LECTURE 3

VLSI –(CMOS)

REFs

Introduction to Microelectronics to Nanoelectronics

Design and Technology

Manoj Kumar Majumde

LECTURE OUTLINE

- ✓ CMOS
- ✓ VLSI
- ✓ Interconnect Parasitic

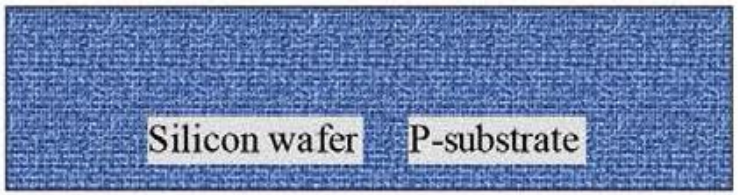
BASIC CMOS TECHNOLOGY

- In VLSI technology, the fabrication of an IC is primarily used to integrate several different components such as a resistor, transistor, diode, and capacitor within the die area based on the requirements.
- This can be only possible with the help of several processes such as oxide growth, epitaxial growth, masked impurity diffusion, photolithography, oxide etching, and metallization.
- Moreover, the transistor in its various forms, such as CMOS, BJT, and FET, etc., is the most complicated element to fabricate over a silicon wafer The different technology approach is used based on the type of transistor built over an IC.

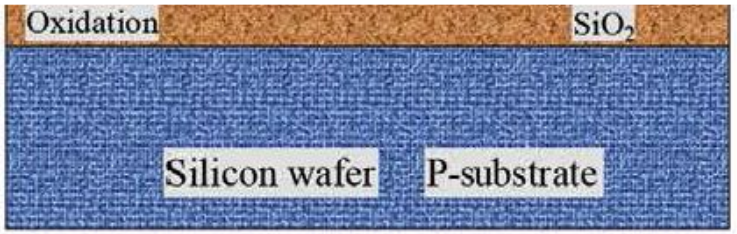
N-WELL AND *P*-WELL CMOS PROCESS

- ✓ The fabrication of CMOS transistors can be obtained by integrating NMOS and PMOS over the same silicon wafer surface by using either *N*-well or *P*-well technology.
- ✓ In *N*-well technology, the formation of CMOS transistor using *N*-well can be obtained, as illustrated in the steps below.

Step-I: This step involved the selection of a *P*-type silicon wafer in order to use *N*-well technology to form the CMOS transistor. Then, the silicon dioxide as a barrier is used to deposit on top of the silicon wafer using an oxidation process such that it gets protected from any external or internal contaminant.



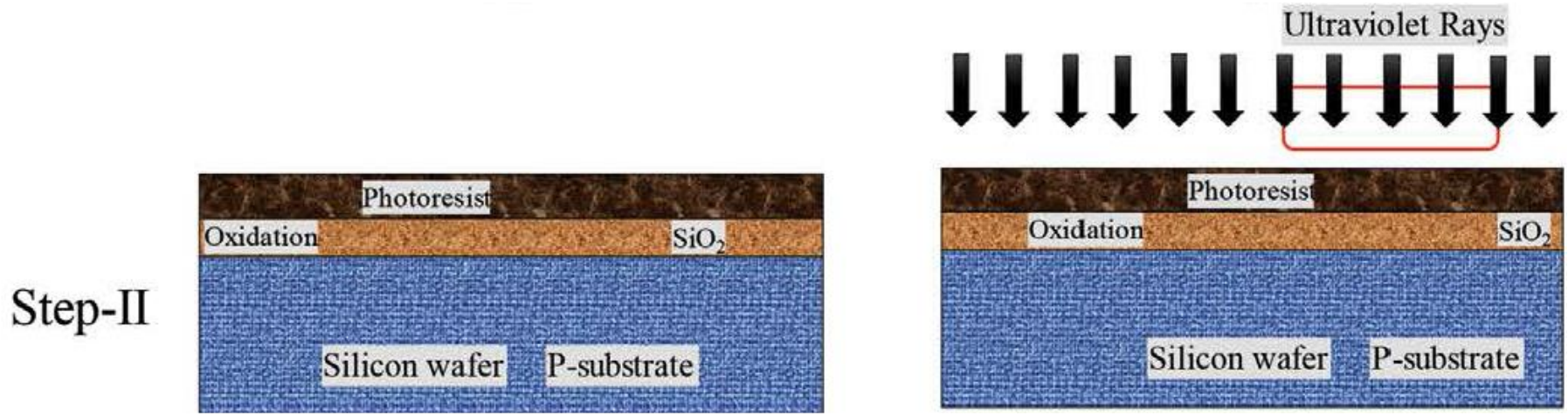
(a)



(b)

Step-II: In order to perform the photolithography process, the layer of SiO_2 is coated uniformly with photoresist material with an appropriate thickness based on the technology requirements. Further, the desired pattern is produced using a stencil that is used for masking on top of the resist material.

Furthermore, the ultraviolet (UV) rays are passed through photoresist material and the masked area gets polymerized.

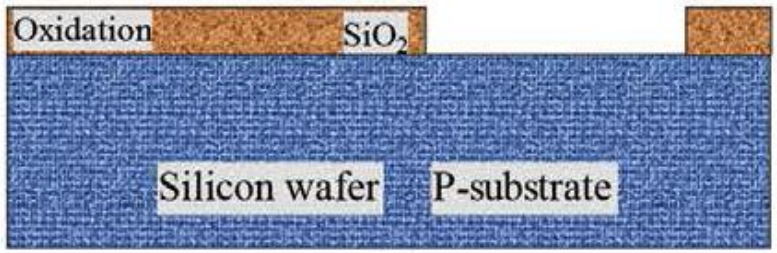


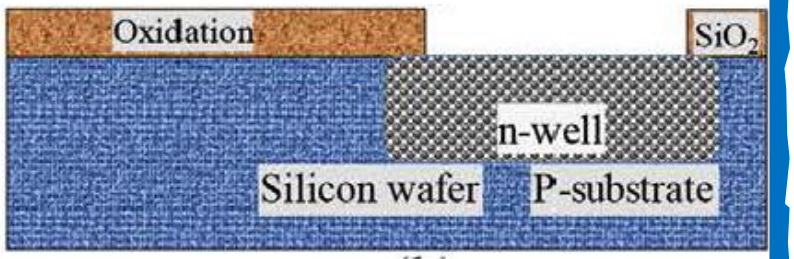
Step-III: The unexposed region is eliminated using the chemical such as trichloroethylene. Later, the silicon wafer is dropped inside the etchant solution of hydrofuoric acid that is used to eliminate only the oxide from the selected area and the area covered by photoresist will not have any effect.

Step-III

Step-IV: After the completion of the etching process, the whole photoresist mask is removed using the H_2SO_4 chemical solvent. Further, the diffusion of *N*-type impurities is performed in order to form *N*-well on top of the *P*-type silicon substrate.

