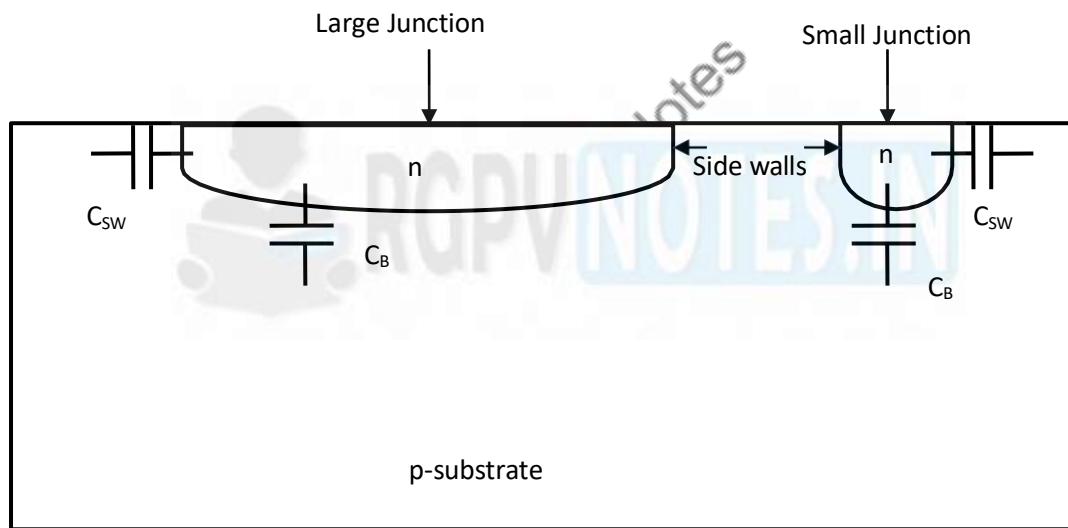


Fig . 3.3 : Cross section of diffused region showing parasitic junction capacitors

The junction is not planar. The impurity concentrations in the p and n type materials at the bottom of the junctions are different than the concentrations along the sidewalls causing the grading coefficient, n , to vary as a function of position. Although the side wall capacitances become negligible for large structures, they may actually dominate for small or narrow junctions. The total junction capacitance associated with the "reverse biased" (actually for $V_F < FC. \phi_B$) junction of either the source or drain is thus

$$C_{RB\ TOTAL} = C_j A / [1 - (V_F / \Phi_B)^{MJ}] + C_{JSW} P / [1 - (V_F / \Phi_B)^{MJSW}]$$

where C_j is the zero-bias bottom capacitance area density, C_{JSW} is the zero-bias sidewall capacitance per unit length of the perimeter, MJ is the bottom junction grading coefficient. FC is the forward bias coefficient and $MJSW$ is the sidewall grading coefficient. A is the junction bottom area and P is the junction sidewall perimeter. The parameters FC , C_j , C_{JSW} , MJ , $MJSW$ and Φ_B can be entered in SPICE on the .MODEL card (line). The parameters A and P for the drain and source junctions must be entered on the device card (line). C_j is calculated Under the assumption of a step graded junction, from the expression $C_j = \sqrt{\epsilon_s q N_{SUB} / 2\Phi_B}$

Under forward biased ($V_F > FC \cdot \phi_B$), the total junction capacitance are modeled by

$$C_{FB\ TOTAL} = [(CJ.A) / (1-FC)^{(1+MJ)}] [(1-FC)(1+MJ) + (V_{BS} / \phi_B).MJ] + [(CJSW.P) / (1-FC)^{(1+MJSW)}] [(1-FC)(1+MJSW) + (V_{BS} / \phi_B).MJSW]$$

Noise Model of the MOSFET :

There are four noise current generator in MOSFET device model in SPICE . Two of these represent thermal noise associated with the parasitic series resistances in the drain and source. These are modeled by spectral densities of $S_{IRD} = 4kT/RD$ and $S_{IRS} = 4kT/RS$.

Where k is Boltzmann constant , T is temperature in Kelvin and RD and RS are drain and source parasitic resistance .

The other two noise current generators are modeled as current sources from drain to source. One represent white shot noise and other flicker (1/f) noise. These are characterized in the saturation region by spectral densities of $S_w = 8kTg_m/3$ and $S_f = (K_f)I_{dq,AF}/(f \cdot C_{ox} \cdot W \cdot L_{eff})$

Where K_f and A_F are user enterable parameters, g_m is small signal trans conductance gain at Q – point, I_{dq} is the quiescent drain current, L_{eff} is the effective channel length and f is the frequency in Hz. All noise sources are assumed to be uncorrelated. S_w and S_f add to obtain the overall noise spectral density .

Diode Model :

Individual diodes are modeled by two step process in SPICE. The device element line (card) contains information about the nodal location of the device in the circuit as well as geometric information and optional initial condition variables useful in transient analysis.

In the device element line, reference is made to a specific device modal. The model line card contains generic information about the electrical characteristics of the device formed in the process based on the characterizing process parameters. Each diode has separate device elements lines.

A philosophy distinctly different from that used for characterizing MOS processes has been adopted for characterizing the process used to fabricate diodes. The difference lies is that the characteristics of the specific reference diode are specified in model line. The size of the reference is conceptually random, although it is advice to select a reference that is geometrically similar, in size and shape of those devices that will be modeled. In the device element line (card) the relative area of the reference diode used to characterize the model is specified.

The diode is modeled as series combination of resistors, r_s and a non ideal diode, D_2 , shunted by parasitic capacitor, C_D as indicated in figure 3.4 (a). The resistor r_s accounts for both the series resistance of the diode as well as high level injection effects.

The current of diode D_2 is modeled as sum of dc (large signal) current be denoted by I_{DC} , and a current that flows through the parasitic shunting capacitance associated with the pn junction.

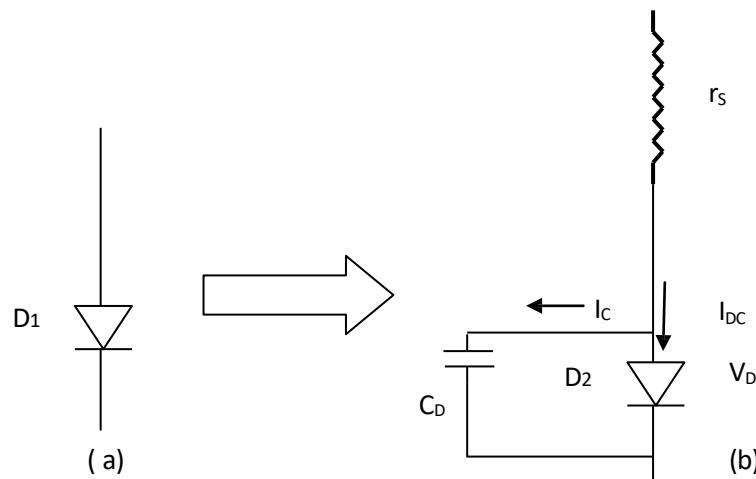


Fig.3.4 : Diode model in spice (a) Physical two terminal device (b)Equivalent circuit



Large Signal Diode Current:

The large signal diode current, I_{DC} , is modeled by that given in standard diode equation.

$$I_{DC} = I_s A_n (e^{V_D/(nV_T)} - 1) \text{ where } V_T = KT/q$$

I_s = Saturation current and n = emission coefficient

The parameter A_n represents the normalized cross section area of the junctions. It is a dimensional parameter that is entered on the device element line; it represents the ratio of the cross sectional area of the device on the device element line. If not specified, SPICE adopt a default value of $A_n=1$. The reverse breakdown voltage, B_V is not major concern, in VLSI, design since B_V generally exceeds the maximum voltage which is permitted in VLSI circuit.

BJT Model:

The SPICE model of the BJT is based upon a modified Gummel - Poon model. Simulations based upon the simpler Gummel-Poon model made by setting the appropriate Ebers-Moll parameters to zero . Use of the Ebers-Moll model in the early stages of a simulation may be useful for verifying theoretical hand calculations.

The bipolar junction transistor is modeled by two input lines (cards) in SPICE. The device line (card) is used for indicating the nodal connections of the BJT in a circuit. It is also used to reference a specific model, by name, which contains process information. An optional normalized parameter on the device line, A_n , is used to indicate the ratio of the area of the emitter to the emitter area of the device of the referenced model. An optional initial condition parameter can also be included on the device line if desired.

The second input line (card) that is used to model the BJT is the MODEL line. SPICE version 2G.6 provides for user entry of up to 40 parameters on the MODEL line for characterizing the device. As was the case with the diode, a user-selected reference transistor is used to characterize the bipolar process. The BJT model information is contained primarily in subroutines MODCHK and BJT in the SPICE source code. MODCHK does pre-processing of some of the input parameters. The model itself basically appears in subroutine BJT.

The case for the other semiconductor models in SPICE, the BJT model is characterized by four types of input parameters. These are 1) large signal or DC parameters, 2) charge storage or capacitance parameters, 3) noise parameters, 4) temperature characterization parameters.

The small signal low-frequency model used in SPICE is obtained from a symbolic differentiation of the large signal current equations evaluated at the DC operation point. As is the case for all small signal analyses in SPICE, the small signal model of the BJT is linear and thus the small signal simulation gives no distortion information. Distortion information at a specific frequency can be obtained from a much more time consuming transient response with a sinusoidal excitation of fixed frequency and amplitude. To obtain reliable distortion information, it is crucial that the transient analysis interval is long enough to guarantee essentially steady state operation.

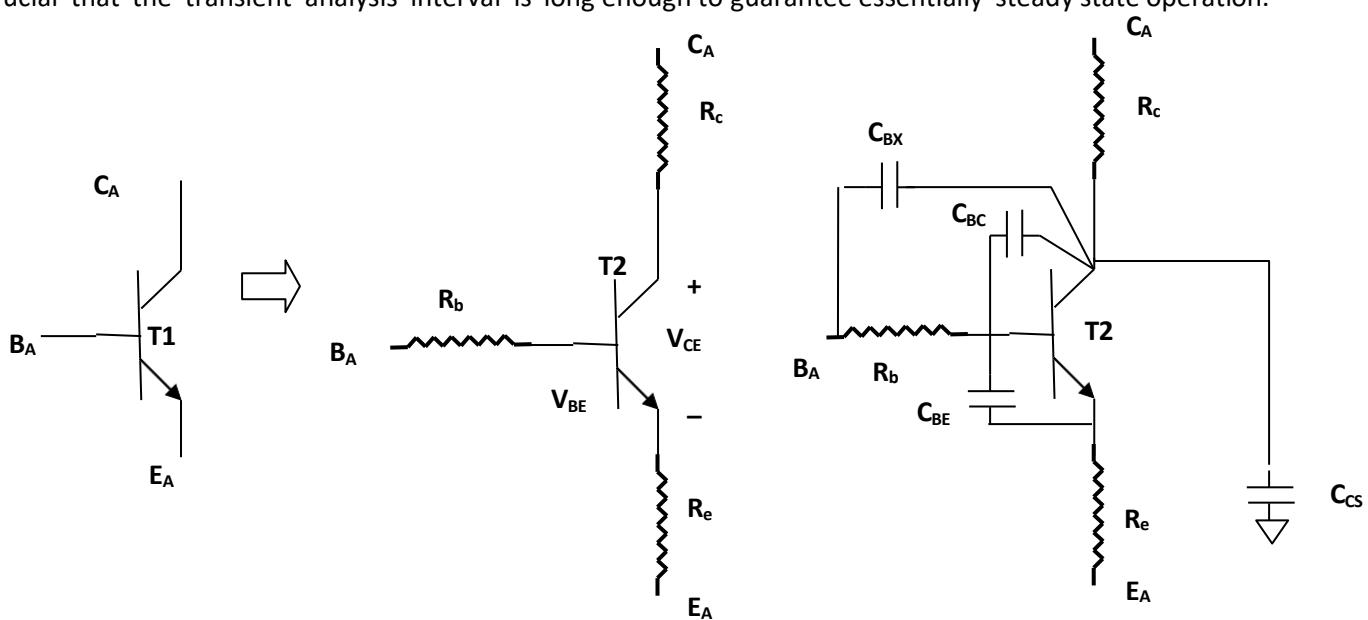


Fig. 3.5 : Modeling of BJT in SPICE(A)DC Modeling (B)High frequency BJT Modeling

A series resistance is modeled in series with each lead of the BJT as shown in fig. 3.5. These resistances are input parameters in SPICE. In the model, all formulation is relative to transistor T2.

