Department of Electronics & Communication Engineering

QUESTION BANK

CEC345 - OPTICAL COMMUNICATION & NETWORKS

V SEMESTER
Regulation – 2021

CEC345 OPTICAL COMMUNICATION & NETWORKS L T P C 3 0 0 3

COURSE OBJECTIVES:

- To Study About The Various Optical Fiber Modes, Configuration Of Optical Fibers
- To Study Transmission Characteristics Of Optical Fibers.
- To Learn About The Various Optical Sources, Detectors And Transmission Techniques.
- To Explore Various Idea About Optical Fiber Measurements & Various Coupling Techniques.
- To Enrich The Knowledge About Optical Communication Systems And Networks.

UNIT I INTRODUCTION TO OPTICAL FIBER COMMUNICATION 9

Introduction - The General Systems - Advantages of Optical Fiber Communication- **Ray Theory Transmission** : Total Internal Reflection, Acceptance Angle, Numerical Aperture, Skew Rays - **Electromagnetic Mode Theory for Optical Propagation**: Modes in a Planar Guide, Phase and group velocity - **Cylindrical Fiber**: Step index fibers, Graded index fibers - **Single mode fibers**: Cutoff wavelength.

UNIT II TRANSMISSION CHARACTERISTICS OF OPTICAL FIBERS 9

Attenuation - **Material absorption losses in silica glass fibers:** Intrinsic absorption, Extrinsic absorption - **Linear scattering losses**: Rayleigh Scattering, Mie Scattering -**Nonlinear scattering losses**: Stimulated Brillouin Scattering, Stimulated Raman Scattering – Fiber Bend Loss – Dispersion- **Chromatic dispersion**: Material dispersion, Waveguide dispersion- **Intermodal dispersion** : Multimode step index fiber, Multimode graded index fiber.

UNIT III OPTICAL SOURCES AND OPTICAL DETECTORS 9

The laser : Introduction - **Basic concepts**: Absorption and emission of radiation, Population inversion , Optical feedback and laser oscillation, Threshold condition for laser oscillation- **Optical emission from semiconductors**: The PN junction, Spontaneous emission, Carrier recombination, Stimulated emission and lasing, Hetero junctions- **LED**: Introduction- Power and Efficiency - **LED structures**: Planar LED, Dome LED, Surface emitter LED, Edge emitter LED- LED Characteristics. **Optical Detectors:**Introduction ,Optical Detection Principles, Quantum Efficiency, Resposivity, P-N Photodiode ,P-I-N Photo Diode and Avalanche Photodiode.

UNIT IV OPTICAL FIBER MEASUREMENTS 9

Introduction- Total Fiber Attenuation Measurement, Fiber Dispersion Measurements In Time Domain and Frequency Domain, Fiber Cut off Wavelength Measurements, Numerical Aperture Measurements. Fiber Diameter Measurements,.Reflectance And Optical Return Loss, Field Measurements

UNIT V OPTICAL NETWORKS 9

Introduction- **Optical Network Concepts**: Optical Networking Terminology, Optical Network Node And Switching Elements, Wavelength Division Multiplexed Networks, Public Telecommunications Network Overview- **Optical Network Transmission Modes**, **Layers And Protocols**: Synchronous Networks, Asynchronous Transfer Mode, Open System Interconnection Reference Model, Optical Transport Network, Internet Protocol- **Wavelength Routing Networks**: Routing And Wavelength Assignment- **Optical Switching Networks**: Optical Circuit Switched Networks, Optical Packet Switched Networks, Multiprotocol Label Switching, Optical Burst Switching Networks- **Optical Network Deployment** : Long Haul Networks, Metropolitan area networks, Access networks, Local Area Networks- **Optical Ethernet**: Network protection, restoration and survivability

TOTAL: 45 PERIODS

COURSE OUTCOMES: At the end of the course, the student will be able to understand the

CO1-Realize Basic Elements In Optical Fibers, Different Modes And Configurations.
CO2-Analyze Transmission Characteristics Associated With Dispersion And Polarization Techniques.
CO3-Design Optical Sources And Detectors With Their Use In Optical Communication System.
CO4-Construct Fiber Optic Receiver Systems, Measurements And Techniques.
CO5-Design Optical Communication Systems And Its Networks

TEXT BOOKS:

1. John M.Senior, “Optical Fiber Communication”, Pearson Education, Fouth Edition.2010.

REFERENCES:

1. Gred Keiser,"Optical Fiber Communication”, McGraw Hill Education (India) Private Limited. Fifth Edition, Reprint 2013.
2. Govind P. Agrawal, “Fiber-Optic Communication Systems”, 3rd Edition, John Wiley & Sons, 2004.
3. J.Gower, “Optical Communication System”, Prentice Hall Of India, 2001
4. Rajiv Ramaswami, “Optical Networks “ , Second Edition, Elsevier , 2004.
5. P Chakrabarti, "Optical Fiber Communication”, McGraw Hill Education (India)Private Limited**,** 2016

UNIT -1 INTRODUCTION
PART A

1. **why do we prefer step index single mode fiber for long distance communication?[APR/MAY 2019]**
   Step index single mode fiber has (1) low attenuation due to smaller core diameter, (2) Higher bandwidth, (3)Low dispersion.
2. **What is the necessity of cladding for an optical fiber? [APR/MAY 2019]**
   a) To provide proper light guidance inside the core b) To avoid leakage of light from the fiber c) To avoid mechanical strength for the fiber d) To protect the core from scratches and other mechanical damages
3. **Distinguish between meridional rays from skew Rays. [NOV/DEC 2018]**
   A **skew ray** is a **ray** that travels in a non planar zig zag path and never crosses the axis of an optical fibre! A **meridional ray** is a **ray** that passes through the axis of an optical fiber.
4. **Manufacturing engineer wants to to make an optical fibre that has core intex of 1.40 and cladding intex of 1.47 8 what should be the core size for single[NOV/DEC 2018]**

$$V = \frac{2\pi a}{\lambda} NA$$

5. A silica optical fiber with a large core diameter has a core refractive index of 1.5 and a cladding refractive index of 1.47.Determine the acceptance angle in air for the fiber. [April 2017, APR 2018]

Given data:

$n_1 = 1.5$

$n_2 = 1.47$

Formula:

$$\theta_a = \sin^{-1}\sqrt{n_1^2 - n_2^2}$$

Solution:

$$\theta_a = \sin^{-1}\sqrt{1.5^2 - 1.47^2}$$

$$\theta_a = 17.36^\circ$$

6. Write short notes on ray optics theory.

Laws governing the nature of light are called as ray optics. These laws are stated as:

1. Light rays in homogenous media travel in straight lines.
2. Laws of reflection: Angle of reflection θ_r equals angle of incidence θ_i
3. Snell's Law: The angle of refraction θ_t is related to angle of incidence θ_i by

$$n_1 \sin\theta_i = n_2 \sin\theta_t$$

7. What are the advantages and disadvantages of the ray optics? [April 2017]

The advantages of ray optics are:

a) Ray optics is used to develop some of the fundamental parameters like acceptance angle, numerical aperture that are associated with optical fiber transmission.

b) It provides an excellent approximation, when the wavelength is very small compared with the size of structures, with which the light interacts.

The disadvantages of the ray optics are:

a) Ray optics fails to account for optical effects such as diffraction and interference.

8. Define Phase and group velocity. (Nov-Dec 2015)

The group velocity of a wave is the velocity with which the overall shape of the waves amplitude known as modulation or envelope of wave propagates through space.

$$v_g = \frac{d\omega}{d\beta}$$

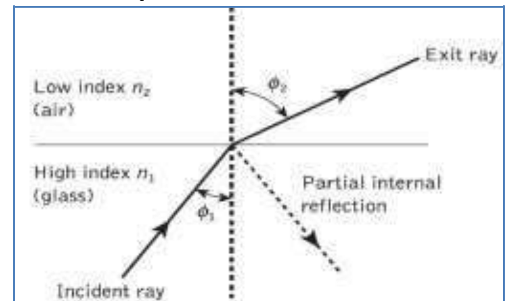
The Phase velocity of a wave is the rate at which the phase of the wave propagates in space. This is the velocity at which the phase of any one frequency component of wave travel.

$$v_p = \frac{\omega}{\beta}$$

9. **Define – Critical Angle [DEC 2016]**

The critical angle is defined as the minimum angle of incidence ($\varphi 1$) at which the ray strikes the interface of the two medium and causes an angle of refraction ($\varphi 2$) equal to 90°.

10. Assume that there is a glass rod of refractive index 1.5, surrounded by air. Find the critical incidence angle.[MAY 2016]



Given data:

$n_1 = 1.5$

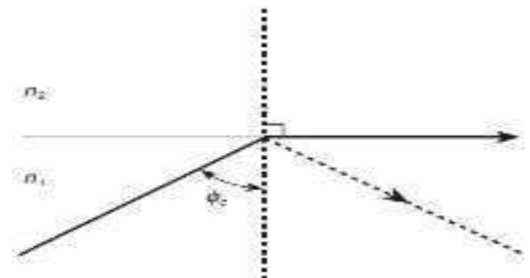
$n_2 = 1$

Formula: $\phi_c = \sin^{-1}\frac{n2}{n1}$

Solution:

$$\phi_c = \sin^{-1}\frac{1}{1.5} = 41.81°$$

11. State Snell's law.[DEC 2013, MAY 2016]



The Snell's law is an expression that describes the relationship between the angles of incidence $\theta 1$ and refraction $\theta 2$ and to the refractive indices of the dielectrics, when referring to waves passing through a boundary between two isotropic medium.

$$n1 \ \sin\theta 1 = n2 \ \sin\theta 2$$

where $n1$ is the refractive index of the core and $n2$ is the refractive index of the cladding.