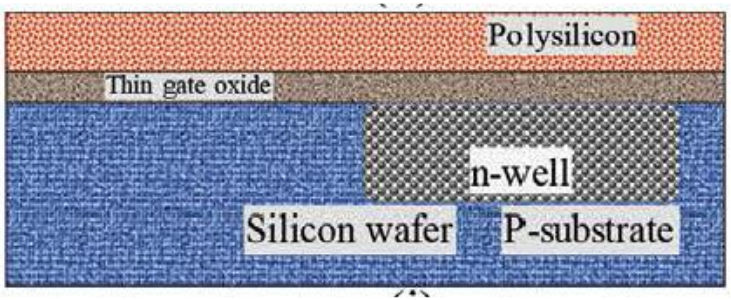
Step-IV





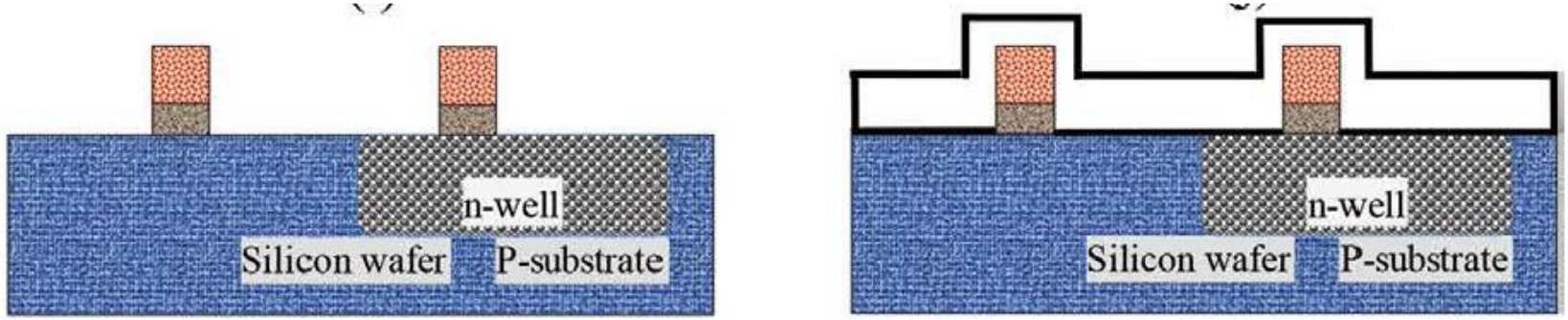
Step-V: In the later stage of the formation of *N*-well, this step is further carried out by eliminating the silicon dioxide layer from the top of the silicon wafer using the hydrofuoric acid. Later, a thin layer of gate oxide is coated on a silicon wafer in order to perform the self-aligned gate process.

Step-V



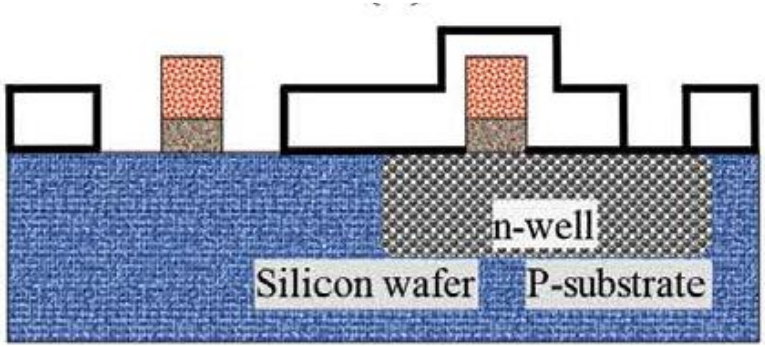
Step-VI: The whole portion of the polysilicon is eliminated except the two regions where the formation of the gate region is required for the NMOS and PMOS transistors. This process is followed by an oxidation process that is used as a protective layer before executing the diffusion and metallization.

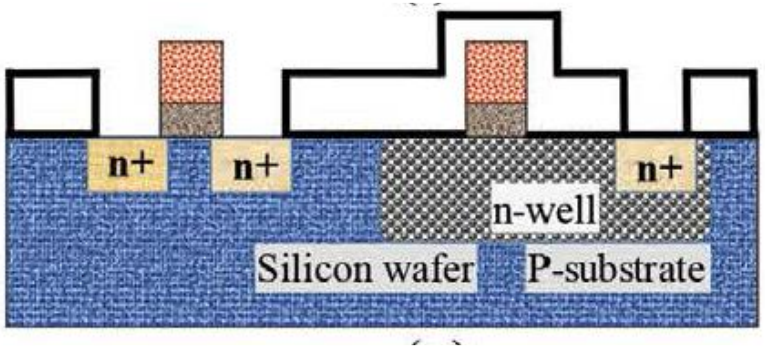
Step-VI



Step-VII: Further, few small regions are formed by masking the silicon wafer in order to diffuse *N*-type impurities on *P*-substrate and *N*-well. Moreover, the three *N*+ regions are formed using a diffusion process to create the NMOS terminal.

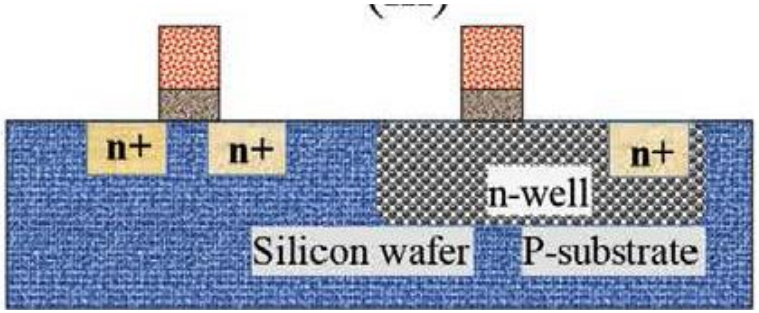
Step-VII

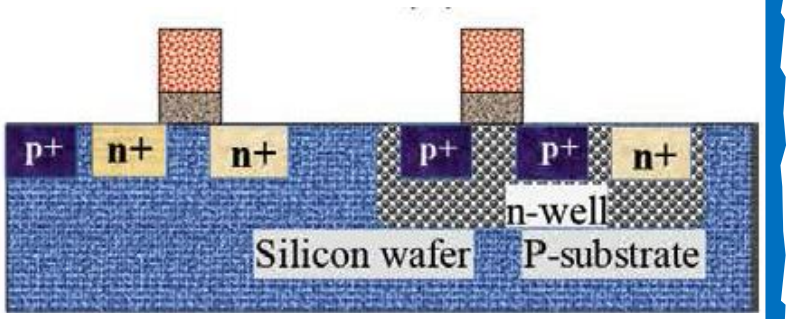




Step-VIII: This step is used to remove the oxide layer followed by diffusion of three *P*+ regions in order to form the PMOS terminal similar to *N*-type diffusion as observed in step-VII.