High-Frequency BJT Model

The high-frequency model of the BJT used in SPICE is obtained by adding three parasitic capacitors to the transistor T2 in Fig. 3.5 and one capacitor, CBX, between the base of T1 (node BA) and the collector of T2 (node C). The parasitic capacitors are, in general, voltage dependent. The capacitor values used in SPICE are defined by the derivative

$$C = \frac{dQ}{dV}$$

where Q is the charge on the capacitor and V is the corresponding port voltage. Seventeen parameters are used to characterize the four parasitic capacitors in SPICE.

The parasitic capacitors are, in general, modeled by the parallel combination (sum) of a depletion region capacitor and a capacitor which occurs due to charge accumulation in the depletion region. The capacitor attributed to the charge accumulation is current dependent. Finally, under forward bias the model for the parasitic capacitors has a different parametric form, than under reverse bias. The voltage where transition must be made from the reverse bias parametric capacitance model to the forward bias parametric model is termed the transition voltage.

The depletion region reverse-biased junction capacitors are defined by the

$$C_{BEDR} = A_n \cdot C_{JE} (1 - V_{BE} / \Phi_B)^{-M_{JE}}$$

$$C_{BCDR} = \theta A_n \cdot C_{JC} (1 - V_{BC} / \Phi_C)^{-M_{JC}}$$

$$C_{CSDR} = A_n \cdot C_{JS} (1 - V_{CS} / \Phi_S)^{-M_{JS}}$$

$$C_{XSDR} = (1 - \Phi_S) A_n \cdot C_{JC} (1 - V_{BC} / \Phi_C)^{-M_{JC}}$$

where M_{JE} , M_{JS} , M_{JC} , C_{JE} , C_{JS} and C_{JC} are input parameters. Φ_B , Φ_C , Φ_S and θ are also SPICE input parameters, denoted by PB, PS, PC and XCJC. The parameter θ represents the percentage of the B-C capacitance which is associated with the internal base node (base of T2 in fig. 3.5). V_{BC} is the voltage from the base of T1 to the collector of T2 in fig. 3.5. The transition voltages V_{BET} , V_{BCT} , V_{CST} and V_{XST} are FC, Φ_B , FC, Φ_C , 0 V and FC, Φ_C respectively where FC is a SPICE input parameter used to characterize the transition.

BJT NOISE MODEL:

Five noise sources are used to model the noise characteristics of BJT. Thermal resistance noise source are characterized by current source with spectral density of

$$S_{NR} = \frac{4kT}{R_x} \quad \text{for } x \in \{b, e, c\}$$

These are modeled in parallel with three resistors, R_b , R_e , R_c , of fig.3.5. Shot and flicker noise are modeled by two current sources, the first with spectral density of

$$S_{NB} = 2q I_{CQ}$$

Is connected from the base of emitter, of T₂. The second current source has spectral density of

$$S_{NC} = 2qI_{BQ} + \frac{K_F I_{BQ}}{f}^{AF}$$

and is connected from the collector to the emitter of T₂ in the same figure. K_F and AF are SPICE, input parameters and f is frequency in hertz. I_{BQ} and I_{CQ} represent the quiescent values of I_B and I_C respectively. All noise source the BJT are assumed to be uncorrelated.

Temperature Dependence of BJT:

Several of the parameters that characterize the BJT are temperature dependent. SPICE models the temperature dependence of the saturation currents I_s , I_{SE} and I_{SC} , betas (BF and BR), the junction capacitance parameters (C_{JE} , C_{JC} , C_{JS} , Φ_B , Φ_C , Φ_S) and the noise coefficient K_F and A_F . The saturation currents at temperature T are characterized by equations

$$I_s(T) = I_s(T_1) \frac{T}{T_1}^{XTI} \exp \left[\frac{T}{T_1} - 1 \right] \frac{EG(T).q}{kT}$$

$$I_{SE}(T) = I_{SE}(T_1) \frac{T}{T_1}^{\{(XTI/NE)-XTB\}} \exp \left[\frac{T}{T_1} - 1 \right] \left[\frac{EG(T).q}{kT} \right]$$

$$I_{SC}(T) = I_{SC}(T_1) \frac{T}{T_1}^{\{(XTI/NC)-XTB\}} \exp \left[\frac{T}{T_1} - 1 \right] \left[\frac{EG(T).q}{NC.kT} \right]$$

Where T_1 is any reference temperature.

The parameters BF and BR are given by expressions



$$BF = BF(T) \frac{T}{T_1} \text{ XTB}$$

$$BR = BR(T) \frac{T}{T_1} \text{ XTB}$$

The temperature dependence C_{JE} , C_{JC} , C_{JS} , ϕ_B , ϕ_C , ϕ_S is given for $Y \in \{C, E, S\}$ by

$$C_{JY}(T) = C_{JY}T_1 \cdot \{1 + \theta_Y(T)\} \text{ where}$$

$$\theta_Y(T) = M_{JY} \{0.0004(T - T_1) + 1 - * \phi_Y(T) / \phi_Y(T_1)\} \text{ and}$$

$$\phi_Y(T) = \phi_Y(T_1) \cdot (T/T_1) + \phi_{BF}(T)$$



**Subject Name:** VLSI Design

Subject Notes
UNIT-IV

Structured Digital Circuits and Systems:

Random Logic and Structured Logic Forms, Register Storage Circuits, Quasi Static Register Cells, A Static Register Cell, Micro coded Controllers, Microprocessor Design, Systolic Arrays, Bit-Serial Processing Elements, Algotronix .

Random logic forms:

Random logic is describing a particular style design of any digital circuit. Some IC circuit are placed within a layout in much the same way that small scale chip are placed on wire wrap circuit board and then interconnected. The particular types of small- scale logic functions may be needed at irregular places within a circuit, the circuit packages and their interconnection wiring sometimes appear to have been randomly placed.

Random logic is a tag commonly used to describe digital circuits that lack regularity of circuit function, placement and interconnection.

Advantages of random logic:

Efficient use of silicon. Fast operation.

Disadvantages:

Lengthy IC layout times. Difficulty of testing. Costly modification step

Structured Logic Forms:

Structure logic is term used to distinguish logic forms that do show reliability in their layout and interconnection.

Digital integrated circuits have been designed with highly structured layouts for many years. Most notable among there are all forms of memory chips. Memory chips, such as the 1M-bit dynamic RAM from Texas Instruments are composed of many identical memory cells and are naturally structured as regular arrays of these cells. Because of the potential sales volume for many parts, considerable effort is expended in reducing the size of the basic memory cell, causing memory chips to be among the densest of all IC's.

Because of the complexity of many new chips, random logic design is no longer feasible for large chips.

Most new digital integrated circuits increasingly use structured logic forms such as PLA's, micro program ROMs, gate arrays and standard cells to displace random logic design.

Regularity factor

Regularity factor is defined as the ratio of the total number of transistors on the chip to the drawn transistors, where total transistors includes all possible ROM and PLA transistor placements.

A design that requires a unique layout for a circuit element and then uses this circuit element n times without change would exhibit a regularity factor of n.

A design with unique layout for each circuit component would exhibit a regularity factor of 1. For a given complexity of design, a higher regularity factor normally indicates reduced design and layout costs.

Table 4.1: Regularity Factor for microprocessor

Chip name	Numbered of devices	Regularity Factor
8080	4600	1.1
8085	6200	3.1
8086	29000	4.4
68000	68000	12.1

Register Storage Circuits:

1) Quasi Static Register Cells: Figure 4.1 shows a way to combine two inverters, two pass transistor and non overlapping two phase clock to provide a quasi static register cell. Quasi-static register cell uses exactly the same components as a two stage dynamic shift register and interconnected in a different way.

The output of a first inverter is connected directly to the input of a second invertors one pass transistor called the input pas transistor, controls the input to the first inverter. The second pass transistor called the feedback transistor controls a feedback path from the output of the input of the first inverter.

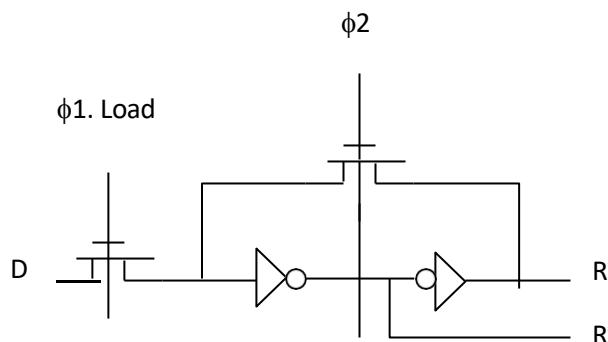


Fig. 4.1 Quasi Static Binary storage cell

Operation: When a binary value is to be stored in the register cell, the input pass transistor is turned on and the feedback transistor is turned off. This is accomplished through use of a load signal ANDed with clock phase ϕ_1 to control the gate of the pass transistor. ϕ_2 is low so that the feedback path is broken at this time.

When the input pass transistor is turned on, any signal applied to D input of the register cell is passed to the gate of the first inverter, resulting in the same logic value at the output R of the second inverter (after two successive inversions)

When the input pass transistor is turned off, the value at the input node of the first inverter is stored dynamically on the parasitic capacitance of that node. The value at the output of the-second inverter is actively driven and is logically equivalent to the stored value at the input of the first inverter.

During the ϕ_2 clock phase the output of the second inverter is fed back to the input of the first inverter, thus reinforcing its logic value. As long as this feedback condition is applied often enough, the quasi-static register cell will maintain its stored value.

Quasi-static register cells were common in early microprocessors. For example, registers in the Motorola 6800 series of microprocessors were composed of an extension of the basic quasi-static cell that permitted dual-port read and write as shown in fig 4.2 . This cell, provides two gated load (write) signals on one clock phase, so the register can be loaded from either of two buses. A feedback path to refresh the stored logic value is provided on the alternate clock phase. The controller (not shown) that generates the write signals should logically AND them with ϕ_1 to avoid conflict with the feedback path that is controlled by ϕ_2 . The register output, taken from the center of the register cell, drives a pull-down transistor. The output of this transistor is directed through pass transistors to one of two possible buses providing dual-port read. This cell requires four control signals (each externally gated by clock phase), an alternate clock signal ϕ_2 to control the feedback path, two bus lines (each bus line is common to one input and one output path), power, and ground.