12. What are the conditions for total internal reflection? [NOV 2015]

The conditions for total internal reflection are:

a) The ray should travel from denser to rarer medium i.e. from core to clad region of the optical fiber.

b) The angle of incidence in the denser medium should be greater than the critical angle of that medium.

13. Calculate the critical angle of incidence between two substances with different refractive indices, where n1 =1.5 and n2 =1.46. (Apr-May 2015)

$n_1 = 1.5$

$n_2 = 1.46$

$\emptyset = \sin^{-1}\frac{n_2}{n_1}$

$= \sin^{-1}\frac{1.46}{1.5}$

$\emptyset = 76.74°$

14. List any two advantages of single mode fibers.(Nov-Dec 2014)

Single mode fiber has only one ray passes through fiber. Ray passes along the axis-axial ray. Core diameter is small (typically 10 - 12 μm). Intermodal dispersion is not present. Coupling efficiency is less.

15. **Define - Numerical Aperture [NOV2014]**

Numerical Aperture (NA) of the fiber is the light collecting efficiency of the fiber and is the measure of the amount of

light rays that can be accepted by the fiber. It is equal to the sine of acceptance angle θ_a

$$NA = \sin \theta_a = (n_1^2 - n_2^2)^{1/2}$$

where n_1 and n_2 are the refractive indices of core and cladding respectively.

16. For n1 = 1.55 and n2 =1.52 , Calculate the critical angle and numerical aperture.(May-June 2013)

Critical angle $\emptyset_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) = \sin^{-1}\left(\frac{1.52}{1.55}\right) = 78.7°$ Numerical aperture $NA = \sqrt{n_1^2 - n_2^2} = 0.3$

17. Define – Relative Refractive Index Difference

The relative refractive index difference is the ratio of the refractive index difference between core and cladding and refractive index of core.

Where, $$\Delta = \frac{n_1 - n_2}{2n_1}$$

Δ is the relative refractive index
n_1 is the numerical aperture of the core
n_2 is the numerical aperture of the cladding

18. What is the energy of the single photon of the light whose λ = 1550nm in eV? (N/D2011)

The energy of the single photon of the light is given by the equation

$E = h \times f$

Sub $f = \frac{c}{\lambda}$ in the above equation

$E = h \times \frac{c}{\lambda}$

Given data:

$h = 6.625 \times 10^{-34}$

$c = 3 \times 10^8$ m/sec

$\lambda = 1550 \times 10^{-9}$ m

$$E = \frac{6.625 \times 10^{-34} \times 3 \times 10^8}{1550 \times 10^{-9}}$$

$= \mathbf{0.0128 \times 10^{-17}\ J}$

19. step index fiber has the normalized frequency of 26.6 at 1300nm. If the core radius is 25µm, find the numerical aperture.

Given data: V= 26.6 , $\lambda = 1300 \times 10^{-9}$m a = 25×10^{-6}m
Formula:
Normalized frequency V is given by V= $2\pi a$ (NA) / λ
NA = $\lambda V / 2\pi a$
Solution:
Numerical Aperture = $\lambda V / 2\pi a$

$$NA = \frac{1300 \times 10^{-9} \times 26.6}{2 \times 3.14 \times 25 \times 10^{-6}}$$

NA = **0.22**

20. What is meant by refractive index of the material?

The refractive index (or index of refraction) 'n' is defined as the ratio of the velocity of light in vacuum to the

velocity of light in the medium.

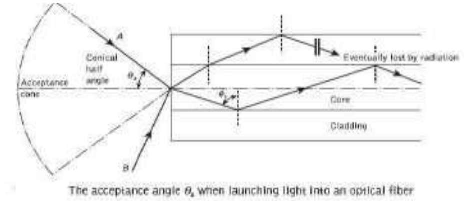
$$n = \frac{c}{v}$$

c = speed of light in free space

v = speed of light in a given material

21. Define - Acceptance angle

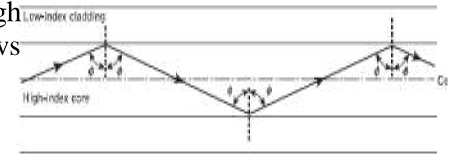
The maximum angle 'θ_a' with which a ray of light can enter through the fiber and still be totally internally reflected is called acceptance angle of the fiber.



The acceptance angle θ_a when launching light into an optical fiber

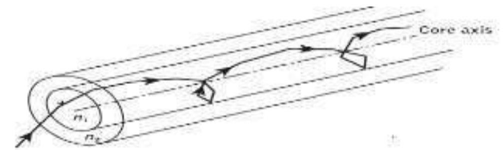
22. What are meridional rays?

Meridional rays are the rays following zig- zag path when they travel through fiber and for every reflection it will cross the fiber axis. The figure below shows the meridional rays.



23. What are skew rays?

Skew rays are the rays following the helical path around the fiber axis when they travel through the fiber and they would not cross the fiber axis at any time. The figure below shows the propagation of skew rays.



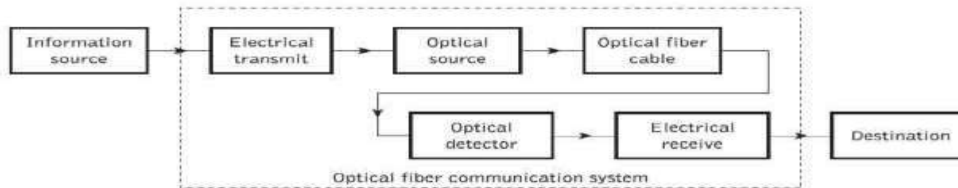
24. Write the acceptance angle condition for the skew rays.

The acceptance conditions for skew rays is given by the equation

$$\theta_{as} = \sin^{-1} \frac{NA}{\cos\gamma}$$

where NA is the numerical aperture and γ is the angle between the projection of the ray in two dimensions and the radius of the fiber core at the point of reflection.

25. Draw the block diagram of an optical communication system.



The block diagram of an optical communication system is represented above:

26. The relative refractive index difference (Δ) for an optical fiber is 1%. Determine the critical angle at the core cladding interface if the core refractive index is 1.46.

$$\Delta = \frac{n1 - n2}{n1} \qquad (1)$$

$$\phi_c = \sin^{-1} \frac{n2}{n1} \qquad (2)$$

Find n_2 from equation (1)

$$n_2 = n1(1 - \Delta) \quad (3)$$

Solution:

$$n_2 = 1.46(1 - 0.01)$$

$$= 1.4454$$

$$\phi_c = \sin^{-1} \frac{1.4454}{1.46}$$

$$\boldsymbol{\phi_c = 81.19^\circ}$$

27. Which photodiode is used for a low power optical signal and Why?

Avalanche Photo Diode (APD) is used for a low power optical signal because it has a greater sensitivity due to an inherent internal gain mechanism produced by avalanche effect.

28. What is V number of a fiber?

Normalized frequency or V number is a dimensionless parameter and represents the relationship among three design variables of the fiber i.e. core radius **a**, relative refractive index Δ and the operating wavelength **λ**. It is expressed as $V= 2\pi a\,(NA)/\lambda$.

29. What are guided modes?

Guided modes are a pattern of electric and magnetic field distributions that is repeated along the fiber at equal intervals.

30. Define – Phase Velocity

As a monochromatic light wave propagates along a waveguide in the *z* direction the points of constant phase travel at a phase velocity V_p given by

$$V_p = \frac{\omega}{\beta}$$

where ω is the angular frequency and β is the propagation constant

31. Define – Group Velocity

Group of waves with closely similar frequencies propagate so that their resultant forms packet of waves. This wave packet does not travel at the phase velocity of individual but it moves with the group velocity V_g given by

$$V_p = \frac{\omega}{\beta}$$

where ω is the angular frequency and β is the propagation constant

32. What is meant by mode coupling? What causes it?

The effect of coupling energy from one mode to another mode is known as mode coupling. The cause of mode coupling is due to waveguide perturbations such as deviations of the fiber axis from straightness, variations in the core diameter, irregularities at the core -cladding interface and refractive index variations.

33. What are the uses of optical fibers?

The uses of optical fiber are

- To transmit analog and digital information.
- To transmit the optical images.(Endoscopy Images)
- To act as a light source at the inaccessible places.
- To act as sensors for mechanical, electrical and magnetic measurements.

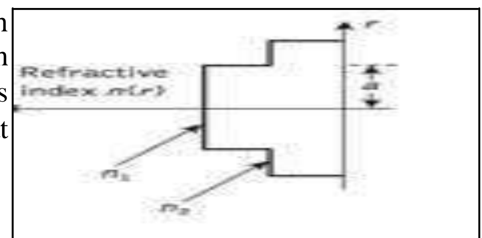
34. What is the necessity of cladding for an optical fiber?