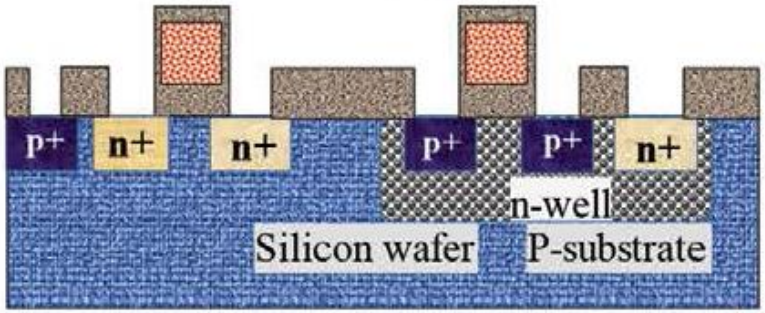
Step-VIII

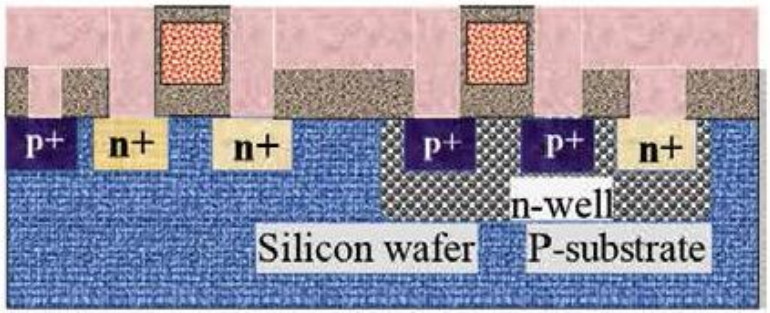




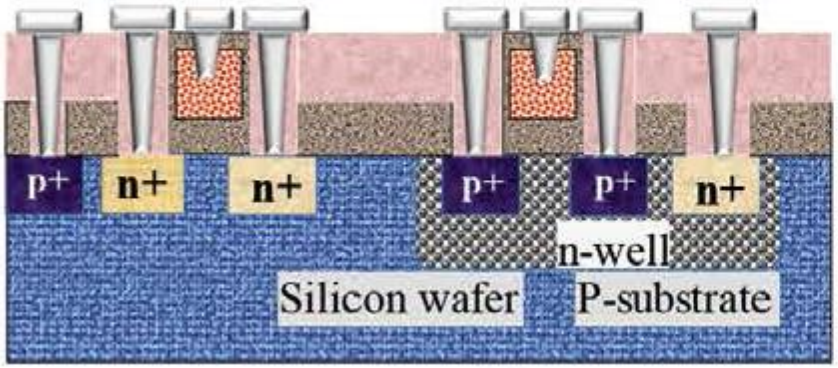
Step-IX: Before forming the metal terminal using the metallization process, a layer of thick oxide needs to lay out to form a protected region where no terminal needs to form. After this process, a suitable material is used to form the metal terminal in the whole wafer area in order to develop the interconnection.