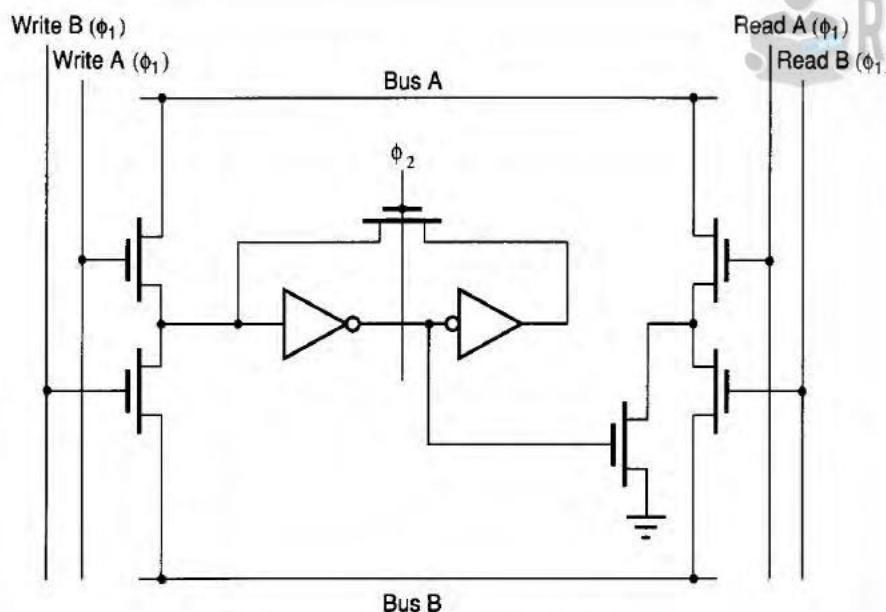


Fig. 4.2 Microprocessor Register cell

2) A Static Register Cell :

Fully static register cells are frequently used within finite state machines and within microprocessor register arrays. One such static cell is based on the classical cross-coupled set-reset (SR) latch shown in Fig.4.3 (a). This latch uses two cross-coupled NOR gates to achieve data storage. An equivalent NMOS transistor level circuit for this latch is given in Fig. 4.3 (b). That this is a static register cell is obvious because the storage does not depend on clock signals, but only on a directly coupled feedback path.

To explain the static register cell operation the SR latch circuit of fig. 4.3(b) will be transformed into a static register cell in two steps.

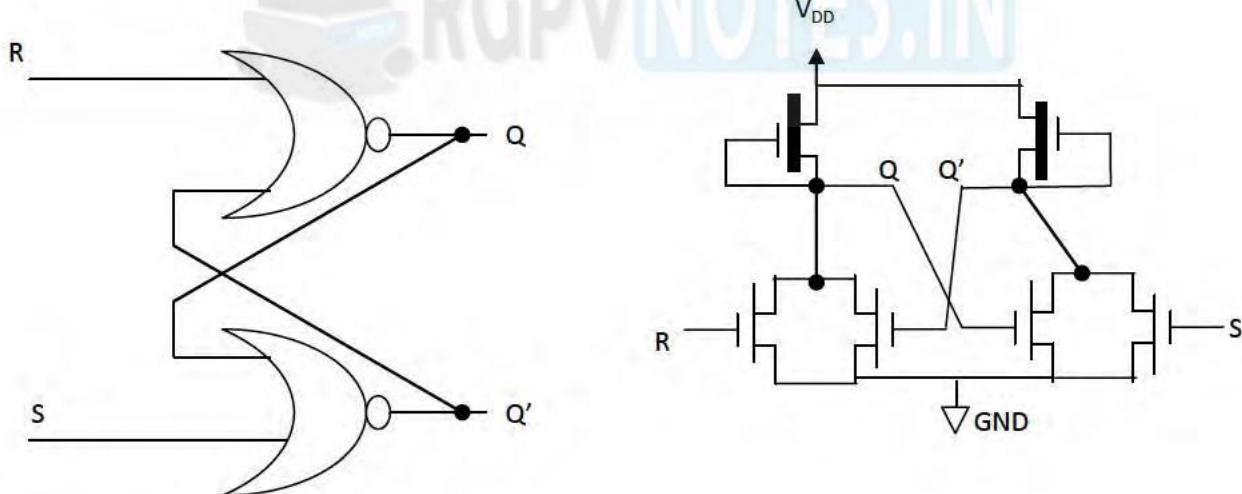


Fig.4.3: Cross coupled NOR latch (a) Logic (b) circuit

Fig. 4.4 (a) shows the previous circuit split into a cross coupled inverter pair with the set and reset pull-down transistors physically separated from the storage element by bus lines. These buses hold signals representing the register cells logic state-and-its complement. Fig.4.4 (b) completes the transformation by including pass transistors between the outputs of the cross-coupled inverter pair and the buses to the set and reset pull-down transistors. The pass transistors provide a way to isolate the register cell from the buses. This basic static register cell consists of six transistors, four for the cross-coupled inverters and two for the connections to the buses. Because the basic register cell of Fig. 4.4 (b) can be isolated from the buses, additional six-transistor register cells can be attached between the same two buses. Then a particular register cell is selected for read or write by selecting (turning on) both pass transistors associated with that cell.

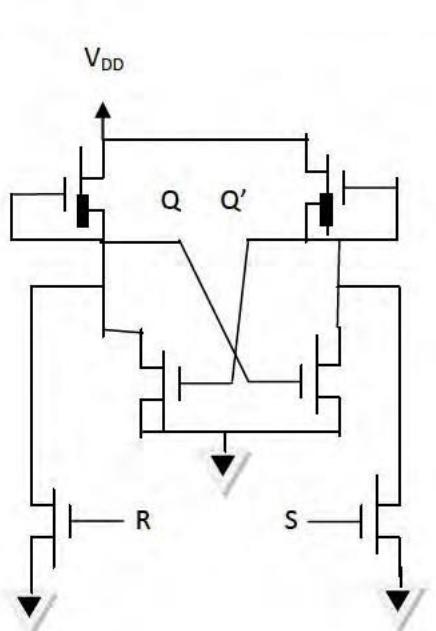


Fig. 4.4 (a) : NMOS static cell storage cell

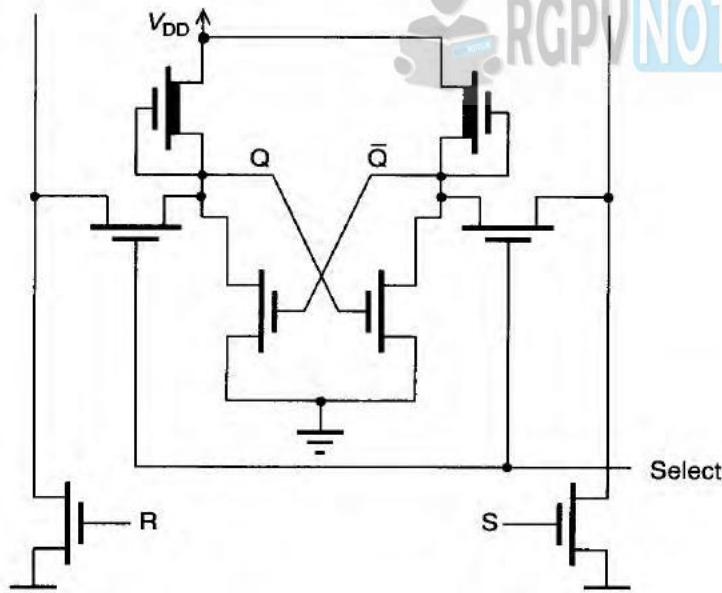


Fig. 4.4 (b) : NMOS static cell storage cell

Microcoded Controllers:

The clocked PLA structure for FSMs is an excellent means to implement small digital controllers. The layout structure is regular and can be generated from the logic equations for a system. For larger digital systems, the logic design to implement a clocked PLA FSM becomes unnecessarily complex and the resulting large PLA, if generated, would be slow. These larger systems require a method that overcomes the disadvantages of a large clocked PLA FSM yet emphasizes regularity in design and layout. A common method to implement complex digital systems in a regular way is to use a memory-based structure known as a microcoded controller. A microcoded controller consists a memory unit and a next address sequencer as shown In Fig. 4.5. The memory unit has microinstructions and next address sequencer directs the execution sequence of micro instructions. A microcoded controller is a special form of computer. A microinstruction is a set of encoded control bits that direct the operation of the logic during a clock cycle.

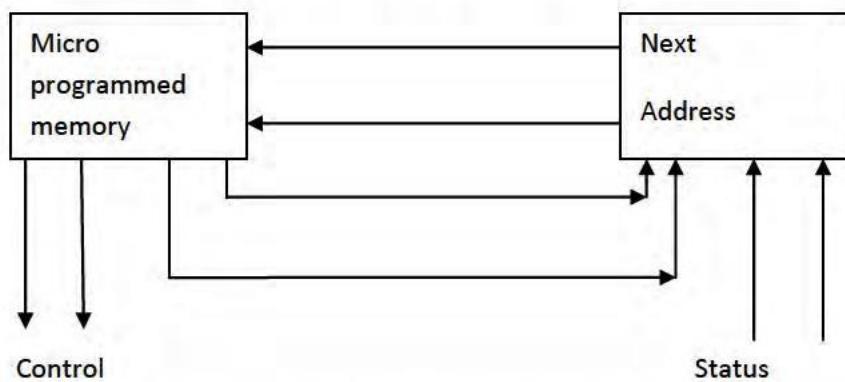


Fig.4.5 Block diagram of simple micro coded controller

The execution hardware is fixed and the functions are a result of instruction placed in the microinstruction memory. This memory is frequently ROM and is thus called micro-ROM.

A microcoded controller is used primarily for larger machines because this controller require the overhead of a next-address sequence that requires design time and integrated circuit area.

The next-address sequence simply generates the next instruction addresses in a fixed pattern for example by incrementing a counter. A microcoded controller configured in this way functions as an open loop controller with a fixed execution sequence. If status inputs are provided the next address sequencer can modify the address of the next instruction, depending on conditions presented by the status inputs. This provides a conditional branching capability.

1) Memory Organization / MicroROM:

Figure 4.6 shows a typical memory organization for a microcoded controller consisting of a microROM and a memory address register (MAR). While most semiconductor memory chips are organized with a wide address bus and a narrow data bus (nine multiplexed address lines and one data line for most 256k DRAMs), the memory (microROM) for a microcoded controller usually has a wide data bus relative to its address bus (perhaps 72 or more data lines compared with 12 or fewer address lines). Most of these data lines are dedicated to driving control points within the system. A few data lines are used to provide next-address information to the next-address sequencer. The next-address sequencer uses this address information along with status inputs from the controlled process to calculate the address of the next microinstruction.

A microROM organized as in Fig. 4.6 would contain almost 300k bits ($2^{12} \times 72 = 294,912$) of control information and would consume a correspondingly large silicon area.