The necessity of cladding for an optical fiber is:

- To provide proper light guidance inside the core.
- To avoid leakage of light from the fiber.
- To provide mechanical strength for the fiber.
- To protect the core from scratches and other mechanical damages

35. What is step index fiber?

Step index fiber is a cylindrical waveguide that has the central core with uniform refractive index of n_1, surrounded by outer cladding with refractive index of n_2.The refractive index of the core is constant and is larger than the refractive index of the cladding. It makes a step change at the



core cladding interface as indicated in the figure,

36. Write the refractive index expression for step index fiber.

In step index fiber, the refractive index of a core is constant and is larger than the refractive index of the cladding. The refractive index profile is defined as

n(r) = n1; r < a (core)

n2; r ≥ a (cladding)

37. What are the advantages of Graded Index Fiber?

The advantages of Graded Index Fiber are

- It exhibits **less intermodal dispersion** because the different group velocities of the modes tend to be normalized by the index grading.
- It provides **higher bandwidth**

38. Write the refractive index expression for graded index fiber.

Graded index fibers does not have a constant refractive index in the core but a gradually decreasing core index $n(r)$ with radial distance from a maximum value of $n1$ at the axis to a constant value $n2$ beyond the core radius 'a' in the cladding. This index variation may be represented as:

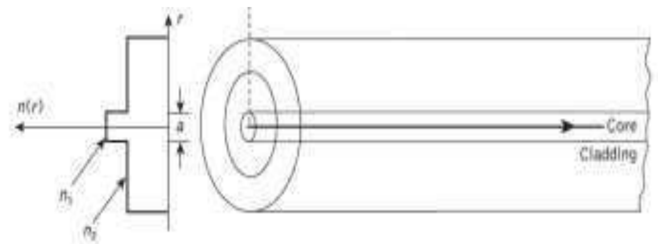
$$n(r) = \{ n1 = (1 - 2\Delta(r/a)^{\alpha})^{1/2} \quad ; \quad r < a \text{ (core)}\}$$

$$\{n1 (1 - 2\Delta)^{1/2} = n2 ; \quad r \geq a \text{ (cladding)}\}$$

Where,n1 is the refractive index of the core and n2 is the refractive index of the cladding Δ is the index difference, α is the index profile

39. Write a short note on single mode fiber.

For single-mode operation, only one mode (the fundamental LP01) can exist and it does not suffer from mode delay. The core diameter is small so that there is only one path for light ray to propagate inside the core. Typical core sizes are 2μm to 5μm. It provides larger bandwidth and less coupling efficiency. It is used for long haul transmission.



40. List out the advantages of multimode fiber over single mode fibers. (A/M2008) The advantages of multimode fiber are:[DEC 2016]

- The larger core radii of multimode fibers make it easier to launch optical power into the fiber. Connecting together of similar fibers is easy.
- Light can be launched into a multimode fiber using an LED source, whereas single-mode fibers with LASER diodes. LED's are easier to make, less expensive, less complex circuitry and have longer life times.

41. List the advantages and disadvantages of monomode fiber.

The advantages of single mode fiber are:

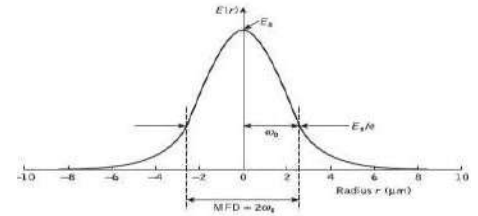
- No intermodal dispersion
- Information capacity of single mode fiber is large

The disadvantages of single mode fiber are:

- Launching of light into single mode and joining of two fibers are very difficult
- Fabrication is very difficult and so that fiber is so costly

42. Define – Mode Field Diameter

Mode-Field Diameter is an important parameter for characterizing single mode fiber properties that accounts the wavelength dependent field penetration into the fiber cladding. This can be determined from the mode field distribution of the fundamental LP_{01} mode. The MFD equals $2\omega_0$ where ω_0 is the nominal half width of the input excitation.



43. Why is step index single mode fiber preferred for long distance communication?

The step index single mode fiber is preferred for long distance communication because,

- They exhibit higher transmission bandwidth because of low fiber losses.
- They have superior transmission quality because of the absence of modal noise.
- The installation of single mode fiber is easy and will not require any fiber replacement over twenty plus years.

44. Define – Birefringence

Manufactured optical fibers have imperfections, such as asymmetrical lateral stresses, non circular cores, and variations in refractive index profiles. These imperfections break the circular symmetry of the ideal fiber and lift the degeneracy of the two modes. These modes propagate with different phase velocity and it is called as fiber birefringence. Birefringence is expressed as

$$B_f = \beta_x - \beta_y / \frac{2\pi}{\lambda}$$ where β is the propagation constant.

45. **State the reasons to opt for optical fiber communication.**

- **Broad bandwidth**: A single optical fiber can carry over 3,000,000 full-duplex voice calls or 90,000 TV channels.
- **Immunity to electromagnetic interference:** Light transmission through optical fibers is unaffected by otherelectromagnetic radiation nearby. The optical fiber is electrically non-conductive, so it does not act as an antenna to pick up electromagnetic signals. Information traveling inside the optical fiber is immune to electromagnetic interference, evenelectromagnetic pulses generated by nuclear devices.
- **Low attenuation loss over long distances:** Attenuation loss can be as low as 0.2 dB/km in optical fiber cables, allowing transmission over long distances without the need for repeaters.
- **Electrical insulator:** Optical fibers do not conduct electricity, preventing problems with ground loops and conduction oflightning. Optical fibers can be strung on poles alongside high voltage power cables.
- **Material cost and theft prevention:** Conventional cable systems use large amounts of copper. Global copper prices experienced a boom in the 2000s, and copper has been a target of metal theft.
- **Security of information passed down the cable:** Copper can be tapped with very little chance of detection.

45. Differentiate between Mono Mode Fiber and Multimode Fiber.

S. No	Mono Mode Fiber	Multi Mode Fiber
1	Only one ray passes	More than one ray passes
2	Ray passes along the axis-axial ray	MMSI – Meridional and Skew MMGI – Paraxial

3	Core diameter is small typically10 - 12µm	Core diameter is large typically 50 - 200µm
4	Intermodal dispersion is not present	Intermodal dispersion is present
5	Fabricating single mode fiber is difficult	Fabricating multimode fiber is easy
6	Coupling efficiency is less	Coupling efficiency is large
7	LED is not suitable source for single mode	LED is suitable for multimode

46. Point out the limitations of Optical Fiber Communication system?

- Optical fiber is made up of glass because of the impurities present within the fiber result in absorption leads to loss of light in the Optical fiber.
- Maximum limitation of the bandwidth of the signals can be carried by the fiber due to spreading of pulse.
- It is costly.
- Optical fiber has limited band radius (□ 10mm)
-

47. Distinguish between Step Index fiber and Graded Index fiber.

S. No	**Step Index Fiber**	**Graded Index Fiber**
1	The core has uniform refractive index but step change in core-cladding interface.	The core has high refractive index along the axis which gradually decreases towards the clad-core interface (radially decreases)
2	Axial ray – SMSI, Meridonial rays & Skew - MMSI	Paraxial rays – MMGI
3	Intermodal dispersion is present in MMSI	Intermodal dispersion is reduced in MMGI
4	Numerical Aperture is constant	Numerical Aperture is a function of radius
5	Step index profile	Graded index profile □ □profile Factor
6	No of modes, m □ $v^2 / 2$. Step index supports twice the number of modes than GI	No of modes, m □ $v^2 / 4$
7	Fabrication is easy	Fabrication is difficult

48. Compare Ray Optics with Wave Optics.

S. No	**Ray Optics**	**Wave Optics**
1	It is used to represent the direction of light propagation	It is used to analyze mode theory
2	It is used to study reflection and refraction of light	It is used to analyze diffraction and Interference of light waves

49. Define Mode.

Mode is the pattern of distribution of electric and magnetic fields

- Transfers Electric Mode TE_{02}
- Transfers Magnetic Mode TM_{02}

50. List out the ways to minimize leaky modes.

A mode remains guided as long as β satisfies the condition $n2K<\beta n1K$.

n1,n2 → Refractive index of core are cladding $K = 2\pi/\lambda$

$\beta \geq n2K$ = To prevent power leaks out of the core.

51. What are the three windows of Optical Communication?

The three wave lengths 850nm, 1300nm and 1500nm are three optical windows of optical communication system. Since only at this wavelength silica fiber loss is minimum.

PART B & C

1. **a. Explain a neat block diagram of fundamentals of optical fibre communication.[8]**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. NO.8**[DEC 2016, APR 2018]**
   b. Discuss the mode theory of circular waveguides. [Nov 2008]
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 43
2. **For multi-mode step-index fibre with glass core (n1 =1.5) and a fused quartz cladding (n2 =1.46), determine the acceptance angle θ_{in}) and numerical aperture. The source to fibre medium is air. (Apr-May 2015, NOV/DEC 2018)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 37
3. **Explain the ray propagation into and down an optical fibre cable. Also derive the expression for acceptance angle. (Apr-May 2015 , 2019)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 37
4. **Contrast the advantages and disadvantages of step-index, graded- index, single-mode propagation and multi-mode propagation. (Apr-May 2015)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 37
5. **Classify fibers and explain them. (Nov-Dec 2015)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 37
6. **Describe and derive the modes in planar guide. (Nov-Dec 2015)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 43
7. **Define the normalized frequency for an optical fiber and explain its use.(Nov-Dec 2014)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 25
8. **Explain the features of multimode and single mode step index fiber and compose them. (Nov-Dec 2014)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 37
9. **A Single mode step index fiber has a core diameter of 7μm and a core refractive index of 1.49. Estimate the shortest wavelength of light which allows single mode operation when the relative refractive index difference for fiber is 1%. (Nov-Dec 2014)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 37
10. **a. Discuss briefly about linearly polarized modes. [6][APR/MAY 2019]**
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 56
    b. Draw the structure of single and multi mode step index fibres and graded index fibres with typical dimensions. [6] [Apr 2018]
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 37
    c. Mention the advantage of optical fibre system [4]. [Nov 2008]
11. **Explain mode propagation in circular waveguides. Obtain its wave equation and modal equations for step index fibers. [Nov-2009]**

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 43

12. **A typical refractive index difference for an optical fiber designated for long distance transmission is 1% .Determine the NA and the solid acceptance angle in air for the fiber when the core index is 1.46.calculate the critical angle at the core-cladding interface with in the fiber [Nov2009]**

    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 23

13. **Draw and explain the working principle of single mode and multimode fiber(Nov 2010)**

    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 37

14. **Distinguish step index and graded index fiber [DEC2016] (Nov 2011)**

    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 37

15. **Derive the mode equations for a circular fibre using maxwell's equations (May 2012)**

    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 43

UNIT- II

TRANSMISSION CHARACTERISTICS OF OPTICAL FIBER

1. **Distinguish between intra model and inter modal dispersion[NOV/DEC 2018]**

   Intramodal dispersion:

   Pulse broadening within a single mode is called as intramodal dispersion or chromatic dispersion

   Intermodal dispersion:

   • Dispersion caused by multipath propagation of light energy is referred to as intermodal dispersion.