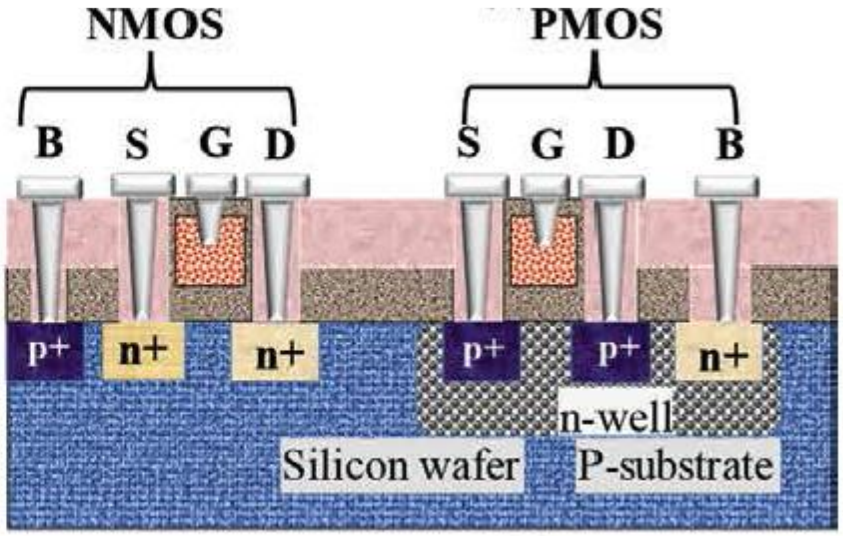
Step-IX



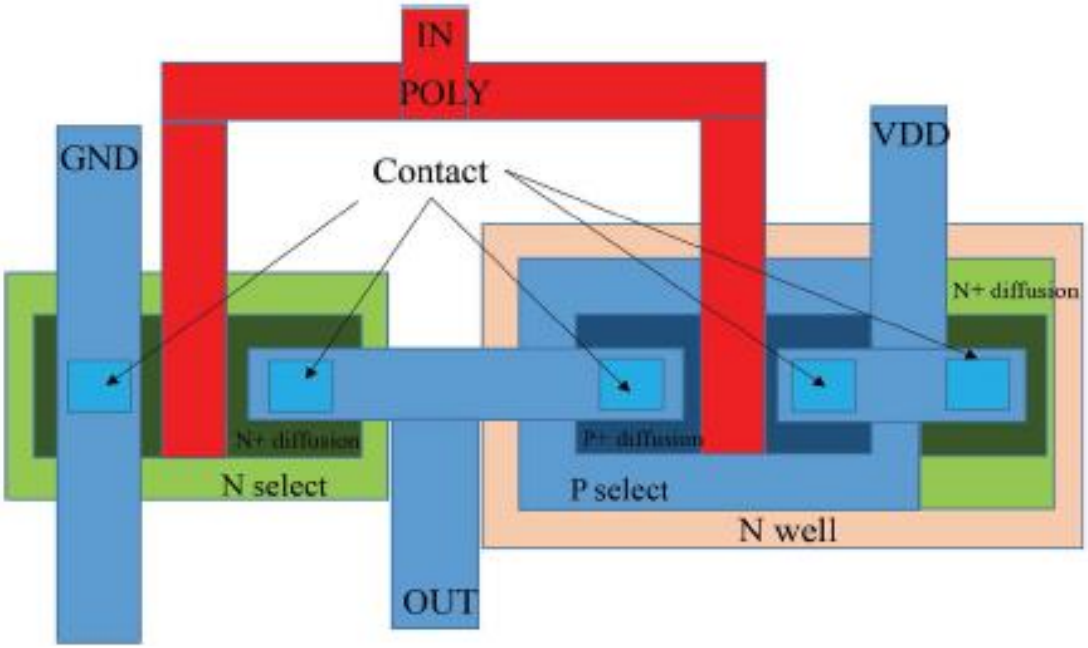


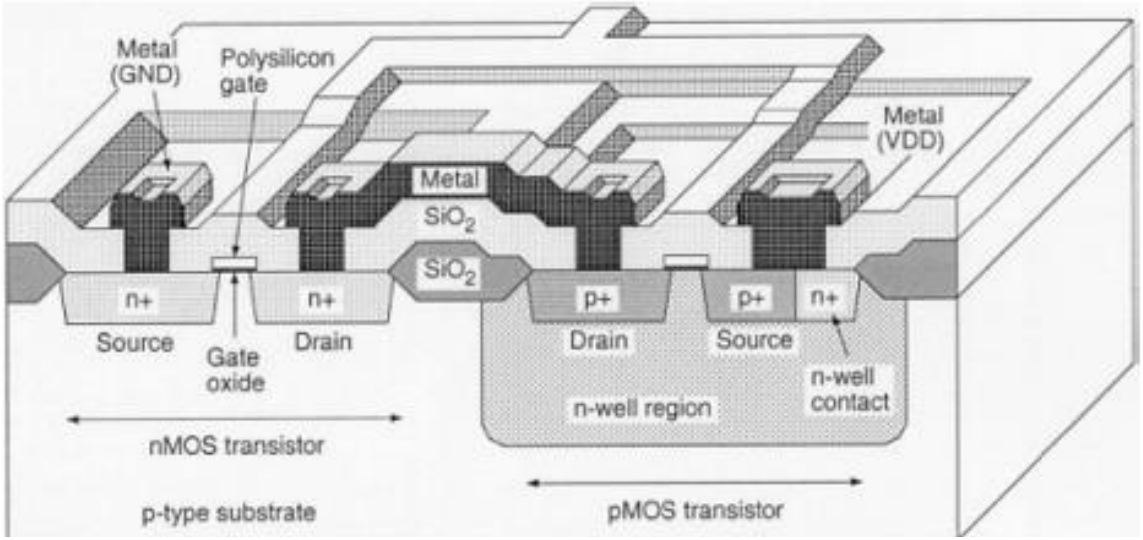
Step X: In the final stage, the etching process is used to remove any excess metal present on the top of the silicon wafer and the metal terminals are formed in the gap after the etch process. Thereafter, all the terminals of NMOS and PMOS transistors are assigned with a unique name for the final product.





Final product

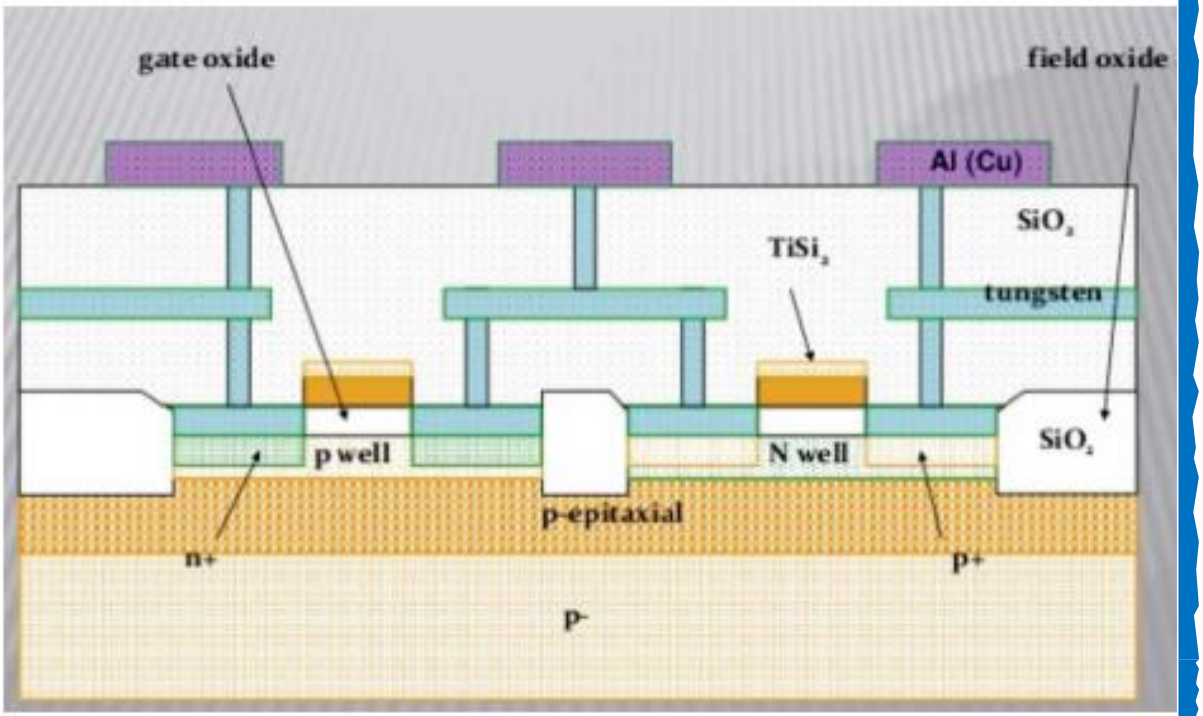




TWIN-TUB PROCESS

In the twin-tub process, the separate optimization can be performed for NMOS and PMOS transistor parameters such as body effect, channel transconductance, and threshold voltage. In this technology, the primary substrate is considered either *N*-type or *P*-type substrate with a lightly doped epitaxial layer on top of it. The formation of *N*-well and *P*-well is the first step toward the beginning of the twin-tub process.

An appropriate dopant concentration is applied in order to produce the desired characteristics of the device. The complete dual-well or twin-tub process is demonstrated



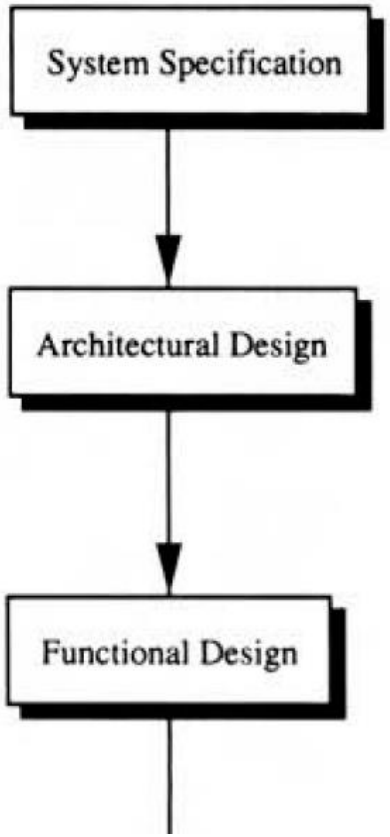
VLSI Design Cycle

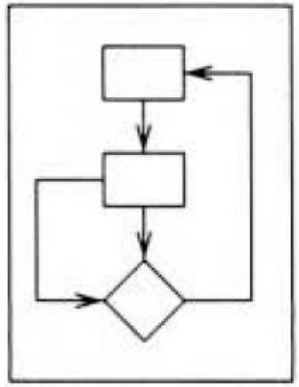
The VLSI design cycle starts with a formal specification of a VLSI chip, follows a series of steps, and eventually produces a packaged chip. A typical design cycle may be represented by the flow chart shown.

The emphasis is on the physical design step of the VLSI design cycle. A brief outline all the steps of the VLSI design cycle.