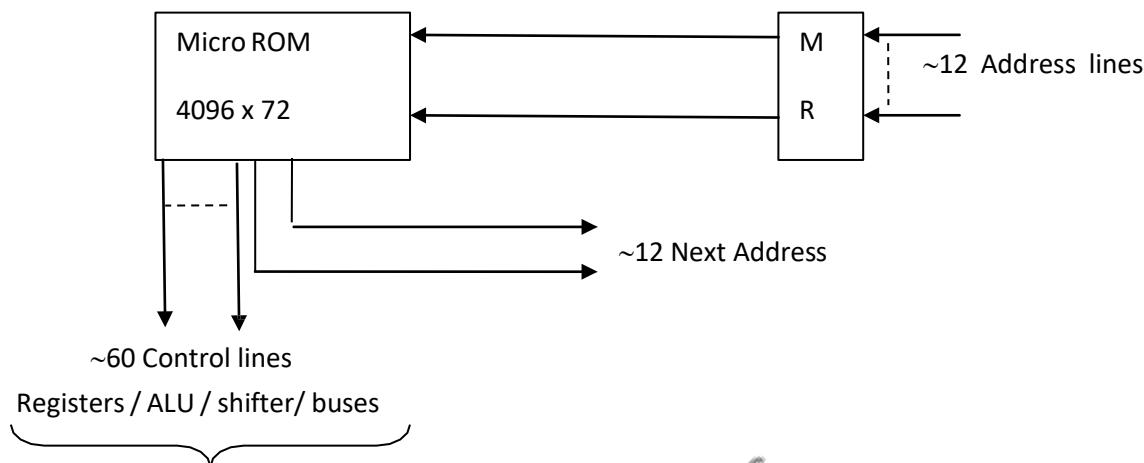


Fig .4.6 :Block diagram of MicroROM architecture

An alternative form for the microROM is shown in Fig. 4.7. This two-level microprogram memory consists of a relatively small micro ROM driving a secondary memory called a nanoROM. This organization is based on two reasonable assumptions. First, only a few of the 272 possible control word combinations of Fig. 4.6 are necessary in a given system. Second, many of the control words that are necessary will be required repeatedly. If fewer than 256 unique control words are necessary, for example, and the microprogram memory is organized as shown in Fig. 4.7, only about 50k bits ($2^{12} \times 8 + 2^8 \times 72 = 51,200$) of control memory are required. This reduction in memory size is not free; the two-level microROM is slower than a single-level memory because a memory access must traverse two memory units to produce data.

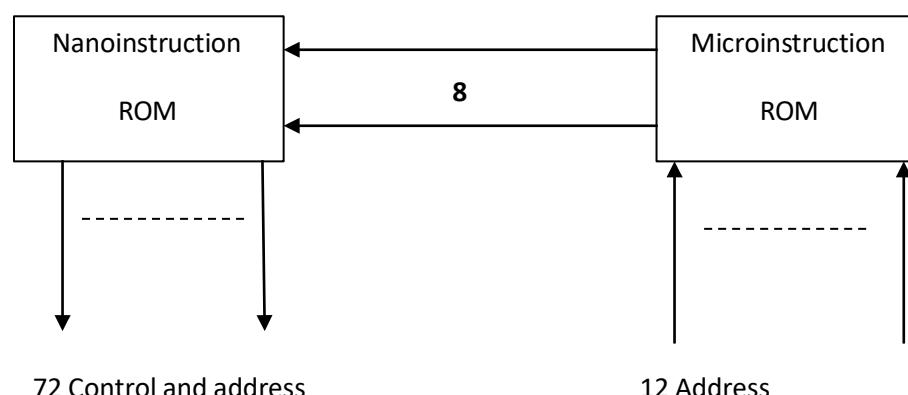


Fig .4.7 : Two level Microprogram memory



2) Comparison Between PLA and Micro-programmed Forms of FMSs:

- (1) A Micro-programmed control unit is more complex than the corresponding PLA FSM because of the next address generation circuitry.
- (2) A PLA FSM can be compiled automatically once the state equation for the system are determined this is much more difficult for a micro-programmed FSM.
- (3) The PLA FSM is normally best for small, simple systems where minimum design time and circuit area are required because the peripheral circuitry for a micro-programmed FSM depends on the application and is thus not automatically generated.
- (4) The microprogram machine is usually more desirable for larger systems over the PLA FSM.

Microprocessor Design:

The evolution of microprocessors provides an interesting study in the development of structured logic forms.

The earliest microprocessors is the Intel 4004 and 8008 were born to counteract the high development costs for custom large-scale integrated circuits.

As the complexity of microprocessor has increased, the design time and costs have also expanded. Development of structured designs using regular logic forms has been required to allow the evolution of microprocessor data path. In which, the data path consists of registers, shifter and ALU and the control unit consists of the micro ROM, MAR, next-address sequences, IR and PC.

The data path for microprocessor usually formed with 8, 16 or 32 bit identical bit path. As a result of their identical bit paths there is an inherent regularity within the data path for microprocessor. The control unit has varied structures with most manufacturers choosing microcoded or PLA controllers.

1) Data path description:

The data path is the place where the microprocessor executes operations such as addition, subtraction, shifts, rotates and Boolean logical functions on data. The data path is also known as execution unit.

A typical n-bit data path structure consisting of a dual-port register array, a barrel shifter, an ALU, interconnection buses and support circuitry. The block diagram of Microprocessor data path is shown in Fig4.8.

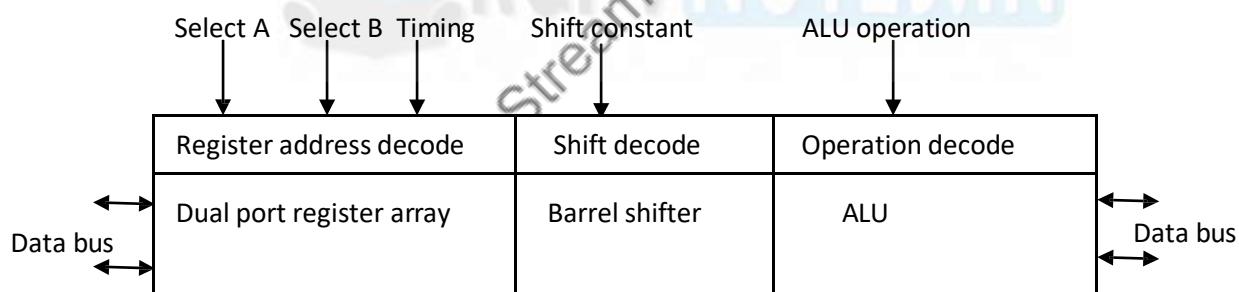


Fig. 4.8 Block diagram of microprocessor data path

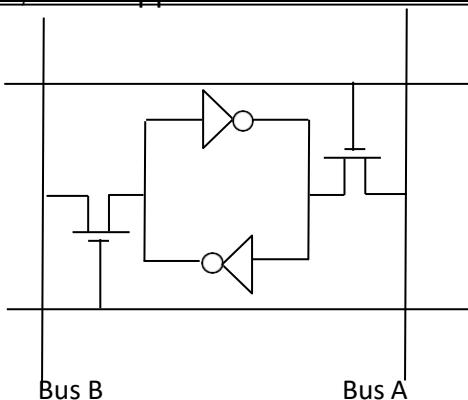
Data flows along n parallel paths in the horizontal direction, while control of the data flow and ALU operations is provided vertically from the top of the data path. Execution of a typical data path operation requires selection of operands from two registers, execution of an operation on the two selected operands, and placement of the result in a register.

The use of a **dual port register array** is convenient for the fast execution of microprocessor programs. This local storage is usually provided within the data path as a small array of static memory cells. These are organized as an n x m structure where n is the width of bits of microprocessor data bus and m is the number of register provided. With a dual-port register array contents from two separate registers can be fetched simultaneous to minimize the delay before execution of an operation can begin.

The memory cell structure of a dual-port register array is quite similar to that of an SRAM cell. The memory cell used for the dual-port register array of the Berkeley RISC processor is shown in Fig.4.9 . In this circuit, the designers took advantage of the provision of double line data access to allow the contents from two registers to be obtained simultaneously.

Select A

RGPVNOTES.IN

**Fig. 4.9 : Memory structure of dual port register cell**

Both data lines must be gated to the storage cell with complementary values for a operation. However the contents of the storage cell may be read by getting the cell to data line. If provision is made to drive the two select lines A and B, separately for operation, then it is possible to obtain the data from a first register along one line of the data bus (bus A), while the data from a second register is obtain along the other rail of the data bus (bus B). So the data from the second bus will be the complement of the cell data must be inverted.

2) Barrel Shifter :

A second component that is included in the data path for many microprocessors is a structure that allows the contents of the data path to be shifted or rotated. A variable-length shift of a bit on the data path requires the possibility of connecting the selected bit to anyone of several other bit paths. A 1-to-n multiplexer circuit for each bit will accomplish the desired connection. An ideal means of implementing multiplexer circuits is provided by the pass transistor available within MOS integrated circuits.

A particularly useful circuit structure to implement a shift or rotate is known as a barrel shifter. This circuit structure can be explained by first considering Fig.4.10, which shows the circuit diagram of a general-purpose bus multiplexer for a 4-bit data path. This multiplexer circuit requires 16 pass transistors to allow connection of any bit line to any other bit line. If each pass transistor could be selected individually, 16 control lines would be required. Because most requirements are for parallel shifts with all bits moved the same number of bit positions, only four shift possibilities are really necessary.

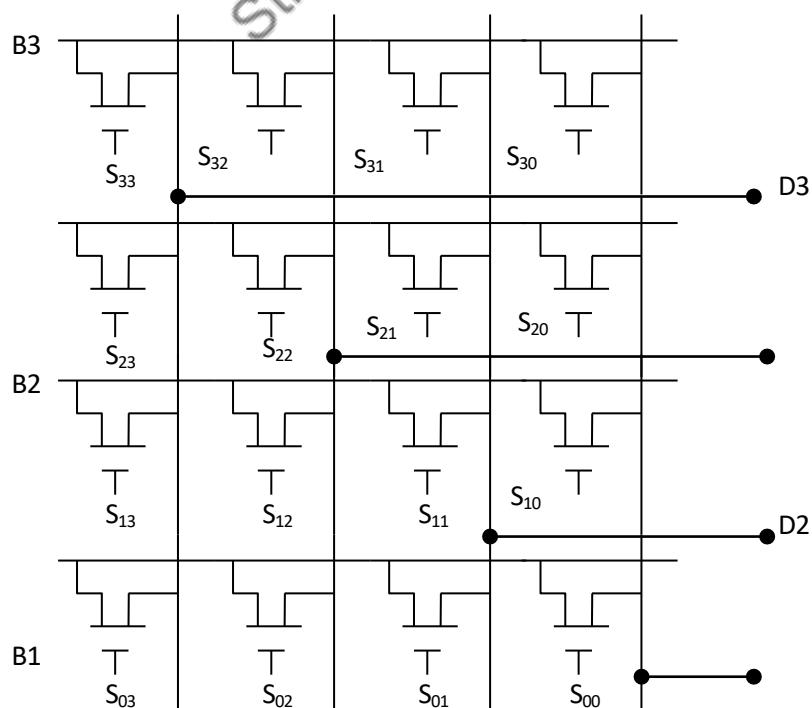
**Fig. 4.10 : General bus multiplexer(16 control line, S₀₀ to S₃₃)**

Figure 4.11 shows a better circuit with the pass transistors connected in groups of four, reducing control line requirements from 16 to 4 separate control lines, 50-53. A particular control line might be selected by encoding a 2-

bit control field to drive a 2-to-4 decoder circuit. The individual decoder output would enable the proper shift control line. For a 32-bit data path, 1024 pass transistors are necessary to allow the desired shift operations. Assuming only parallel shifts, 32 control lines selected by a 5-bit encoded control field are sufficient.