2. **Give the measure of information capacity in optical wave guide.[Apr /May 2019]**

   It is usually specified by bandwidth distance product in MHz.For a step index fiber the various distortion effects tend to limit the bandwidth distance product to 20MHz.

3. **What is elastic and inelastic scattering? Give examples. [Apr 2018]**

   Purely elastic scattering means all the pre-collision kinetic energy of the colliding objects goes into kinetic energy of the post-collision objects. A collision between two hard things, like billiard balls, is a good example of a collision that's *mostly* elastic.

   Inelastic scattering means that at least *some* of the pre-collision kinetic energy ends up somewhere else, besides post-collision kinetic energy. For example, the pre-collision kinetic energy can be used to cause an internal state change in one of the colliding objects.

4. **What is meant by attenuation coefficient of a fiber? (N/D2011)[DEC2016]**

   Attenuation coefficient is defined as the ratio of the input optical power P_i launched into the to the output optical

   $$\propto_{dB} = \frac{10}{L} \log_{10}\frac{P_i}{P_o}$$

   power P_o from the fiber.

   where $\propto_{dB}$ is the attenuation coefficient in decibels per kilometer.

5. **A 30 km long optical fiber has an attenuation of 0.8 dB/km. If 7 dBm of optical power is launched into the fiber, determine the output optical power in dBm. (M/J 2012)**

   Given Data: $P_i = 7$ dBm

   $L = 30$ km $= 3\times10^4$ m;

   $\alpha_{dB} = 0.8$dB/km $= 0.8\times10^{-3}$ dB/m

   Solution:

   $$\alpha_{dB} = \frac{10}{L}\log_{10}\frac{Pi}{Po}$$

   $$= \frac{10}{3\times10^4 \log\frac{7}{P_o}}$$

   $\log(7/P_0) = 2.4$;

   $P_0 = \left(\frac{7}{e^{2.4}}\right) =$ **0.63dBm**

6. **What are the types of material absorption losses in silica glass fibers? [DEC 2016]**

   The types of material absorption losses in silica fiber are:

   - Absorption by atomic defects in the glass composition
   - Extrinsic absorption by impurity atoms in the glass material

     Intrinsic absorption by the basic constituent atoms in the glass material

7. **Compare Rayleigh scattering and Mie scattering.**

S.No	Rayleigh Scattering	Mie Scattering
1	Caused due to refractive index variation in the core glass.	Caused by fiber imperfections such as irregularities in the core –cladding interface, core- cladding refractive index difference along the fiber length, diameter fluctuations, strains and bubbles.
2.	When the inhomegenetics size is smaller than the wavelength of light Rayleigh scattering occurs.	When the in homegenetics size is greater than the Wavelength of light, Mie scattering occurs
3	Scattering occurs both is forward and backward direction.	Scattering is mainly in the forward direction.
4.	Rayleigh scattering can be reduced by minimizing the compositional fluctuations by using best manufacturer methods.	Mie scattering can be reduced by Removing imperfections due to the glass manufacturing process. Carefully controlled extrusion and coating of the fiber Increasing the fiber guidance by increasing the relative refractive index difference.

8. Compare Linear scattering and Non- Linear scattering.

S.No	Linear Scattering	Non-Linear Scattering
1.	Linear scatterings are observed only at low optical power densities below the threshold power levels.	Non-Linear scattering are only observed at high optical power densities above the threshold power levels in long single mode fibers.
2.	There are two types are Linear Scattering namely, Rayleigh Scattering Mie Scattering	There are two types of Non-Linear scattering namely, Stimulated Brillouin Scattering (SBS) Stimulated Raman Scattering (SRS)
3.	The Incident light frequency and scattered light frequency is the same. There is no frequency shift during scattering.	The Incident light frequency and scattered light frequency are different. There is a frequency shift during scattering.

9. Compare SRS and SBS.

S.No	SRS	SBS
1.	SRS can occur both in forward and backward direction.	It is mainly backward process.
2.	The threshold power level of SRS is three times higher than SBS threshold in a particular fiber.	The SBS threshold power level is less.
3.	The scattering process produces high frequency optical phonon.	The scattering process produces acoustic phonon as well as a scattered photon.

10. What is meant by intrinsic absorption in optical fibers?

The absorption caused by the interaction of one or more of the major components of the glass is known as intrinsic absorption.

11. What is meant by extrinsic absorption in optical fibers?

The absorption caused by the impurities within the glass is known as extrinsic absorption.

12. Differentiate linear scattering from nonlinear scattering.

- Linear scattering mechanisms transfers linearly some or all of the optical power contained within one propagating mode to a different mode.
- Non-linear scattering causes the optical power from one mode to be transferred in either the forward or backward direction to the same or other modes at different frequencies.

13. What are the types of linear scattering losses? [MAY 2016]

Linear scattering is of two types. They are:

- Rayleigh scattering
- Mie scattering

14.What are the types of nonlinear scattering losses?

Non-linear scattering is of two types. They are

- Stimulated Brillouin Scattering (SBS)
- Stimulated Raman Scattering (SRS)

15. What is meant by Fresnel Reflection? (N/D 2011)

When the two joined fiber ends are smooth and perpendicular to the axes, and the two fiber axes are perfectly aligned, small proportion of the light may be reflected back into the transmitting fiber causing attenuation at joint. This is known as Fresnel reflection.

16. What is meant by linear scattering?

Linear scattering mechanisms transfers linearly some or all of the optical power contained within one propagating mode to a different mode.

17. What are the factors that cause Rayleigh scattering in optical fibers? (M/J 2012)

The inhomogenetics of a random nature occurring on a small scale compared with the wavelength of the light in optical fiber causes Rayleigh scattering. These inhomogeneities manifest themselves as refractive index fluctuations and arise from density and compositional variations that are frozen into the glass lattice on cooling.

18. What are the factors that cause Mie scattering in optical fibers?

The factors that cause Mie scattering in optical fibers are:

- Fiber imperfections such as irregularities in the core – cladding interface
- Core – cladding refractive index differences along the fiber length, diameter fluctuations

19. What are the ways to reduce macro bending losses? (N/D 2009) (N/D 2010)

The ways to reduce macro bending losses are

- Designing fibers with large relative refractive index differences
- Operating at the shortest wavelength possible.

20. What is meant by dispersion in optical fiber? (A/M 2008)

Different spectral components of the optical pulse travel at slightly different group velocities and cause pulse broadening within the fiber. This phenomenon is referred as dispersion.

21. What are the different types of dispersion? (N/D 2008)

There are two types of dispersion. They are

- Intramodal Dispersion:
  - Material Dispersion
  - Waveguide Dispersion
- Intermodal Dispersion:
  - Multimode step index
  - Multimode graded index

22. What is meant by intermodal dispersion? (A/M 2010 A/M 2008)

Pulse broadening due to propagation delay differences between modes within a multimode fiber is known as intermodal dispersion.

23. Define – Group Velocity Dispersion (GVD) (A/M 2011), (N/D 2010)

Intra-modal dispersion is pulse spreading that occurs within a single mode. The spreading arises from the finite spectral emission width of an optical source. This phenomenon is known as Group Velocity Dispersion (GVD).

24. What is meant by modal noise? (A/M 2011)

The speckle patterns are observed in multimode fiber as fluctuations which have characteristic times longer than the resolution time of the detector. This is known as modal or speckle noise.

25. What is meant by chromatic dispersion? (N/D2011)

The dispersion due to the variation of the refractive index of the core material as a function of wavelength is known chromatic dispersion. This causes a wavelength dependence of the group velocity of any given mode. Pulse spreading occurs even when different wavelengths follow the same path.

26. What is meant by polarization mode dispersion? (N/D 2007) [Apr 2018]

Polarization refers to the electric - field orientation of a light signal, which can vary significantly along the length of the fiber.

27. Distinguish between dispersion shifted and dispersion flattened fibers. (N/D 2007)

Reduction in the fiber core diameter with an increase in the relative or fractional index difference to create dispersion is known a dispersion shifted fiber. Fibers which relax the spectral requirements for optical sources and allow flexible wavelength division multiplexing are known as dispersion flattened fibers.

28. What are the two types of fiber joints? The two types of fiber joints are:

(i) Fiber splices: These are semi permanent or permanent joints.
(ii) Demountable fiber connectors or simple connectors: These are removable joints.

29. What is meant by fiber splicing?

A permanent joint formed between two individual optical fibers in the field or factory is known as fiber splice.

30. What are the techniques used in splicing?

Generally used splicing techniques are:

- Fusion splice
- V-groove mecha/ nical splice
- Elastic tube splice

31. List the types of mechanical misalignments that occur between two joined fibers.

There are three types of mechanical misalignments:

- Lateral/radial/axial misalignment
- Longitudinal misalignment
- Angular misalignment

32. What are the causes of absorption?

- Absorption by atomic defects in glass composition.
- Extrinsic absorption by impurity atoms in the glass materials.
- Intrinsic absorption by basic constituent atoms.

33. What is polarization mode dispersion?

The difference in propagation times between the two orthogonal polarization modes will result pulse spreading. This is called as polarization mode dispersion. (PMD)

34. Define signal attenuation.

If P(0) is the optical power in a fiber at the origin (at Z = 0), then the power P(Z) at a distance z further down the fiber is

$P(z) = P(0)\, e^{-\alpha_p z}$.