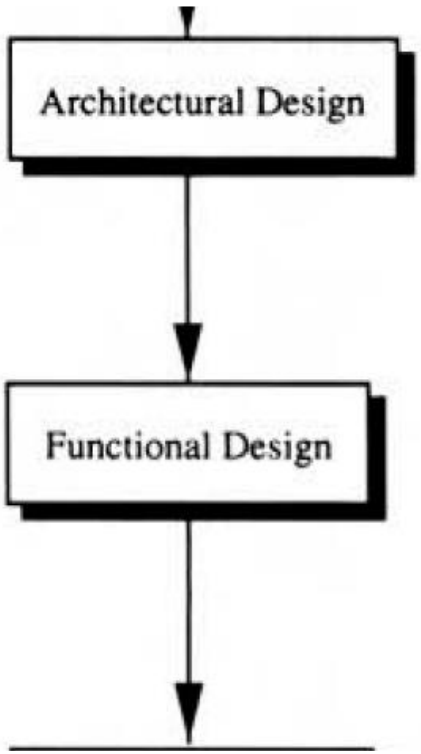
System Specification: The first step of any design process is to lay down the specifications of the system. System specification is a high level representation of the system.

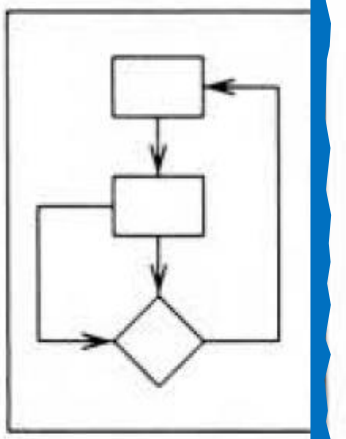
The factors to be considered in this process include: performance, functionality, and physical dimensions (size of the die (chip)). The fabrication technology and design techniques are also considered. The specification of a system is a compromise between market requirements, technology and economic viability. The end results are specifications for the size, speed, power, and functionality of the VLSI system.



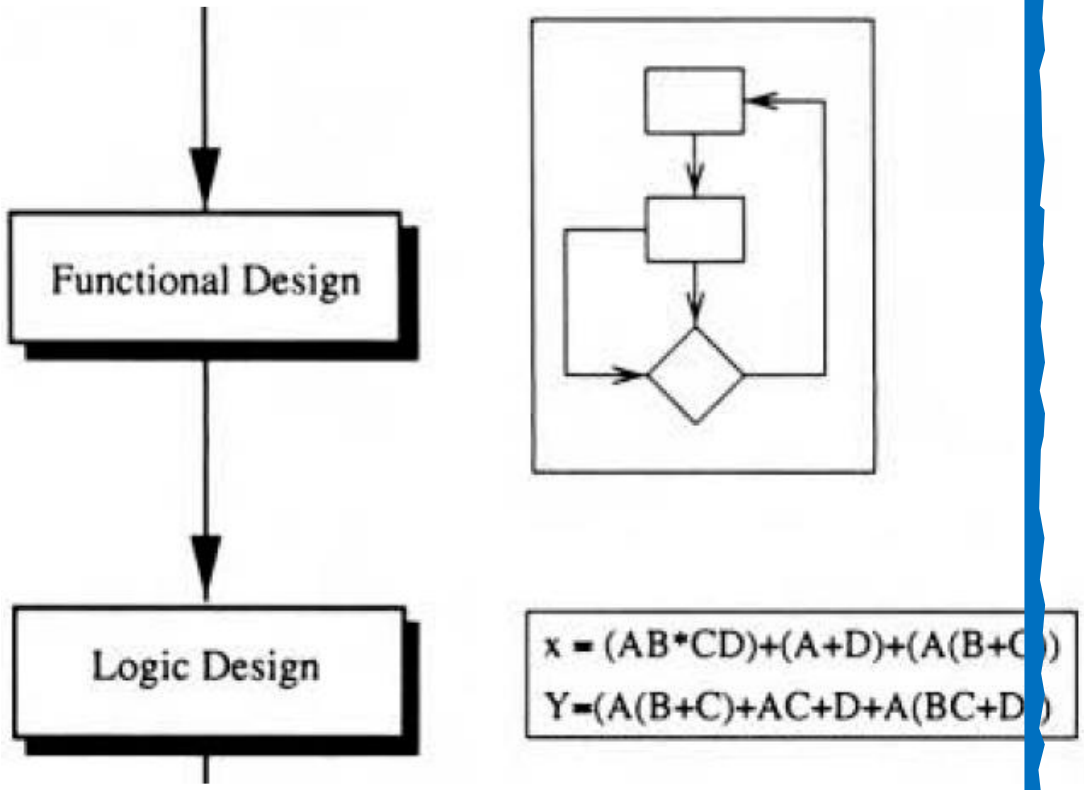


Architectural Design: The basic architecture of the system is designed in this step. This includes, such decisions as RISC (Reduced Instruction Set Computer) versus CISC (Complex Instruction Set Computer), number of ALUs, Floating Point units, number and structure of pipelines, and size of caches among others. The outcome of architectural design is a Micro-Architectural Specification (MAS). While MAS is a textual (English like) description, architects can accurately predict the performance,
power and die size of the design
based on such a description.





Behavioral or Functional Design: In this step, main functional units of the system are identified. This also identifies the interconnect requirements between the units. The area, power, and other parameters of each unit are estimated. The behavioral aspects of the system are considered without implementation specific information. For example, it may specify that a multiplication is required, but exactly in which mode such multiplication may be executed is not specified.



Logic Design: In this step the control flow, word widths, register allocation, arithmetic operations, and logic operations of the design that represent the functional design are derived and tested.

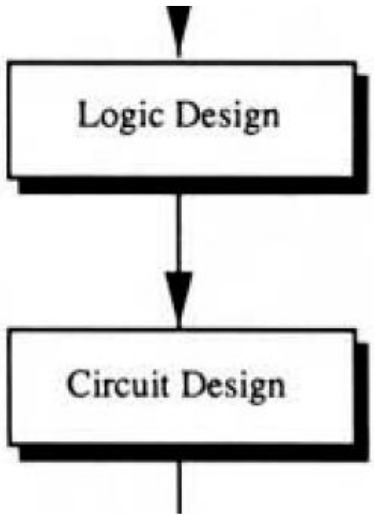
This description is called Register Transfer Level (RTL) description.

RTL is expressed in a Hardware Description Language (HDL), such as VHDL or Verilog.

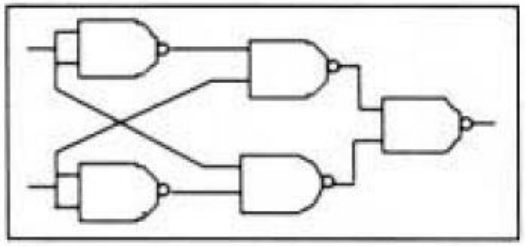
This description can be used in simulation and verification. This description consists of Boolean expressions and timing information.

The Boolean expressions are minimized to achieve the smallest logic design which conforms to the functional design.