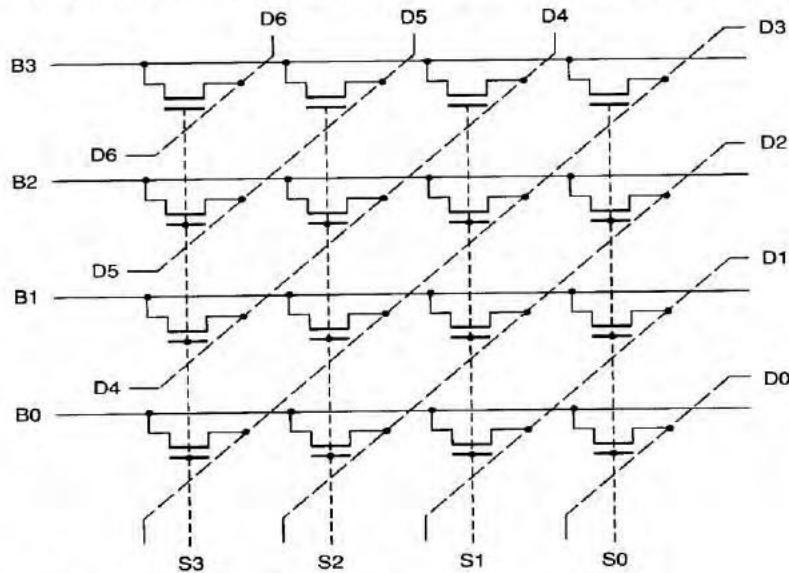


Fig. 4.11 : Barrel Shifter (04 control lines, S0 to S3)

Arithmetic Logic Unit :

The Arithmetic Logic Unit must provide arithmetic and logic operations on data furnished from the data path. A ALU is very important part of the data path. The ALU accepts two operands to perform a specified operation and outputs the result. The block diagram of 32 bit wide ALU is shown in figure 4.12

A and B input of the ALU along with dual port register array having 23 bit buses these input should be provided into the data path between the register array and the ALU.

One of the parallel buses used to return the result to the register array after the ALU operation completed.

The ALU execution time may limit the maximum clock frequency of the microprocessor unless special care is taken for arithmetic operations. These operations are slowed by carry or borrow propagation delays across the width of the ALU. Most microprocessors use carry line with each bit position of the ALU required to generate a carry propagate or a carry-generate signal. In addition newer microprocessors include one or more levels of carry skip circuits to speed carry propagation across groups of adjacent stages.

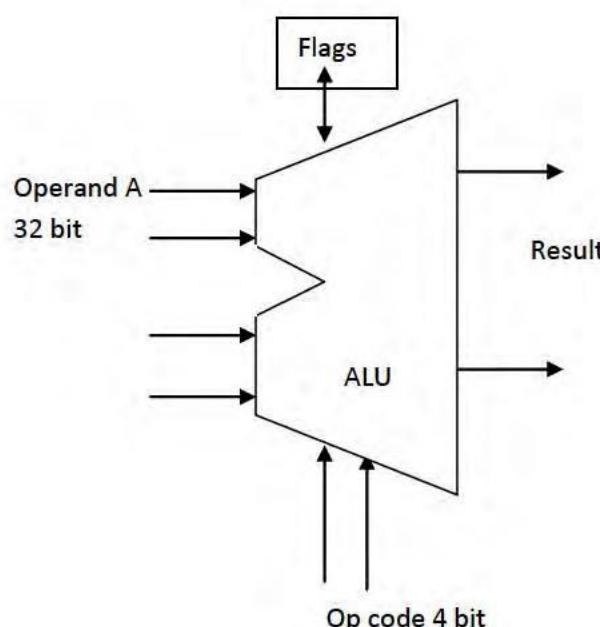


Fig.4.12 : Arithmetic Logic Unit

Systolic Array:

The ability to place hundreds of thousands of transistors on a single integrated circuit and then replicate that circuit inexpensively has led many researchers to propose the use of connected sets of identical integrated circuits to solve problems in parallel. One class of parallel processors has been given the name systolic arrays because the data flow through the array.

The concept of systolic processing combines a highly parallel array of identical processors with local interconnections and rhythmic data flow. The array of processors may span several IC's. The connections are formed so that data the data is accepted and processed at each level stage with result ready for output to the next stage as new data arrives at the current stage.

The objective is to keep most processors busy doing useful work to reduce the time to achieve a result.

1) Bit-Serial Processing Elements:

The prospect for placing n parallel processing stages on a silicon die for large n . The size of the typical processing stages and the interconnection buses require too much silicon area. A partial solution to this problem uses bit-serial processing stages. These stages are smaller than their m -bit parallel counterparts by at least a factor of m , allowing m times as many stages to be placed on a silicon die.

If the processing stages are designed properly, individual stages will interconnect directly as they are placed, thereby eliminating interconnection buses. This technique of interconnection by default is generally useful within IC design saving both layout time and silicon area.

Initially it might appear that bit-serial processing stages are a factor of $1/m$ as fast as parallel processing stages. This could negate the gain achieved by placing m times as many processors on a silicon die. However, bit-serial multipliers and adders can be designed to eliminate carry propagation delay. This allows a net processing speed advantage when m bit-serial processing stages m -bit parallel processing stage.

The ability to create special purpose processors to provide parallel solution of time consuming problems may be the next major step in increasing computational speeds. The concept of systolic processing with many processors working in lock step fashion is an important means to achieve this goal.

An example of a simple bit-serial multiplier is given in Fig. 4.13. The algorithm for this particular bit-serial multiplier requires the multiplicand b in parallel and the multiplier a in bit-sequential form. During the first iteration, the first bit of the multiplier a_0 is input and ANDed with the parallel multiplicand b , producing a set of variables called summands. The summands are added by the full adders to compute a set of partial product bits and the first product bit P_0 . Carries and partial product bits are saved and shifted through unit delay registers so they are available for the next step. As the second multiplier bit a_1 is shifted in, it is ANDed with the multiplicand and added to previous carries and partial product bits. This produces a second bit of the product P_1 . At each iteration the weight of each stage doubles, allowing the carry to be fed back within the same stage. This operation continues until the multiplier a is exhausted and the entire product has been shifted out of the multiplier.

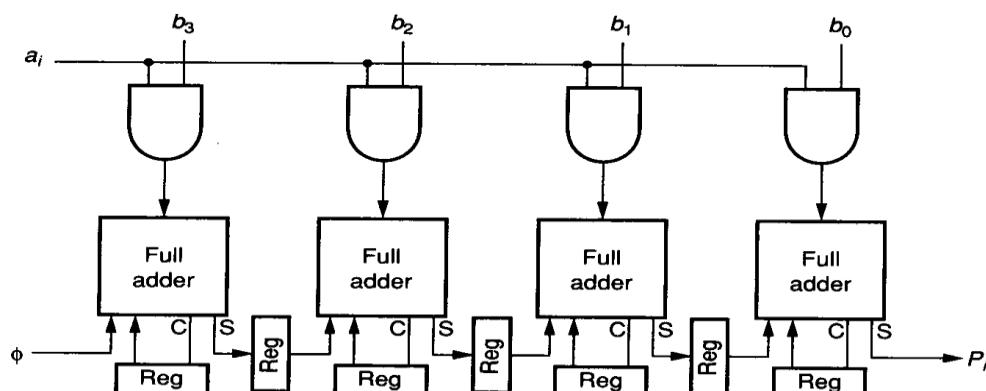


Fig. 4.13: Simple Bit-Serial Multiplier

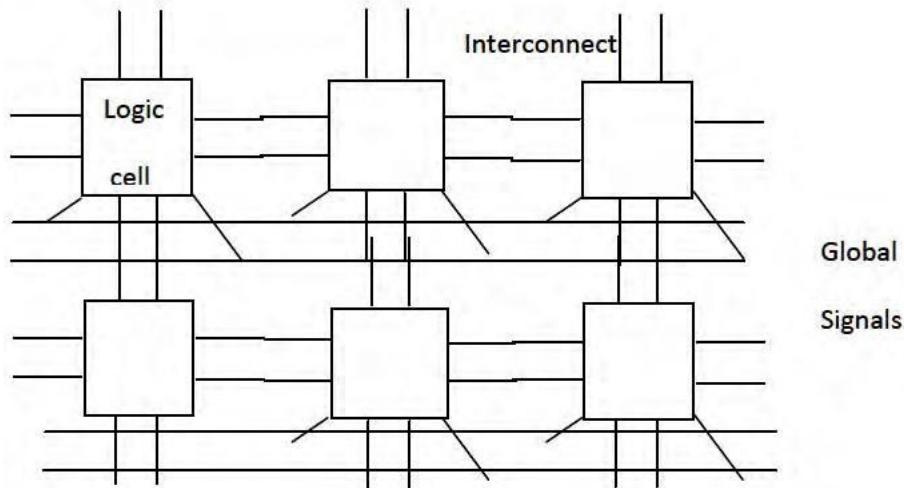
Algotronix:

An example of a regular programmable array is the CAL 1024 (Configurable Array logic) from Algotronix. This architecture contains 1024 identical logic cells arranged in a 32 matrix.

At the boundary of the chip 128 programmable Input output pins allow cascading the chips in even larger arrays.

The cell interconnect is shown in figure 4.14. Each cell is connected to the neighbor at four corner. In addition two global interconnect signal connect to each cell. These are used to supply a low skew signal to all cells for clocking.

Each cell also receives row select lines and bit lines that are used to program RAM bit within the logic cells that dynamically customize the logic cell.



Each logic cell also has RAM data and address lines passing through it.

Fig.4.14 : Algotronix FPGA chip architecture

The input- output pads are very interesting. The trick is to use only one pin for input- output and out of the array but have the communicating chips automatically deal with two pins that are outputs.

The pads achieve this by using three level logic scheme to sense when two outputs are driving each other via a contention circuit. This is then used with an XOR gate as shown in figure 4.15 to deduce the correct input value.

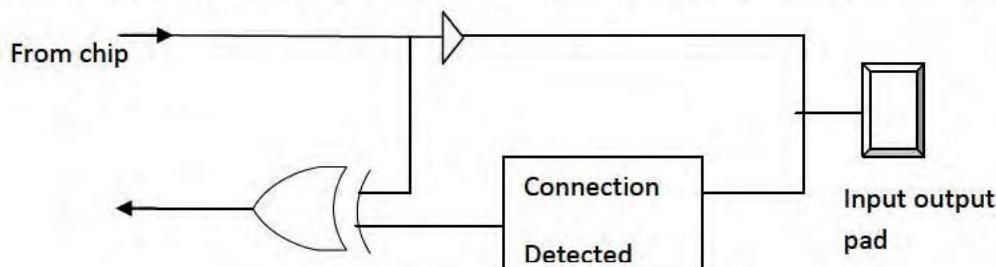


Fig. 4.15 : Algotronix input /output circuit



Subject Name: VLSI Design

Subject NotesUNIT-V**CMOS Processing Technology**

Basic CMOS Technology, A basic n-well CMOS Process, Twin Tub Processes, CMOS Process Enhancement, Interconnects and Circuit Elements, Layout Design Rules, Latch up, Physical Origin of latchup, Latch up Triggering, Latch up Prevention, Internal Latch up Prevention Techniques.

5.1 Basic CMOS Technology:

For fabrication of devices such as nMOS, CMOS etc. following technological process step by step:

Oxidation: Oxidation refers to chemical process of reaction of silicon with oxygen to form silicon dioxide.

Diffusion: Diffusion is the process of introducing controlled amount of dopants into semiconductor.

Ion-Implantation: The process of introduction of high energy charged particles into the substrate. The dopant atoms are vaporized, accelerated and targeted at the substrate.

Epitaxy: The growth of thin ordered crystalline layer on a crystalline substrate is known as epitaxy. During epitaxy the substrate acts as the seed crystal and the crystal can be grown below the melting point. Epitaxy is generally used to enhance the performance of bipolar-transistor and other CMOS IC's also.