The above equation can be rewritten as

$\alpha_p = (1 / z) \{ P(0) / P(z)\}$. Where α_p is the fiber attenuation coefficient given in units of km^{-1}

35. What are bending losses? Name any two types,

Micro bend losses

Macro bend losses

36. What is meant by Polarization of light?

The polarization of light describes by a specifying the orientation of the waves electric field at a point in space over one period of the oscillation. When light travels in free space, it propagates as a transverse wave, i.e. the polarization is perpendicular to the wave‟s direction of travel.

37. What is fiber Bi - refrigence?

Fiber bi-refrigence is the optical property of a material having a refractive index that depends on the polarization and propagation direction of light. These optically anisotropic materials are said to be bi-refringence. The bi-refringence is often quantified by the maximum difference in the refractive index within the material.

38. Define Beat length?

Beat length is defined as the period of interference effects in a bi- refringence medium. When two waves with different linear polarization states propagate in a bi-refringence medium, their phases will evolve differently. It is assumed that the polarization of each wave is along the principle directions of the medium (x – axis (or) y – axis), so that this polarization will be preserved during propagation. This means that the phase relation between both waves is restored after integer multiples called the polarization beat length.

39. Define PMF (Polarization Mode Fiber)?

PMF is an optical fiber in which the polarization of linearly polarized light waves launched into the fiber is maintained during propagation, with less or no cross-coupling of optical power between the polarization modes. Such fiber is used in special application where processing the polarization is essential.

40. What are Dispersion Flattened Fibers (DFF)?

DFF is a type of glass optical fiber that provides low pulse Dispersion over a broad portion of the light spectrum and as a result can operate at 1300 nm and 1550 nm wavelength simultaneously.

41. What are Dispersion Shifted Fibers (DSF)?

DSF is a type of optical fiber made to optimize both low dispersion and low attenuation. DSF is a type of single mode optical fiber with a core-clad index profile tailored to shift the zero-dispersion wavelength from the natural 1300 nm in silica-glass fibers to the minimum loss at 1550 nm.

42. What is meant by Fresnel reflection in Fiber cable?

Fresnel reflection at the air-glass interfaces at the entrance and exit of an optical fiber.

43. List out the advantages of elastic tube splicing?

The advantages of elastic tube splicing are,

a) This type of splicing allows accurate and automatic alignment of axes of the two fibers to be joined.

b) In this method the fibers to be splices do not have to be equal in diameter.

44. List out the advantages of V-groove splicing?

a) There is no thermal stress.

b) Nochangeinrefractionindexofthetwofibers. c)

45. What are bending losses? Name any two types. (Apr-May 2015)

(i) Micro bending losses - The light power is dissipated through the micro bends because if the respective coupling of energy between guided modes and leaky modes.

(ii) Macro bending losses - Macrobending losses occur when fibres are physically bent beyond the point at which the critical angle is exceeded.

46. What are the types of fiber losses which are given per unit distance?(Nov-Dec 2014)

(i)Absorption

(ii) Scattering

(iii)Bending Loss

47. List the factors that cause intrinsic joint losses in a fiber. (N0v-Dec 2014)

(i) Different core and / or cladding diameters

(ii) Differentnumericalaperturesand/orrelativerefractive index differences.

(iii) Different refractive indexprofiles.

(iv) Fiber faults.

48. Define dispersion in multimode fibers. What is its effect? (Nov-Dec 2013) (R)

In multimode fiber many modes are propagating along the fiber at a time. Different modes are taking different ray path and they reach at different time at the output end of the fiber. So a time delay is experienced between modes. This is called intermodal delay and pulse broadening occurs due to intermodal delay is called intermodal dispersion

Effect:

1. It restricts bandwidth of the optical fibercable.
2. The intermodal dispersion causes the light rays to spread out through the fiber.
3. It accounts for a significant loss occurring in thefiber.

49. What are the two reasons for Chromatic Dispersion? (Nov-Dec 2012)

- Dispersive Properties of the waveguide material – **Material Dispersion**
- Guidance effects within the fiber structure– **Waveguide Dispersion**

50. What are the most important non-linear effects of optical fibre communication? (Nov-Dec 2012)

Non linear effects of Scattering are:

- Stimulated Brillouin Scattering (SBS)
- Stimulated Raman Scattering (SRS)

51. A fiber has an attenuation of 0.5dB/Km at 1500nm. If 0.5mW of optical power is initially launched into the fiber estimate the power level after 25km.

☐ $$P_{out}\,(dBm) = P_{in}(dBm) - \alpha\left(\frac{dB}{Km}\right) \times l$$

$$\log_{10}\frac{P_{in}(dBm)}{1mW}$$

$$P_{in}(dBm) = 10\log_{10}\frac{(0.5 \times 10^{-3})}{(1 \times 10^{-3})}$$

$-$

$$P_{out}\,(dBm) = - = 10 - (0.5 \times 25)$$

$$P_{out}\,(dBm) = -$$

PART B & C

1. **Discuss about the design optimization of single mode fiber.(Nov-Dec 2016)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 115·
2. **What is waveguide dispersion? Derive and expression for time delay produced due to waveguide dispersion.(Nov-Dec 2016,APR/MAY 2019)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 109·
3. **With necessary diagrams, explain the causes and types of fiber attenuation loss. (Nov-Dec 2015)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 94·
4. **With diagram, derive the expression for intra modal dispersion. (Nov-Dec 2015)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 109·
5. **What are the loss or signal attenuation mechanism in a fibre? Explain.(Apr-May 2015, NOV/DEC 2018)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 94·
6. **Discuss the pulse broadening in graded index fiber (April 2005,Nov 08)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 105
7. **Derive the expression for pulse broadening due to material dispersion.**
   Refer Book: Optical fiber Communications - Gerd Kaiser -Pg. No. 108
8. **What is meant by waveguide dispersion ? Derive the expression for the same (Nov 2006)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 109
9. **Explain the attenuation mechanisms in Optical fibers. (Dec 2007)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 94
10. **Explain bending losses and type of dispersion (Nov 2010)**
    Refer Book: Optical fiber Communications - Gerd Kaiser -Pg. No. 108
11. **Describe the linear and non-linear scattering losses in optical fibers (Nov 2012)**
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 94
12. **Discuss the attenuation encountered in optical fiber communication due to:**
    1. Bending 2. Scattering 3. Absorption.
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 94
13. **Explain the attenuation and losses in fiber. (May 2014) [DEC2016]**
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 94
14. **With diagram, explain intra and inter modal dispersion. (May 2014, AAPR/MAY2019)**
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 115
15. **Explain signal distortion in single mode fibers.**
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 115·

UNIT-III

SOURCES AND DETECTORS

1. **What is meant by heterojunction? List out the advantages of heterojunction.(A/M 2011) (N/D 2007)**
   A heterojunction is an interface between two adjoining single crystal semiconductors with different bandgap energies. Devices that are fabricated with heterojunction are said to have hetrostructure.
   Advantages of heterojunction are:
   - Carrier and optical confinement
   - High output power
   - High coherence and stability

2. **Distinguish between direct and indirect band gap materials. (N/D2010), (N/D 2008)**

Direct bandgap materials	Indirect bandgap material
The electron and hole have the same momentum value	The conduction band minimum and the valence band maximum energy level occur at different values of momentum.
Direct transition is possible from valence band to conduction	Direct transition is not possible from valence band to conduction

3. **Why is silicon not used to fabricate LED or Laser diode? (N/D2011, 2018)[DEC2016]**
   Silicon is not used to fabricate LED or Laser diode because
   - It is an indirect bandgap semiconductor
   - It has Eg level of 1.1eV, the radiated emission corresponds to infrared but not the visible light.
4. **What are the advantages of LED?(M/J2012)**
   The advantages of LED are:
   - Less expensive
   - Less complex
   - Long life time
   - Used for short distance communication
5. **When an LED has 2V applied to its terminals, it draws 100mA and produces 2mW of optical power. Determine conversion efficiency of the LED from electrical to optical power. (N/D2008)**

   Given Data: $V_{in} = 2$ V, $I_{in} = 100\times10^{-3}$ A, $P_{out} = 2\times10^{-3}$

   Formula: LED conversion efficiency = $\frac{P_{out}}{P_{in}}$

   Solution:

   $P_{in} = V_{in} \times I_{in} = 2 \times 100\times10^{-3}$

   Conversion Efficiency = $\frac{2\times10^{-3}}{2\times100\times10^{-3}}$ = **0.01**
6. **What are the advantages and disadvantages of LED?**
   Advantages:
   - Small size and light weight;
   - High Speed;
   - Low operating temperature;
   - Longer life;
   - No complex driver capacity required.

   Disadvantages:
   - Quantum efficiency is low;
   - Damages due to over voltage and over current;

- Temperature dependent.

7. What are the three requirements of Laser action? [DEC2016]

The three requirements of Laser action are

- Absorption
- Spontaneous emission
- Stimulated emission

8. What is the principle of operation of LASER? (N/D2008)

The principle of operation of LASER is population inversion, the most photons incident on the system. The population of the upper energy level is greater than lower energy level i.e. $N_2 > N_1$. This condition is known as population inversion.

9. Write the three modes of the cavity of LASER diode. (N/D2009)

The three modes of the cavity of LASER are:

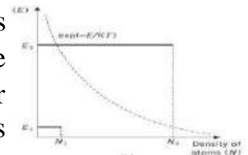
- Longitudinal modes, related to the length L of the cavity
- Lateral Modes lie in the plane of the P-N junction. These modes depend upon the side wall preparation and width of the cavity.
- Transverse modes are associated with the Electro Magnetic Field and beam profile in the direction perpendicular to the plane of the PN junction. These modes determine the radiation pattern of the LASER.

10. What is a DFB Laser? Differentiate DFB LASER from other types of LASER.(N/D2009)

In DFB Laser, the lasing action is obtained by periodic variations of refractive index, which are incorporated into multilayer structure along the length of the diode. DFB LASER does not require optical feedback unlike the other LASERs.