This logic design of the system is simulated and tested to verify its correctness.

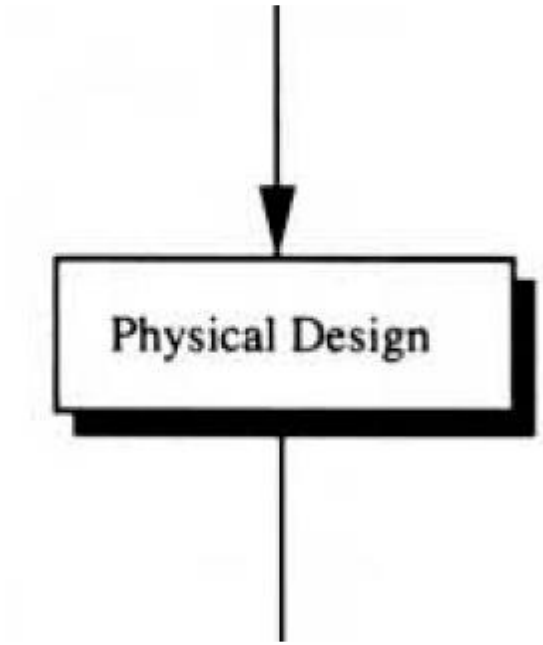


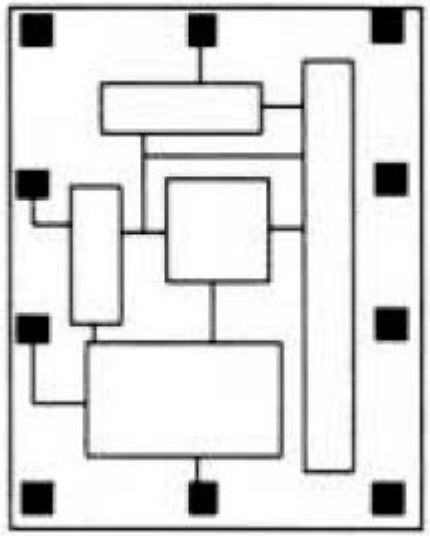
x = (AB*CD)+(A+D)+(A(B+C))
Y=(A(B+C)+AC+D+A(BC+D))



Circuit Design: The purpose of circuit design is to develop a circuit representation based on the logic design. The Boolean expressions are converted into a circuit representation by taking into consideration the speed and power requirements of the original design.

Circuit Simulation is used to verify the correctness and timing of each component. The circuit design is usually expressed in a detailed circuit diagram.

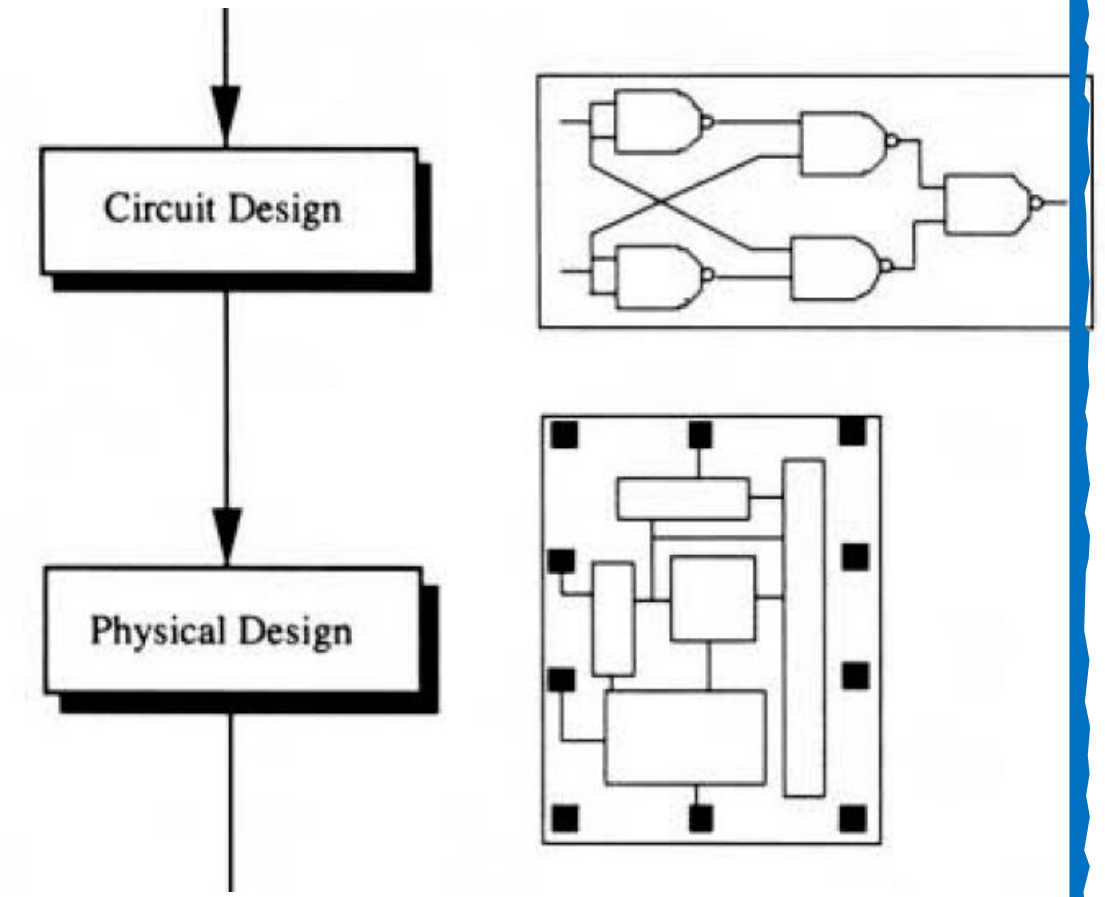




This representation is also called a *netlist.*

Tools used to manually enter such description are called *schematic capture tools.*

Physical Design: In this step the circuit representation (or netlist) is converted into a geometric representation. As stated earlier, this geometric representation of a circuit is called a *layout*. Layout is created by converting each logic component (cells, macros, gates, transistors) into a geometric representation (specific shapes in multiple layers), which perform the intended logic function of the corresponding component.