Lithography: Lithography is the process of transferring the patterns of geometric shapes on a mask. The pattern define the regions of the integrated circuit that have to be fabricated.

Etching: The process of removing material selectively is known as etching.

Metallization: Metallization is the formation of metal films used for inter-connections, ohmic contacts and rectifying metal-semiconductor contacts.

The required process flow steps for patterning of silicon dioxide is shown figure5.1. These steps are discussed below:

Step 1: First, we take a silicon substrate as Fig.(a).

Step 2: The thermal oxidation is used to created oxide layer of about 1 mm thickness on the substrate [Fig. (b)].

Step 3: After that, the entire oxide surface is covered with the a layer of photoresist [Fig. (c)].

This photoresist is essentially a light-sensitive, acid-resistant organic polymer, initially insoluble in the developing solution.

Step 4: The photoresist is then exposed to UV light through a mask, which defines those regions into which diffusion is to take place together with transistor channels. In the areas where the UV light can be pass through the photoresist is exposed and becomes soluble [Fig. (d)].

Step 5: The unexposed portions are subsequently etched away together with the underlying SiO₂ (silicon dioxide) so that the wafer surface is exposed in the window defined by mask [Fig. d]

Step 6: After the Step-5 (or) at the end of Step-5, we obtain an oxide window that reaches down to the silicon surface [Fig. d]

Step 7: The remaining photoresist is removed from the silicon dioxide surface by using another solvent. The leaving patterned silicon dioxide feature on the surface as shown in [Fig. e]

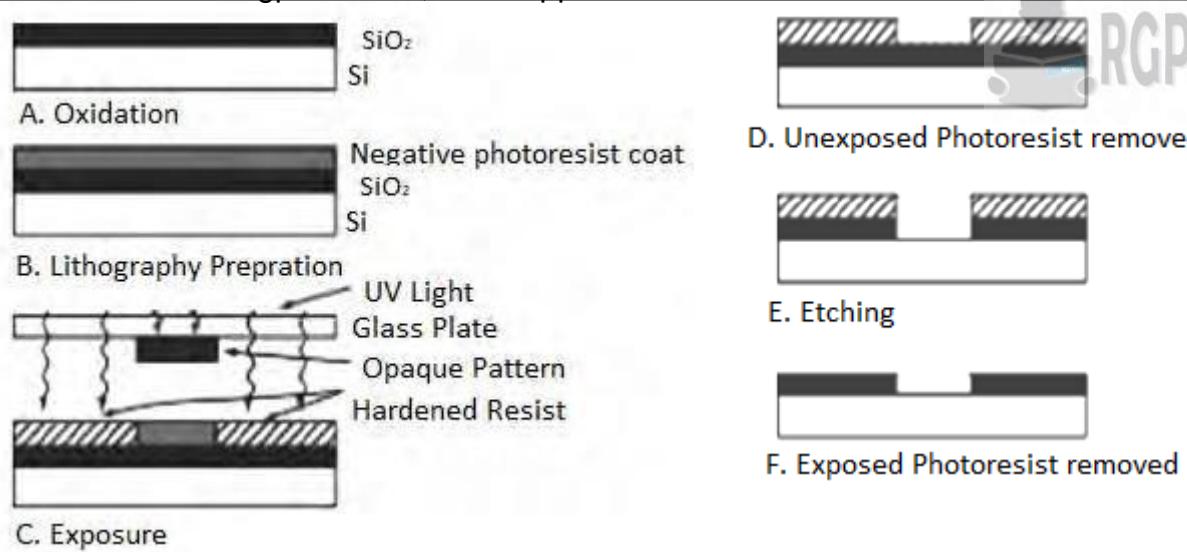


Figure 5.1: Process steps required for patterning silicon dioxide

5.2 A Basic n-well CMOS Process:

There is a requirement of both n-channel (nMOS) and p-channel (pMOS) transistors on the same chip substrate for the fabrication of CMOS. To accommodate both nMOS and pMOS devices, special regions must be created in opposite type substrate of semiconductors. These regions are known as wells or tubs.

CMOS provides an inherently low power static technology that has the capability of providing lower power delay than bipolar, nMOS or GaAs technologies.

The four main CMOS technologies are:

- 1) p-well process
- 2) n-well process
- 3) Twin-tub process
- 4) Silicon on insulator process

The **n-well fabrication** has gained wide acceptance.

N-well CMOS circuits are superior to p-well because of the lower substrate bias effect on transistor threshold voltage and inherently lower parasitic capacitances associated with source and drain regions. The flow diagram of the fabrication for nMOS process are illustrated in Fig.5.2: below

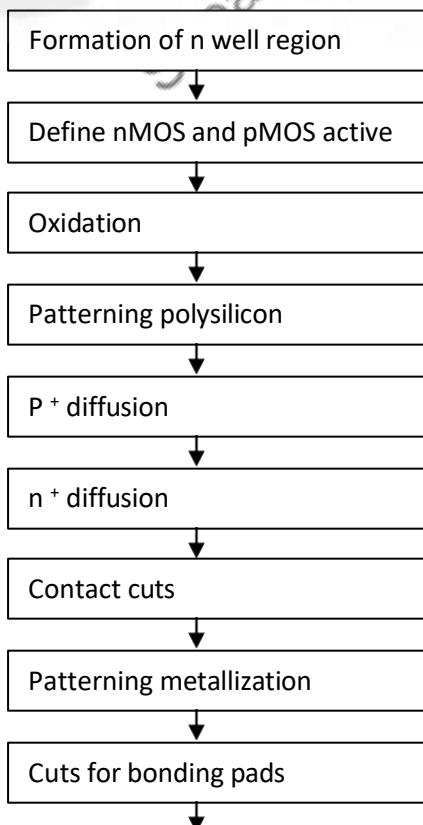


Fig . 5.2 : Fabrication step for NMOS

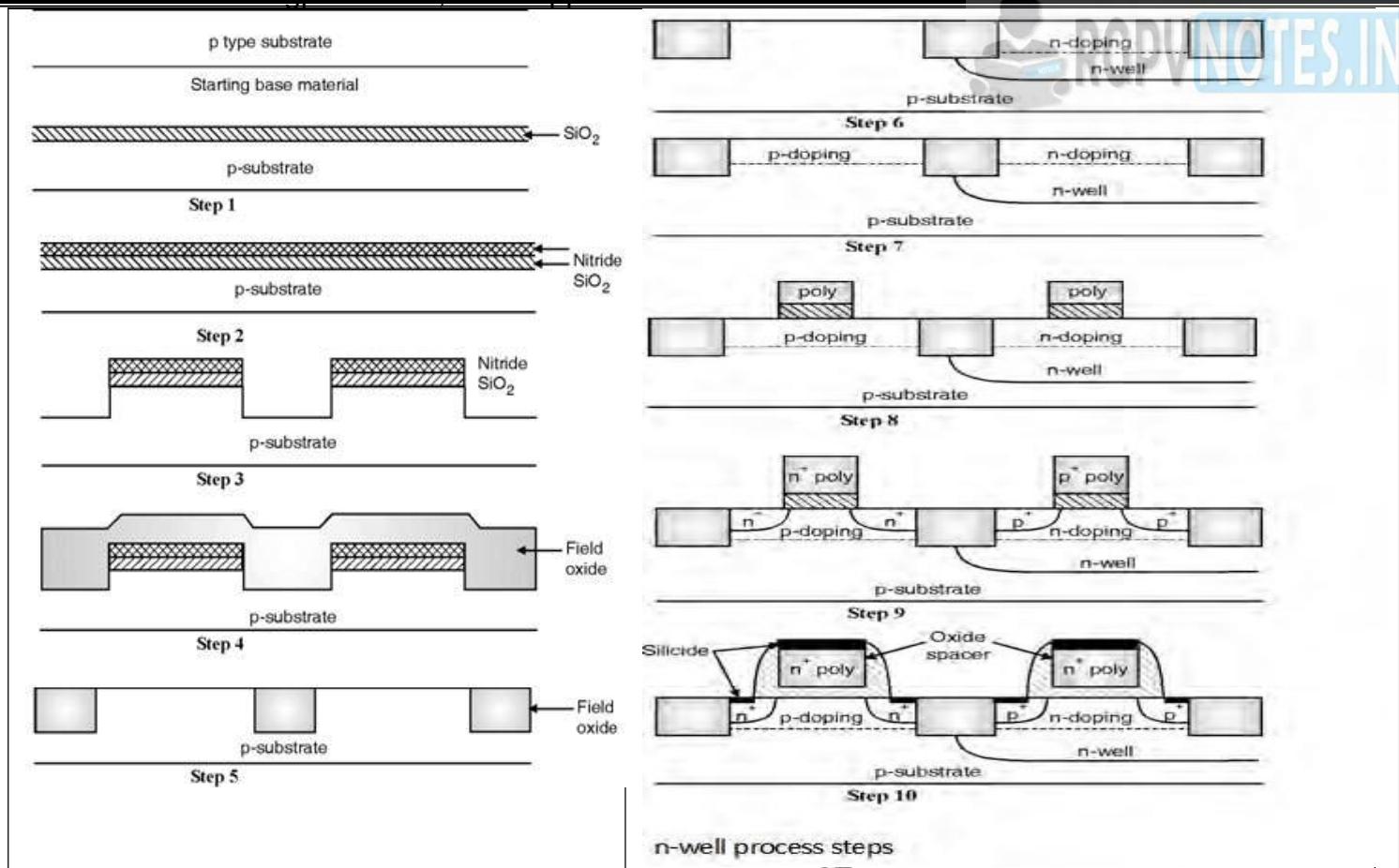


Fig.5.3: CMOS n- Well process step

The n-wells are created in p-type substrate. The typical processing steps shown in figure 5.3 for fabrication of CMOS n-well devices may be summarized as below:

The process starts with a p-substrate.

Step 1 : A thin layer of SiO_2 is deposited which will serve as the pad oxide.

Step 2 : Deposition of a thicker sacrificial silicon nitride layer by chemical vapour deposition (CVD).

Step 3 : A plasma etching process using the complementary of the active area mask to create trenches used for insulating the devices.

Step 4 : The trenches are filled with SiO_2 which is called as the field oxide.

Step 5 : To provide flat surface, chemical mechanical planerization is performed and also sacrificial nitride is removed.

Step 6 : The n-well mask is used to expose only the n-well areas, after this implant and annealing sequence is applied to adjust the well doping. This is followed by second implant step to adjust the threshold voltage of the PMOS transistor.

Step 7 : Implant step is performed to adjust the threshold voltage of NMOS transistor.

Step 8 : A thin layer of gate oxide and polysilicon is chemically deposited and patterned with the help of polysilicon mask.

Step 9 : Ion implantation to dope the source and drain regions of the PMOS (p^+) and NMOS (n^+) transistors, this will also form n^+ polysilicon gate and p^+ polysilicon gate for NMOS and PMOS transistors respectively. Hence this process is called as self aligned process.

Step 10 : Then the oxide or nitride spacers are formed by chemical vapour deposition.

Step 11 : In this step contact or via holes are etched, metal is deposited and patterned. After the deposition of last metal layer final passivation or overglass is deposited for protection.



5.3 Twin Tub Processes:

A logical extension of the p-well and the n-well approaches is the twin-tub fabrication process.

This process, starts with a substrate of high resistivity n-type material and then create both n-well and p-well regions.