11. What is population inversion? (A/M 2008)

Under thermal equilibrium, the lower energy level E_1 of the two level atomic system contains more atoms than upper energy level E_2.To achieve optical amplification it is necessary to create non-equilibrium distributions of atoms such that population of the upper energy level is greater than lower energy level i.e. $N_2 > N_1$ as shown in the figure. This condition is known as population inversion.



12.Compare LED and ILD sources.[Apr 2017] (A/M 2008)

Sl.N	LED	ILD
1.	Incoherent	Coherent
2.	For multimode fibers only	For multi and single mode fibers
3.	Large beam divergence due to s	Low beam divergence due to s

13. Write the three key processes of laser action.(A/M 2008)

The three key processes of laser actions are:

- The atomic system must have population inversion. This means the number of atoms in the excited state should be more than that of ground state
- There should be photons with proper energy to start the stimulated emission
- There should be an arrangement for multiple reflections to increase the intensity of LASER beam

14. What are the advantages of Quantum Well Lasers?(N/D2009)

The advantages of Quantum Well Lasers are:

- High threshold current density
- High modulation speed
- High line width of the device

15. Define – Internal Quantum Efficiency[NOV /DEC 2018]

Internal Quantum Efficiency is defined as the ratio of radiative recombination rate to the total recombination rate.

$$\eta_{in} = \frac{R_r}{R_r + R_{nr}}$$

where R_r is radiative recombination rate, R_{nr} is the non-radiative recombination rate.

16. Define – External Quantum Efficiency

The external quantum efficiency is defined as the ratio of photons emitted from LED to the number of photons generated internally.

17. Define – Quantum efficiency of a photo detector(A/M2008,10) (M/J2009)(N/D2011)

Quantum efficiency is defined as the number of the electron-hole carrier pairs generated per incident photon of energy hv, is given bynumber of electron−hole pairs generated

number of incident photons

$$\eta = \frac{I_p/q}{P_o/hv}$$

where I_p is the photon current
q is the charge of the electron
p_o is the optical output power
h is the Planck's constant
v is the frequency of the optical signal

18. An LED has radiative and nonradiative recombination times of 30 and 100 ns respectively. Determine the internal quantum efficiency. (N/D 2007) (N/D 2010)

Given data: $\tau = 30\times 10^{-9} sec$, $\tau_{nr} = 100\times 10^{-9} sec$

Formula: $\tau = \frac{\tau_r \times \tau_{nr}}{\tau_r + \tau_{nr}} = \frac{30\times10^{-9}\times100\times10^{-9}}{130\times10^{-9}} = 23.1ns$

Solution: $\eta_{int} = \frac{\tau}{\tau_r} = \frac{23.1\ ns}{30ns} = \mathbf{0.77 = 77\%}$

19. Calculate the external differential quantum efficiency of a laser diode operating at 1.33µm.The slope of the straight line portion of the emitted optical power P versus drive current I is given by 15 mW/mA. (N/D2011)

Given data: $\lambda = 1.33\times10^{-6}$

$\frac{q}{Eg} = 0.8065\ \lambda = 0.8065\times1.33\times10^{-6}$ $M = (q)/Eg \times \frac{dP}{dI} = 0.8065\times1.33\times10^{-6}\times 15\times10^{-3}$

$\frac{dP}{dI} = 15\times10^{-3}$ $= \mathbf{16.089\%}$

20. What are the necessary features of a photo detector? (N/D2007)

The necessary features of a photo detector are:

- High Quantum efficiency
- Low rise time or fast response
- Low dark current

21. Define – Responsivity of a photodetector (N/D2008, 2018),(N/D 2010)

Responsitivity is defined as the ratio of output photo current to the incident optical power.

$R = \frac{I_p}{P_O} = \frac{\eta q}{hv}$

where, R=Responsivity.

I_p=Output photo current P_O=Incident optical power

22. Compare the performance of APD with PIN diode. (N/D2008)

APD	PIN
No internal gain	Internal gain is high
Thermal current noise dominates photo detector noise current	Photo detector noise current dominates thermal noise current
Low responsivity	High responsivity
Low dark current	High dark current
Suitable for high intensity application	Suitable for low intensity application

Required low reverse bias voltage	Required high reverse bias voltage

23. List out the operating wavelengths and responsivities of Si, Ge, and InGaAs photodiodes. **(N/D2009)**

The Operating Wavelengths and Responsivities of Si, Ge, and InGaAs photodiodes are:

Silicon (Si) :

- Operating wavelength range λ = 400 - 1100 nm
- Responsivity R = 0.4-0.6

Germanium (Ge) :

- Operating wavelength range λ = 800 - 1650 nm
- Responsivity R = 0.4 - 0.5

Indium Gallium Arsenide (InGaAs):

- Operating wavelength range λ = 1100 - 1700 nm
- Responsivity R = 0.75 - 0.95

24. List the benefits and drawbacks of avalanche photodiodes.

Benefits of APD are:

- Carrier multiplication takes place.
- Sharp threshold

Drawbacks of APD are:

- High biasing voltage.
- Noisy

25. Photons of energy 1.53×10^{-19} J are incident on a photodiode that has the responsivity of 0.65Amps/W. If the optical power level is 10μW, find the photo current generated.(M/J 2012)

Given data : $E = 1.53\times10^{-19}$ J, R = 0.65Amps/W, $P_0 = 10\times10^{-6}$ W

Formula : $I_p = R\times P_0$

Solution : $I_p = 0.65\times 10\times10^{-6}$ = **6.5 μA**

26. GaAs has band gap energy of 1.43ev at 300k. Determine the wavelength above which an intrinsic photo detector fabricated from this material will cease to operate.(A/M 2008)

Given data: Eg(eV) = 1.43ev

Formula: λ(μm) = 1.24/Eg(eV)

Solution: λ(μm) =1.24/1.43

λ(μm) =**0.86 μm**.

27. Define – Photocurrent

The high electric field present in the depletion region causes the carriers to separate and be collected across the reverse- biased junction. This gives to a current flow in the external circuit, with one electron flowing for every carrier pair generated. This current flow is known as photocurrent.

28. Define – Impact Ionization

In order for carrier multiplication to take place, the photo-generated carriers must traverse a region where a very high electric field is present. In this high field region, a photo generated electron or hole can gain energy so that it ionizes bound electrons in the valence band upon colliding with them. This current multiplication mechanism is known as impact ionization.

29. Define – Avalanche Effect

The newly created carriers are accelerated by the high electric field, thus gaining enough energy to cause further impact ionization. This phenomenon is called avalanche effect.

30. llustrate the factor that determine the response time of the photodiode.

The resistance and capacitance of the photodiode and the external circuitry give rise to another response time known as RC time constant{\display style \tau =RC}. This combination of R and C integrates the photo response over time and thus lengthens the impulse response of the photodiode. When used in an optical communication system, the response time determines the bandwidth available for signal modulation and thus data transmission.

31. Define dark current?

The photo diode dark current is the current that continues to flow through the bias circuit of the device when no light is incident on the photo diode.

32. Define Johnson or thermal noise?

When current is flowing continuously across the load resistor, heat will be dissipated. This is called thermal noise.

33. What is known as detector response time? (May 2012, NOV/DEC2018)

It is defined as the time taken for the photo detector to respond to an optical input pulse. The response time determines the bandwidth available for signal modulation and data transmission.

34. What are the factors that limit the response time of the photo detectors?

- Transit time of photo carriers within the depletionregion.
- Diffusion time of photo carriers outside the depletion region.
- RC time constant of the photo diode and its associated circuit.

35. What are inherent connection problems when joining fibers?

The inherent connection problems when jointing fibers are,

- Different core and/or cladding diameters.
- Different numerical apertures and/or relative refractive index differences.
- Different refractive index profiles.
- Fiber faults(core elliptically, core concentricityetc

36. Compare PIN and APD?

S.No	PIN	APD
1	No internal gain.	Internal gain.
2	Thermal noise current dominates photo detector noise current.	Photo detector noise current dominates thermal noise current.
3	Low responsivity.	High responsivity.
4	Low dark current.	High dark current.
5	Suitable for high intensity application.	Suitable for low intensity application.
6	Required low reverse bias voltage.	Required high reverse biasvoltage.

37. List out the different types of mechanical misalignments during fiber connection?

The three possible types of misalignment which may occur when joining compatible optical fibers are,

- Longitudinal misalignment
- Lateral misalignment
- Angular misalignment

38. What is fiber splicing?

Fiber splicing is the process of joining two fibers by melting the fiber ends.

39. Compare splices and connectors.

S.No	Splices	Connectors
1	Permanent or semi permanent joints	Temporary joint
2	Splice loss is low	Connector loss is high

40. Define power- bandwidth product.(Apr-May 2015)

High output power and high bandwidth are two important parameters in the design of photo-detector. The Product of photodetector bandwidth and power at which bandwidth ismeasured.

41.Contrast the advantages of PIN diode with APD diode. (Apr-May 2015)

- Low dark current

- It is affected but only thermal noise
- No speed limitation due to capacitance effect

42. What is meant by Mechanical splicing? (May-June 2013)

Mechanical splicing, in which the fibers are held in alignment by some mechanical means, may be achieved by including the use of V-groove into which the butted fibers are placed (or) the use of tubes around the fiber ends.

43. Calculate the Band gap energy for an LED to emit $850nm$? (May- June 2013)

Solution $\lambda = 850nm = 0.85\mu m$

$$E_g = \frac{hc}{\lambda} = \frac{6.625 \times 10^{-34} \times 3 \times 10^8}{0.85 \times 10^{-6}} = 2.33 \times 10^{-19} = 1.45\ eV$$

$$E_g = 1.45\ eV$$

44. Define external quantum efficiency.(Nov-Dec 2016).

The external quantum efficiency is defined as the ratio of the photons emitted from the LED to the number of internally generated photons.