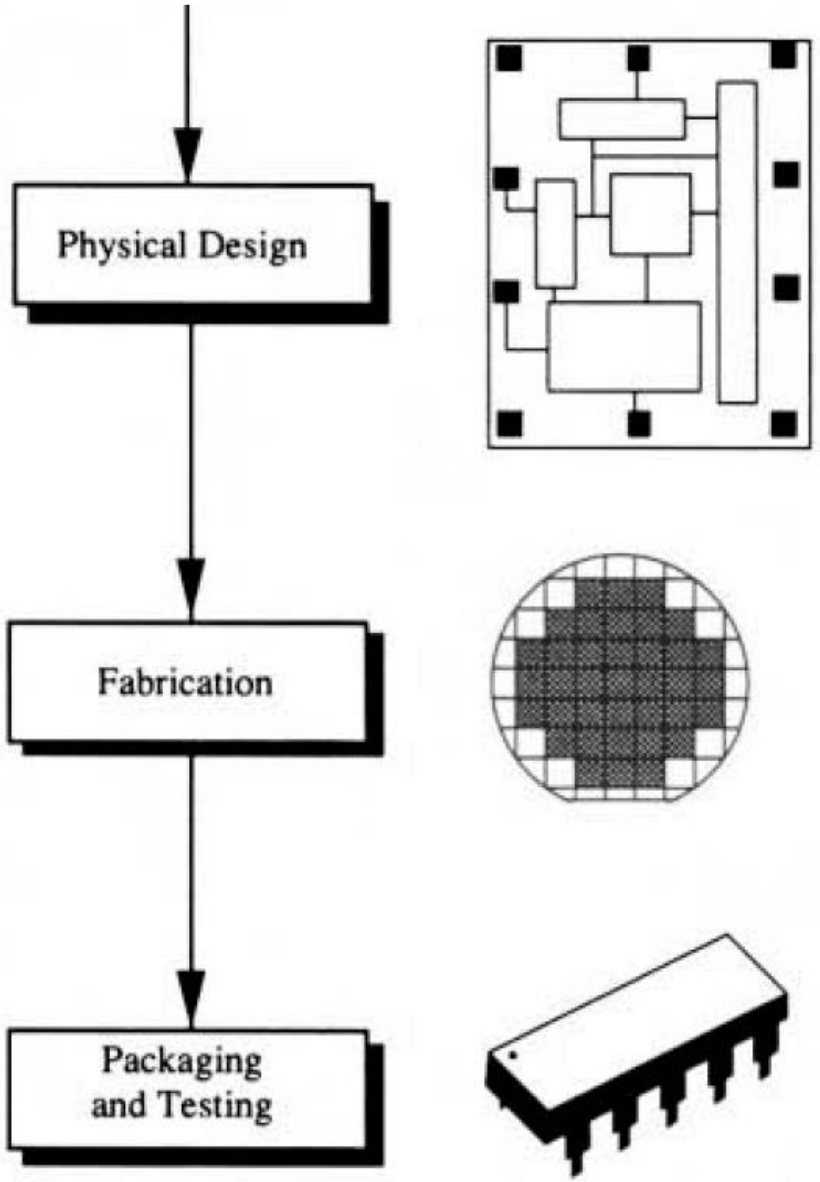


Fabrication: After layout and verification, the design is ready for fabrication.

Since layout data is typically sent to fabrication on a tape, the event of release of data is called *Tape Out*. Layout data is converted (or fractured) into photo-lithographic masks, one for each layer. Masks identify spaces on the wafer, where certain materials need to be deposited, diffused or even removed. Silicon crystals are grown and sliced to produce wafers. Extremely small dimensions of VLSI devices require that the wafers be polished to near perfection. The fabrication process consists of several steps involving deposition, and diffusion of various materials on the wafer. During each step one mask is used.

Packaging, Testing and Debugging:

Finally, the wafer is fabricated and diced into individual chips in a fabrication facility. Each chip is then packaged and tested to ensure that it meets all the design specifications and that it functions properly. Chips used in Printed Circuit Boards (PCBs) are packaged in Dual In-line Package (DIP), Pin Grid Array (PGA), Ball Grid Array (BGA)



IDEAL INTERCONNECT

- ✓ In the early stage of the VLSI technology era, the high-speed operation and large power dissipation were not major concerns.
- ✓ However, the performance was dominated by gate and active device delay instead of interconnect delay.
- ✓ For several decades, the evaluation of circuit modelling has developed to understand the physical behaviour of interconnect in terms of speed and technology scaling.
- ✓ Therefore, circuit modelling is categorized as an ideal interconnect
- ✓ Interconnect are used to understand the performance in terms of propagation delay, crosstalk effect, and power dissipation.
- ✓ Here, we begin with the calculation of resistance, inductance, and capacitance associated with interconnect by assuming a rectangular geometrical structure.

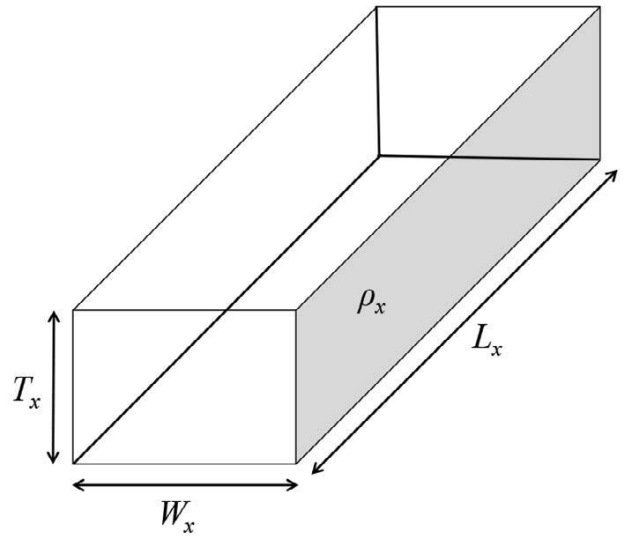
INTERCONNECT RESISTANCE ESTIMATION

Interconnect used in VLSI has different resistance based on the resistivity of metal used for interconnect applications at different technology nodes. Scaling of technology has a major impact on resistance value due to a reduced width and thickness. The basic model of interconnect resistance is shown

$$R_x = \frac{\rho_x L_x}{A_x} = \frac{\rho_x L_x}{T_x W_x}$$

$$= \frac{L_x}{\sigma_x T_x W_x} = \frac{L_x}{(e\mu n) T_x W_x}$$

$$R_x = R_{sqr} \frac{L_x}{W_x}$$



where ρ_x is the resistivity of conducting wire that depends on the charge of electron e, mobility μ, and carrier concentration n, referred to $\sigma_x = e\mu n$. The number of segmented squares of the conducting metal wire is defned as the ratio of length and width

Example

Calculate the sheet resistances of Copper (Cu) having a metal 1 wire thickness of 0.48 μm and metal 5 wire having a thickness of 1.6 μm.

Take

resistivity of Cu ≈ 1.7 μΩ cm

$$R_{sqr} = \frac{\rho_x}{T_x} = \frac{1.7\ \mu\Omega\ \text{cm}}{0.48\ \mu\text{m}} = 35.42\ \text{m}\Omega\ /\square$$

INTERCONNECT INDUCTANCE ESTIMATION

At high operating frequency and lower technology node, the accurate estimation of inductance is essential in order to check the overall system performance. The inductance appears when the current flows through the conducting wire that produces electromagnetic flux around and stored the energy.