Twin-Tub Process: The twin-tub process allows two separate tubs to be implanted into very lightly doped silicon. This allows the doping profiles in each tub region to be tailored independently so that neither type of devices will suffer from excessive doping effects. The lightly doped silicon is an epitaxially grown layer on heavily doped silicon substrate. The substrate can be either n type or p-type. The process sequence for a CMOS twin-tub process is discussed as:

The process starts with a p-substrate surfaced with a lightly doped p-epitaxial layer as shown in figure 5.4.

Step 1 : A thin layer of SiO₂ is deposited which will serve as the pad oxide.

Step 2 : A thicker sacrificial silicon nitride layer is deposited by chemical vapour deposition.

Step 3 : A plasma etching process is used to create trenches used for insulating the devices.

Step 4 : The trenches are filled with SiO₂ which is called as the field oxide.

Step 5 : To provide flat surface chemical mechanical planarization is performed and also sacrificial nitride and pad oxide is removed.

Step 6 : The p-well mask is used to expose only the p-well areas, after this implant and annealing sequence is applied to adjust the well doping. This is followed by second implant step to adjust the threshold NMOS transistor.

Step 7 : The n-well mask is used to expose only the n-well areas, after this implant and annealing sequence is applied to adjust the well doping. This is followed by a second implant step to adjust the threshold voltage of PMOS transistor.

Step 8 : A thin layer of gate oxide and polysilicon is chemically deposited and patterned with the help of polysilicon mask.

Step 9 : Ion implantation to dope the source and drain regions of the PMOS (p⁺) and NMOS (n⁺) transistors is used this will also form n⁺ polysilicon gate and p⁺ polysilicon gate for NMOS and PMOS transistors respectively.

Step 10 : Then the oxide or nitride spacers are formed by chemical vapour deposition (CVD).

Step 11 : In this step contact or holes are etched, metal is deposited and patterned. After the deposition of last metal layer final passivation or overglass is deposited for protection.

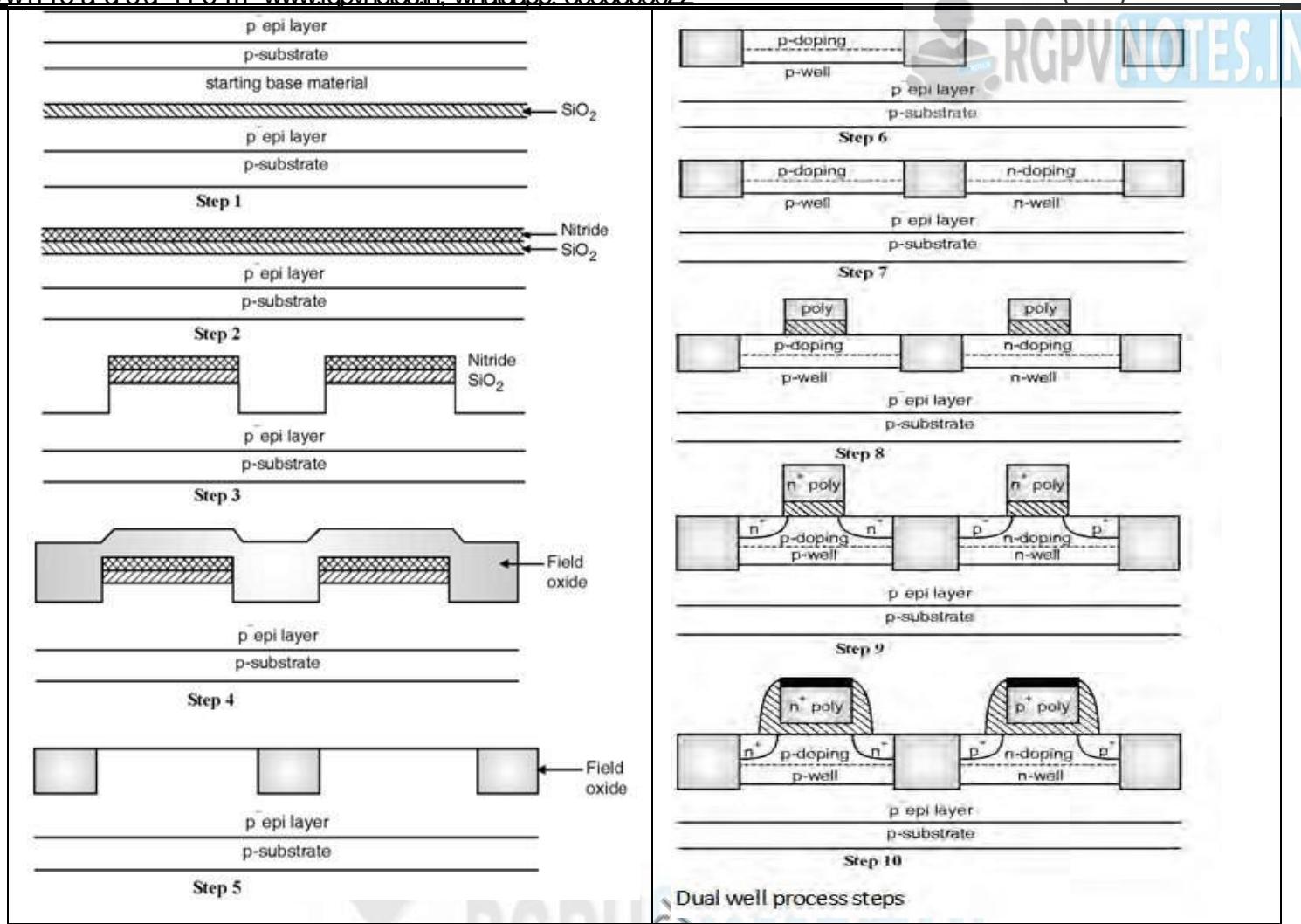


Fig.5. 4 : Twin tub process structure

5.4 CMOS Process Enhancement:

A number of enhancements may be added to the CMOS processes, primarily to increase reliability of circuits provide high-quality capacitors for analog circuits and memories, or provide resistors of variable characteristics.

These enhancements include

- Double- or triple- or quadruple-level metal (or more).
- Double- or triple-level poly (or more) .
- Combinations of the above.

5.4.1 Metal Interconnect

A second level of metal is almost mandatory for modern CMOS digital design. A third layer is becoming common and is certainly required for leading edge high density, high-speed chips. Normally, aluminum is used for the metal layers. If some form of planarization is employed the second-level metal can be the same as the first. As the vertical topology becomes more varied, the width and spacing of metal conductors has to increase so that the conductors do not thin and hence break at vertical topology jumps.

Contacting the second-layer metal to the first-layer metal is achieved by a via, as shown in fig.5.5.

If further contact to diffusion or polysilicon is required, a separation between the via and the contact cut is usually required.

This requires a first-level metal tab to bridge between metal 2 and the lower-level conductor. It is important to realize that in contemporary processes first-level metal must be involved in any contact to underlying areas.

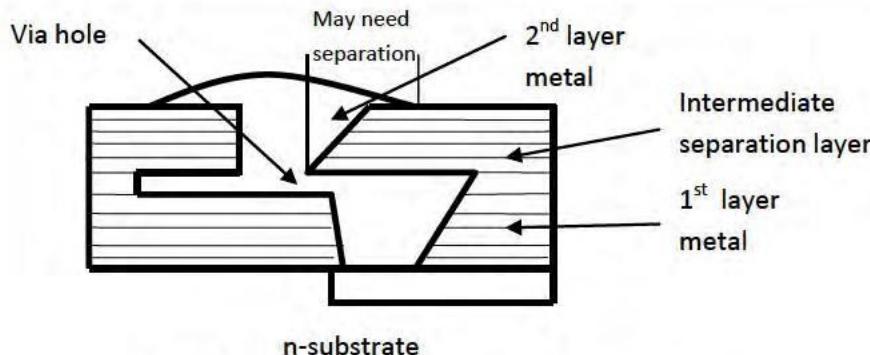


Fig.5.5 Two level metal process cross section

5.4.2 Circuit Elements

Resistors

Polysilicon is highly resistive. This property is used to build resistors that are used in static memory cell. The process step is achieved by preventing the resistor areas from being implanted during normal processing.

Resistors in the tera ohm region are used.

For mixed signal CMOS a resistive metal such as nichrome may be added to produce high value, high quality resistors. The resistor accuracy might be further improved by laser trimming the resulting resistors on each chip to some predetermined test specification.

In this process a high powered laser vaporizes areas of the metal resistor until it meets a measurement constraint.

Capacitors

Good-quality capacitors are required for switched-capacitor analog circuits while small high value area capacitors are required for dynamic memory cell.

Both types of capacitors are usually added by using at least one extra layer of polysilicon, although the process techniques are very different.

Polysilicon capacitors for analog applications are the most straightforward. A second thin oxide layer is required in order to have an oxide sandwich between the two polysilicon layers yielding a high capacitance/unit area. Fig. 5.6 shows a typical polysilicon capacitor.

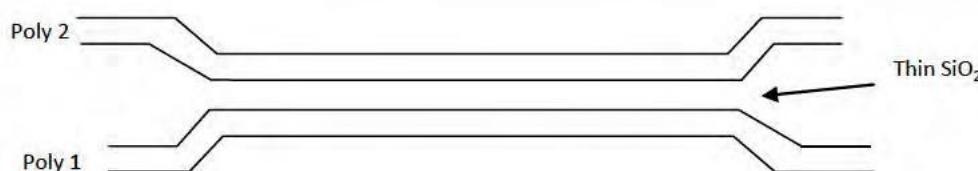


Figure 5.6: Polysilicon capacitor (process cross section)

Bipolar Transistors

The addition of the bipolar transistor to the device forms the basis for BiCMOS processes. Adding an npn-transistor can clearly aid in reducing the delay times of highly loaded signals, such as memory word lines and microprocessor busses. For analog applications bipolar transistors may be used to provide better performance analog functions than MOS alone. To get merged bipolar/CMOS functionality, MOS transistors can be added to a bipolar process or vice versa. In past days, MOS processes always had to have excellent gate oxides while bipolar processes had to have precisely controlled diffusions.

The cross section of a mixed signal BiCMOS process is shown in fig.5.7 . This process features both npn and pnp transistors in addition to pMOS and nMOS transistors.

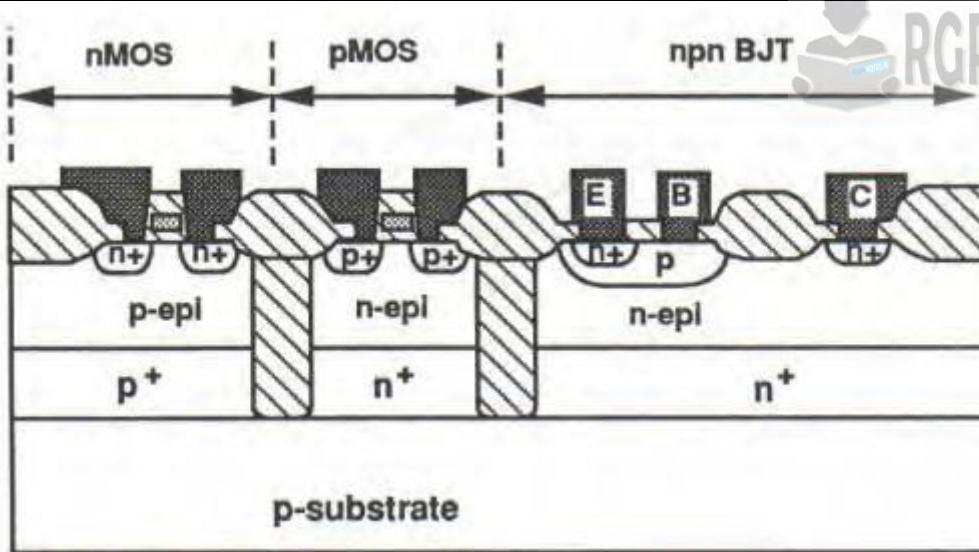


Figure 5.7: Cross section of a mixed signal BiCMOS process

5.5 Layout Design Rules:

Design rules 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.