45. Write two difference between a Laser diode and a LED. (Nov-Dec 2013)

S.no	Laser Diode	LED
1.	Coherent radiation Takes place.	In coherent radiation takes place.
2.	Narrow spectral width	Wider Spectral width

46. Why silicon is preferred for fabrication of photo receiver?

- Silica is used for fabrication photo receiver, because it has larger band gap, it generates low noise and it supports multiple channels as it has larger bandwidth.
- Silicon is available plenty in nature.

47. Why are semiconductor based photo detectors preferred to other types of photo detectors?

Semiconductor laser diode generates low noise and they support multiple channels as they have larger bandwidth.

48. What are the requirements of photo detector?

- The photo detector must have high quantum efficiency to generate a large signal power.
- The photo detector and amplifier noises should be kept as low as possible.

49. What is the significance of intrinsic layer in PIN diode.[APR/MAY 2019]

The intrinsic region in the diode is in contrast to a PN junction diode. This region makes the PIN diode an lower rectifier, but it makes it appropriate for fast switches, attenuators, photo detectors and applications of high voltage power electronics.

50. What are the factors that limit the response time of the photo detectors?

- Transit time of photo carriers within the depletionregion.
- Diffusion time of photo carriers outside the depletion region.
- RC time constant of the photo diode and its associated circuit.

51.What are inherent connection problems when joining fibers?

The inherent connection problems when jointing fibers are,

- Different core and/or cladding diameters.
- Different numerical apertures and/or relative refractive index differences.
- Different refractive index profiles.
- Fiber faults(core elliptically, core concentricity etc)

PART B & C

1. Discuss the LASER diode principle, modes and threshold conditions. (Jun 2007)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 163 **(Apr./May 2008)**

2. **a. Derive the threshold condition for lasing. [DEC2016]**
   b. Explain in detail the fabry perot resonator cavity Laser diode.
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 163 **(Apr./May 2008)**

3. **a. Explain the various lensing schemes. (Nov 2004,Dec 05,07,April 09, Apr 2017. Apr 2018)**
   b. Explain the various splicing techniques.
4. **Discuss about modulation of LED & Quantum LASER (Nov 2010)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 152
5. **With neat diagram explain the working of surface emitting LED (Nov 2011, 2018, may 2012, Apr 2017, 2019)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 152
6. **Explain the structure of silicon ADP**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 249
7. **Explain any two injection laser structure with neat diagram (May 2012)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 163 **(Apr./May 2008)**
8. **Derive laser diode rate equations. (16 marks) [APR/MAY 2019] (Nov 2012)[DEC2016]**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 163 **(Apr./May 2008, Apr 2018)**

9. **(i) Explain the working of n hetero structure LED. (Nov2013, NOV 2018)**
   (ii) Define internal quantum efficiency of a LED. Deduce the expression for the same.
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 152
10. **Explain expanded beam connectors with neat diagram (Nov 2011)**
    Refer Book: Optical fiber Communications –John M.Senior - Pg. No. 238-244.

11. **Write brief note on fiber alignment and joint loss (May 2012)**
    Refer Book: Optical fiber Communications –John M.Senior - Pg. No. 227-234.
12. **Describe about connectors, splices and couplers. (Nov-Dec 2015)**
    Refer Book: Optical fiber Communications –John M.Senior - Pg. No. 238-244.

13. **A Photodiode is constructed of GaAs which has a band gap energy of 1.43eV at 300K. Find the long wavelength cut-off. (Apr-May 2015)**
    Refer Book: Optical fiber Communications –John M.Senior - Pg. No. 238-244.
14. **What do you understand by optical-wave confinement and current confinement in LASER diode? Explain with suitable structures. (Nov-Dec 2013)**
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 163
15. **Discuss about optical detection noise. (Nov-Dec 2015)**
    Refer Book: Optical fiber Communications –John M.Senior - Pg. No. 238-244.

16. **Explain gain guided and Index guided laser diodes.**
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 163

UNIT-IV

FIBER OPTIC RECEIVER AND MEASUREMENTS

1. **What is bit rate?**
The transmitted signal is two level binary data stream consisting of either 0 or 1 in a time slot of duration T. this time slot is referred to a bit period.

2. **What are the error sources of receiver? (M / J 2013)**
The error sources of receiver are
    - Thermal noise
    - Dark current noise
    - Quantum noise

3. **A digital fiber optic link operating at 1310 nm, requires a maximum BER of 10^{-8}. Calculate the required average photons per pulse. (N / D 2013)**
The probability error $P_r(o) = e^{-N} = 10^{-8}$
Solving for $N = 8 \log_e 10 = 18.42$
An average of 18 photons per pulse is required for this BER.

4. **Define: Probability of error**.
Probability of error means that a transmitted' l' is misinterpreted as a '0' or transmitted
'0' is misinterpreted as a '1' by the receiver.

5. **Define – Quantum Limit (M/J2012), (N/D2007)9DEC2016)**
The minimum received power level required to maintain a specific Bit – Error - Rate (BER) of an optical receiver is known as the quantum limit.

6. **What is meant by (1/f) noise corner frequency? (N/D2009)**
The (1/f) noise corner frequency is defined as the frequency at which (1/f) noise, which dominates the FET noise at low frequencies and has (1/f) power spectrum.

7. **Why silicon is preferred to make fiber optical receivers? (N/D2010), (A/M2011)**
Silicon is preferred to make fiber optical receivers because
It has high sensitivity over the 0.8–0.9 μm wavelength band with adequate speed
It provides negligible shunt conductance, low dark current and long-term stability

8. **Define – Modal noise and Mode Partition Noise.(A/M2011)(M/J2009)(A/M2010)**
Disturbances along the fibre such as vibrations, discontinuities, connectors, splices and source/detector coupling may cause fluctuations in the speckle patterns. It is known as
modal noise.
Phenomenon that occurs in multimode semiconductor lasers when the modes are not well stabilized is known as mode partition noise.

9. **Mention the error sources in fiber optical receiver. (N/D2011)**
There are three main error sources in fiber optical receiver. They are:
    - Thermal noise
    - Dark current noise
    - Quantum noise

10. **Define – Bit Error Rate(DEC2016)**
Bit Error Rate (BER) is defined as the ratio of the number of errors occurred over a certain time interval't' to the number of pulses transmitted during this interval.

11. **How does dark current arise?**
When there is no optical power incident on the photo detector a small reverse leakage current flows from the

device terminals known as dark current. Dark current contributes to the total system noise and gives random fluctuations about the average particle flow of the photocurrent

12. **What is Inter Symbol Interference?**

Each pulse broadens and overlaps with its neighbors, eventually becoming indistinguishable at the receiver input. This effect is known as Inter Symbol Interference.

13. **Define – Extinction ratio**

The extinction ratio is usually defined as the ratio of the optical energy emitted in the '0' bit period to that emitted during the '1' bit period.

14. **What are the requirements of an optical receiver?(Nov. / Dec. 2006)**
    - It should have fast response;
    - High sensitivity;
    - Tolerable SNR;
    - Should reproduce the signal without any distortion.

15. Define – Minimum Detectable Optical Power

It is defined as the optical power necessary to produce a photocurrent of the same magnitude as the root mean square of the total current.

16. What are the noise effects on system performance?

The main penalties are modal noise, wavelength chirp, spectral broadening, mode-partition noise.

17. Why the attenuation limit curve slopes towards to the right?

As the minimum optical power required at the receiver for a given BER becomes higher for increasing data rates, the attenuation limit curve slopes downward to the right.

18. What do you mean thermal noise?

Thermal noise is due to the random motion of electrons in a conductor. Thermal noise arising from the detector load resistor and from the amplifier electronics tend to dominate in applications with low signal to noise ratio.

19. What is meant by excess noise factor?

The ratio of the actual noise generated in an avalanche photodiode to the noise that would exist if all carrier pairs were multiplied by exactly m is called the excess noise factor. (F).

20. What are the system requirements?

The key system requirements are as follows

- The desired or possible transmission distance
- The data rate or channel bandwidth
- Bit error rate (BER)

21. Give the two analyses that are used to ensure system performance.

The two analyses that are used to ensure system performance are

- Link power budget analysis
- Rise time budget analysis.

22. What are the requirements of preamplifier?

It should have low noise level, high bandwidth, high dynamic range, and high sensitivity to avoid non linearity and high gain.

23. What are the types of pre - amplifiers?

The types of pre-amplifier are

- Low- impedance preamplifier
- High – impedance preamplifier

Transimpedance preamplifier

24. List the advantages of preamplifiers.

The advantages of pre amplifiers are

- Low noise level
- High Bandwidth
- High dynamic range

- High Sensitivity
- High gain

25. What are the standard fiber measurement techniques?

The standard fiber measurement techniques are

- Fiber attenuation measurement
- Fiber dispersion measurement
- Fiber refractive index profile measurement
- Fiber cutoff wavelength measurement
- Fiber numerical aperture measurement
- Fiber diameter measurement

26. Define – Bend Attenuation

A peak in the wavelength region where the radiation losses resulting from the small loop are much higher than the fundamental mode is known as bend attenuation.

27. **What is the technique used for measuring the total fiber attenuation?**

Total fiber attenuation per unit length can be determined using cut-back method. Taking a set of optical output power measurements over the required spectrum using a long length of fiber usually at least a kilometre is known as cut back technique. The fiber is then cut back to a point 2 meters from the input end and maintaining the same launch conditions, another set of power output measurements are taken.

Relationship for the optical attenuation per unit length α_{db} for the fiber may be obtained from,

$$\alpha_{db} = \frac{10}{(L_1 - L_2) log_{10} \frac{P_{02}}{P_{01}}}$$

where,

L_1, L_2 - original and cut-back fiber length respectively

P_{02}, P_{01} -output optical powers at a specific wavelength from the original and cut back fiber lengths.