According to Faraday's law, the voltage induced across the inductive element can be observed, when there is a change in the flow of current on conducting wire and can be expressed as

$$\Delta V = L\frac{di}{dt}$$

In addition to that, the inductance of the wire that depends on the width, thickness, and length can be calculated as

$$L = \frac{\mu_0}{2\pi}\left[L_x \ln\left(\frac{2l}{w+t}\right) + \frac{L_x}{2} + 0.2235(w+t) \right]$$

Example

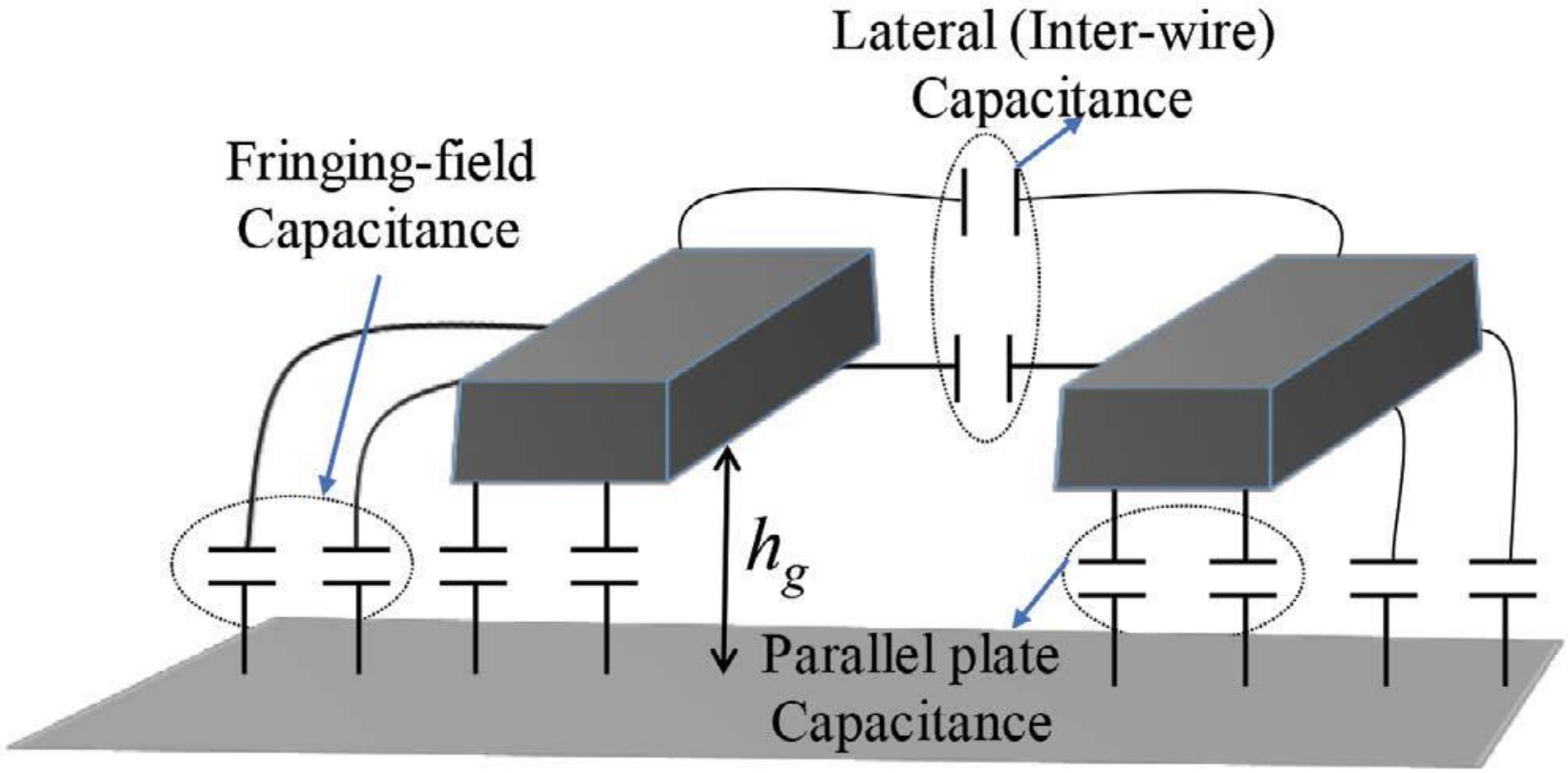
At 180 nm technology, calculate the overall inductance of a 1000 µm metal 1 wire having width and thickness of 0.6 µm and 1.32 µm, respectively. Consider the inductance of a metal 1 wire above the feld oxide of thickness 0.53 µm.

$$L = \frac{4\pi \times 10^{-13}}{2\pi}\left[1000 \ln\left(\frac{2 \times 1000}{0.6 + 1.32}\right) + \frac{1000}{2} + 0.2235(0.6 + 1.32)\right]$$

$$L = 2 \times 10^{-13}\left[6948.577 + 500.89502\right] = 1.4899 \text{ pH}$$

INTERCONNECT CAPACITANCE ESTIMATION

The interconnect capacitance primarily depends on its distance to the substrate, between the wires, distance from the surrounding metal wire, and the shape of interconnect.



Parallel Plate Capacitor

The formation of capacitance between two parallel plates.

Considering the overall area of metal plate A and distance between plates as T_x, the parallel plate capacitance can be obtained as

$$C_{p2p} = \frac{\varepsilon_{r,die}}{T_x} A = \frac{\varepsilon_{r,die}}{T_x} W_x L_x$$

$$C_{p2p} = \frac{\varepsilon_r \times \varepsilon_0}{T_x} W_x L_x$$

Fringing Capacitance

The capacitance between sidewall and substrate occurs which is known as a fringing capacitance.

In the deep submicron technology, the impact of fringing capacitance cannot be ignored.

The fringing capacitance of a wire can be obtained as

$$C_t = C_{p2p} + C_{fringe}$$

Lateral Capacitance

Apart from parallel and fringing capacitance, the lateral capacitance also has a major impact on interconnect and mainly formed due to two-wire placed closely to each other on the same plane. The lateral capacitance in *p.u.l.* for the closely spaced wires can be calculated as

$$C_{lateral} = \frac{\varepsilon_{r,die} T_x}{s}$$

Example

Consider the wire made of metal 1 placed above the dielectric constant with a thickness of 0.62 µm. Calculate the total capacitance in *p.u.l.* and total capacitance for 100 µm length having a width of 0.6 µm and thickness of 0.7 µm, respectively.

Example 2

Using 180 nm technology, assuming the following parameter based on the ITRS benchmark as width = 0.6 µm, thickness Tx = 0.7 µm, dielectric thickness hg = 0.62 µm, and spacing between the wire s = 0.6 µm.

a. Calculate parallel plate, lateral and fringing capacitance.

b. Calculate middle wire capacitance when it is closely packed.

c. Calculate middle wire capacitance when it is separated with wider space.

ASSIGNMENT

For CMOS inverter terminated with a load capacitance of 6 fF and driven by a supply voltage of 2.5 V.

a. Calculate the amount of energy needed to charge and discharge of load capacitance.

b. Assume that the CMOS inverter is switched at the maximum possible rate of 2*tp*, where t_p = 32.5 psec, then find the dynamic power dissipation of the circuit.

Due 3rd Nov 2022 (1:30npm)

NEXT LECTURE

VLSI design

CMOS