Lambda rules, which specify the layout constraints in terms of a single parameter 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. It must be emphasized, however, that most of the submicron CMOS process design rules do not lend themselves to straightforward linear scaling. The use of lambda-based design rules must therefore be handled with caution in sub-micron geometries. In the following, we present a sample set of the lambda-based layout design rules devised for the MOSIS CMOS process and illustrate the implications of these rules on a section of a simple layout which includes two transistors.

Design rules for a typical p-well CMOS process

Sr. No.	Rule number	Description	Dimensions	
			Microns	Scalable
1	p-well Layer			
1.1	Width		5	4λ
1.2	Spacing to well at different potential		15	10λ
1.3	Spacing to well at same potential		9	6λ
2	Active (Diffusion) Layer			
2.1	Width		4	2λ
2.2	Spacing to active		4	2λ
2.3	P+ active in n-subs to p-well edge		8	6λ
2.4	n+ active in n-subs to p-well edge		7	5λ
2.5	n+ active in p-well to p-well edge		4	2λ
2.6	p+ active in p-well to p-well edge		1	λ
3	Poly			
3.1	Width		3	2λ
3.2	Spacing		3	2λ
3.3	Field poly to active		2	λ
3.3	Poly overlap of active		3	2λ

	3.4	Active overlap of poly	4	2λ
4	P+ Select			
	4.1	Overlap of active	2	λ
	4.2	Space to n+ active	2	λ
	4.3	Overlap of channel	3.5	2λ
	4.4	Space to channel	3.5	2λ
	4.5	Space to P+ select	3	2λ
	4.6	width	3	2λ

Table 1 : Layout Design Rules

5.6 Latch up:

In CMOS circuits, the parasitic NPN and PNP bipolar junction transistors (BJT) forms a thyristor or PNPN structure. When the thyristor is triggered by noise or glitch, the BJTs become ON, and it leads to a formation of short circuit between the power and ground lines, known as CMOS latch-up.

Latch-up is a condition in which parasitic components give rise to establishment low-resistance conducting path between V_{DD} and V_{SS} with disastrous results. Careful steps during fabrication are considered in order to avoid these problems. This may be induced by glitches on the supply rails or by incident radiation.

If every silver lining has a cloud, then the cloud that has plagued CMOS is a parasitic circuit effect called latchup. The result of this effect is the shorting of the V_{DD} and V_{SS} lines.

This effect was a critical factor in the lack of acceptance of early CMOS processes, but in current processes it is controlled by process innovations and well-understood circuit techniques.

5.6.1 Physical Origin of latchup:

The source of the latchup effect may be explained by examining the process cross section of a CMOS inverter shown in figure 5.8 (a).

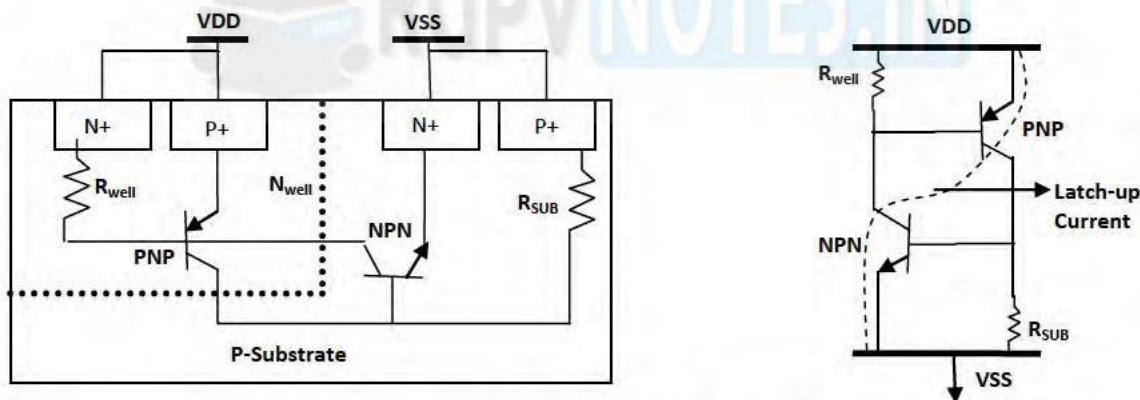


Fig.5.8 (a & b): (a) CMOS inverter (b) model

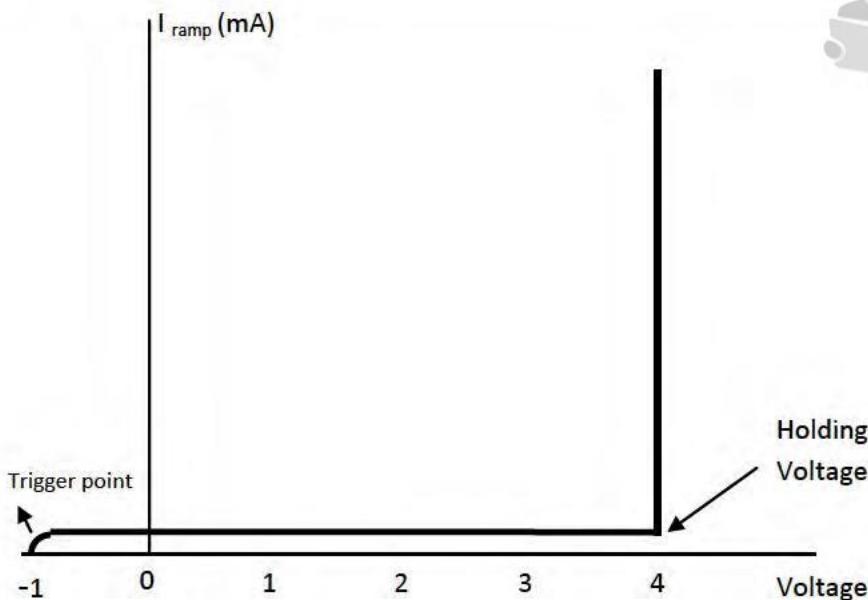


Fig.5.8 (c): VI characteristics of CMOS latchup

The schematic depicts in addition to the expected nMOS and pMOS transistors a circuit composed of an npn transistor a pnp transistor and two resistors connected between the power and ground rails Fig. 5.8(b). Under the right conditions, this parasitic circuit has the VI characteristic shown in fig 5.8 (c), which indicates that above some critical voltage known as the trigger point, the circuit "snaps" and draws a large current while maintaining a low voltage across the terminals known as the holding voltage. This is, in effect, a short circuit. The bipolar devices and resistors is shown in fig 5.8 (b) are parasitical .

Figure 5.8 (a) shows a cross-sectional view of a typical (n-well) CMOS process the pnp transistor has its emitter formed by the p+ source/drain implant used in the pMOS transistors. Note that either the drain or source may act as the emitter although the source is the only terminal that can lay the latchup condition. The base is formed by the n well while the collector is the p-substrate. The emitter of the npn transistor is the n⁺ source or drain implant while the base is the p-substrate and the collector is the n-well. In addition, substrate resistance $R_{substrate}$ and well resistance R_{well} are due to the resistivity of the semiconductors involved.

Consider the circuit shown in figure 5.8 (b). If a current is drawn from the npn emitter, the emitter voltage becomes negative with respect to the base until the base emitter voltage is approximately 0.7 volts. At this point the npn transistor turns on and a current flows in the well resistor due to common emitter current amplification of the npn-transistor. This raises the base emitter voltage of the npn transistor. This in turn raises the npn base voltage causing a positive feedback condition, which has the characteristic shown in figure 5.8 (c).

At a certain npn base-emitter voltage, called the trigger point, the emitter voltage suddenly "snaps back" a nil enters a stable state called the ON state. This state will persist as long as the voltage across the two transistors is greater the holding voltage as shown.

5.6.2 Latch up Triggering:

For latchup to occur, the parasitic npn pnp circuit has to be triggered and the holding state has to be maintained. Latchup can be triggered by transient currents or voltages that may occur internally to a chip during power-up or externally due to voltages or currents beyond normal operating ranges.

Two distinct method of triggering are possible lateral and vertical triggering.

1) Lateral triggering occurs when a current flow in the emitter of the lateral npn transistor.

The static trigger point is set by

$$L_{trigger} = V_{pnp-on} / \alpha_{pnp} R_{well}$$

V_{pnp-on} =0.7 volt turn on voltage of the vertical pnp transistor.

α_{pnp} = common base gain of the lateral npn transistor.

R_{well} =well resistance.



- 2) Vertical triggering occurs when a sufficient current is injected into the emitter of the vertical pnp transistor.

5.6.3 Latch up Prevention

Latchup may be prevented in two basic ways:

- (1) Latchup resistant CMOS processes. (2) Layout techniques

A popular process option that reduces the gain of the parasitic transistors is the use of silicon starling material with a thin epitaxial layer on top of a highly doped substrate. This decreases the value of the substrate resist and also provides a sink for collector current of the vertical pnp transistor.

As the epi layer is thinned, the latchup performance improves until a point where the up-diffusion of the substrate and the down-diffusion of any diffusions in subsequent high-temperature procession steps thwart require device doping profiles. The so called retrograde well structure is also used.

This highly doped area at the bottom of the well, whereas the top of the well is more lightly doped. This preserves good characteristics for the pMOS transistors but reduces the well resistance deep in the well. A technique linked to these two approaches is to increase the holding voltage above the V_{DD} supply.

It is hard to reduce the betas of the bipolar transistors to meet the condition set above. Nominally, for a 1micro n-well process, the vertical pnp has beta of 10-100, depending on the technology.

The lateral npn current gain which is a function of n^+ drain to n-well spacing is between 2 and 5.

5.6.4 Internal Latch up Prevention Techniques:

A few rules may be followed that reduce the possibility of internal latchup :

- Every well must have a substrate contact of the appropriate type.
- Every substrate contact should be connected to metal directly to a supply pad.
- Place substrate contacts as close as possible to the source connection of transistors connected to the supply rails i.e. V_{SS} n-devices, V_{DD} p-devices. This reduces the value of $R_{substrate}$ and R_{well} . A rule would place one substrate contact for every supply (V_{SS} or V_{DD}) connection.
- A less conservative rule is place a substrate contact for every 5-10 transistors or every 25-100 μ .
- Use of P+ and n+ guard ring around nMOS and pMOS connected to the ground and VDD.
- Reduction of R-well and R-substrate as much as possible by placing substrate and well contacts as close as possible to the source connect.
- Keeping sufficient spacing between the nMOS and pMOS transistors.

Some system level approaches to avoid the latchup are:

- Power supplies must be off before plugging. A plus-in power supply ON condition can cause a surge voltage to appear on the signal pins which could trigger the latchup.
- The electrostatic discharge (ESD) can trigger the latchup. ESD protection circuits needs the special care.
- The electron-hole pairs are generated when radiation penetrate into the chip. These carriers can contribute to well or substrate currents, leading to latchup.
- Simultaneous switching of a large number of transistors can cause noise in the power/ground lines, which can drive the circuit into the latchup.