28. What are the factors that produce dispersion in optical fibers?

The factors that produce dispersion in optical fibers are:

- Propagation delay difference between the different spectral components of the transmitted signal.
- Variation in group velocity with wavelength

29. What are the methods used to measure fiber dispersion?

The methods used to measure fiber dispersion are:

- Time domain measurement
- Frequency domain measurement

30. What are the methods used to measure fiber refractive index profile? (M/J2012)

The methods used to measure fiber refractive index profile are

- Interferometric method
- Near infra scanning method
- Refracted near field method

31. List the process associated with fiber optic receiver section.

Although the photo-detector is the major element in the fibre optic receiver, the other elements to the whole unit. Once the light has been received by the fibre optic receiver and converted into electronic pulses, the signals are processed by the electronics in the receiver. Typically these will include various forms of amplification including a limiting amplifier. These serve to generate a suitable square wave that can then be processed in any logic circuitry that may be required.

32. What is Mode Coupling and what are its causes?

It is another type of pulse distortion which is common in optical links.

The pulse distortion will be increased less rapidly after a certain initial length of fiber, due to this mode coupling and differential mode losses occur.

33. Define Quantum limit (Q). (May-June 2013)

The minimum received power level required for a specific BER of digital system is known as Quantum limit.

34. List out the methods used to measure fiber refractive index profile.

1. Inter-ferometric method
2. Near field scanning method
3. End field scanning method

35. What are the error sources in fiber optic receiver? (May-June 2013, Nov-Dec 2012)

The error sources in fiber optic receiver are

- Shot Noise
- DarkCurrent
  Bulk Dark Current SurfaceDarkCurrent
- Thermal Noise.
- Amplifier noise

36. What are the different techniques for determining attenuation in optical fiber?

The different techniques for determining attenuation are

i) Cut-back ii) Insertion-loss

37. Write the expression to measure attenuation using cut back method.

$$\alpha_{dB} = \frac{10}{L_1 - L_2}\log_{10}\frac{V_1}{V_2}$$

38. List any two advantages of trans-impedance amplifiers.(Apr-May 2015)

(i) Reduces thermal noise

(ii) Provide wide bandwidth

39. State the significance of maintaining the fiber outer diameter constant.(Nov-Dec 2014)

It is essential during the fiber manufacturing process (at the drawing stage) that the fiber outer diameter (cladding diameter) is maintained constant to within 1%. Any diameter variations may cause excessive radiation losses and make accurate fiber – fiber connection difficult.

40. Mention few fiber diameter measurement techniques. (Nov-Dec 2015)

There are two very broad classifications of diameter measurements techniques

(i) Contacting or destructive methods

(ii) Non-contacting and nondestructive methods

41. What is dark current?(Nov-Dec 2012)

The photo diode dark current is the current that continues to flow through the bias circuit of the device when no light is incident on the photo diode.

42. A trignometrical measurement is performed in order to determine the numerical aperture of a step index fiber. The screen is positioned 10.0cm from the fiber end face. When illuminated from a wide angled visible source the measurement output pattern size is6.2 cm. Calculate the Numerical Aperture of the fiber.

NA = A / (A^2 +4D^2) = 6.2 / (38.44 + 400) = 0.30

43. In a fiber optic system operating at 1.3 µm, the transmitter power = -3 dB m. Fiber cable loss = 0. 2 dB/km. Total connector loss at the transmitter and receiver = - 2dB. PINreceiver sensitivity when operating at 400 Mbps = - 44 dB m. Safety margin = 6 dB. Calculate the maximum transmission distance without repeaters (i) when there is no dispersion equalization penalty and (ii) when there is a dispersion equalization penalty o f1.5 dB.

(i) P Tr- P Rr= a fL + a c+ P SM= - 3 dB - (- 44 dB)

=0.2 x L + 2 + 641-8=0.2L

L= 33/0.2 = 165 km

(ii) P Tr- P Rr= a fL + a c+ P SM+ D L

41 =0.2xL+2+6+1.5

L = 31.5/0.2 = 157.5 km

45. State detector response time

This is the measure of the photodiode response speed to a stepped light input signal. It is the time required for the photodiode to increase its output from 10% to 90% of final output level. is the rise timeof the device.

46. Give the 2 analysis that are used to ensure system performance?

The 2 analysis that are used to ensure system performance are: • link power budget analysis • rise time budget analysis

47. Explain briefly about link power budget analysis?

In the optical power loss model for a pt-to-pt link, the optical power rxed at the photo detector depends on the amount of light coupled into the fiber & losses occurring in the fiber at the connectors & splices. The link loss budget is derived from the sequential loss contribution of each element in the link. Loss=10 log (Pout) (Pin) The total optical power loss is, PT = PS - PR

48. **Give the range of system margin in link power budget?** The system margin is usually (6-8) dB. A positive system margin ensures proper operation of the circuit. A negative value indicates that insufficient power will reach the detector to achieve the required bit error rate, BER.

49. The specifications of the light sources are converted to equivalent rise time in rise time budget. Why? A rise time budget is a convenient method to determine the dispersion limitation of an optical link. This is particularly useful for digital systems. For this purpose, the specifications of the light sources (both the fiber & the photo detector) are converted to equivalent rise time. The overall system rise time is given in terms of the light source rise time, fiber dispersion time & the photo detector rise time.

50. **What are the system components of system rise time?** The 4 basic system components that contribute to the system rise time are: • transmitter (source) rise time • receiver rise time • material dispersion time of the fiber • modal dispersion time of the fiber link All these 4 basic elements may significantly limit system speed.

51. Why the attenuation limit curve slopes downwards to the right? As the minimum optical power required at the rxer for a given BER becomes higher for increasing data rates, the attenuation limit curve slopes downward to the right.

PART B & C

1. **Draw and explain the high impedence of high impedence pre-amplifier designed based on BJT and FET (8)**
   b) Write a brief note on trans impedence amplifier. [8] (Nov/Dec 2008, 2018)
   Refer Book: Optical fiber Communications - John M.Senior - Pg. No . 377.
2. **Explain the operation of preamplifier built using FET (Nov 2011)**
   Refer Book: Optical fiber Communications - John M.Senior - Pg. No . 377.
3. **Explain measurement technique used in the case of fiber diameter, fiber cut off length, refractive index profile, Numerical aperture (Nov 2011, 2018)**
   Refer Book: Optical fiber Communications - John M.Senior - Pg. No . 779 -781.
4. **Draw and explain the operation of high impedance FET and BJT preamplifiers(May 2012)**
   Refer Book: Optical fiber Communications - John M.Senior - Pg. No . 377.
5. **Explain a) Attenuation measurement using cut back technique**
   b)Frequency domain measurement of fiber dispersion (May 2012) (DEC2016)
   Refer Book: Optical fiber Communications - John M.Senior - Pg. No . 782 -783.
6. **Considering the probability distributions for received logic 0 and 1 signal pulses.**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 282**(Nov. / Dec. 2007)**
7. **Derive the expressions for BER and error function. (Nov 2012)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 282**(Nov. / Dec. 2007)**
8. **Explain the types of preamplifiers used in a receiver.**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 282**(Nov. / Dec. 2007)**
9. **Define the terms- Quantum limit and probability of Error with respect to a receiver with typical values. (Nov 2013, 2018)**
   Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 282**(Nov. / Dec. 2007)**
10. **With suitable diagram, explain optical receiver operation and its performance.(May 2014, May 2015)**
    Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 279
11. **Describe the dispersion and numerical aperture measurements of fiber. (May 2014)**
    Refer Book: Optical fiber Communications - John M.Senior - Pg. No . 779 -781.
12. **With schematic diagram, explain the blocks and their functions of an optical receiver. (Apr-May 2015, Nov-Dec 2014, APR/MAY 2019)**
    Refer Book: Optical fiber Communications - John M.Senior - Pg. No . 779 -781.
13. **A digital fibre optic link operationg at 850nm requires a maximum BER of 10^{-9}. Find the quantum limit in terms of the quantum efficiency of the detector and the energy of the incident photon. (Apr-May 2015)**
    Refer Book: Optical fiber Communications - John M.Senior - Pg. No . 779 -781.
14. **Draw the block diagram of OTDR. Explain the measurements of any two fiber optic measurements with this. (Nov-Dec 2014)**
    Refer Book: Optical fiber Communications - John M.Senior - Pg. No . 779 -781.

UNIT-V OPTICAL NETWORKS

1. **Define power penalty [NOV/DEC 2018]**
   The power penalty may become significant if the semiconductor laser is biased above the threshold. For lasers biased below threshold, the extinction ratio is typically 0.05 and the power penalty is less than 0.4 dB. Timing Jitter In a digital system the signal is generally sampled at the center of the pulse.
2. **Distinguish between fundamental and higher order solitons.[APR/MAY 2019]**
   A fundamental soliton is an optical pulse which can propagate in a dispersive medium (e.g. an optical fiber) with a constant shape of the temporal intensity profile, i.e., without any temporal broadening as is usually caused by dispersion.
   A higher-order soliton is a soliton pulse the energy of which is higher than that of a fundamental soliton by a factor which is the square of an integer number
3. **Mention the drawbacks of Broadcast and select networks for wide area networkapplications.[Apr 2018]**
   The drawbacks of broadcast and select networks for wide area network applications are:
   More wavelengths are needed as the number of nodes in the network grows
   Without the use of optical booster amplifiers splitting losses occurs
4. **What are the three topologies used for fiber optical network? (N/D 2011)**
   The three topologies used for fiber optical network are:
   Bus
   Ring
   Star
5. **Calculate the number of independent signals that can be sent on a single fiber in the 1525-1565 nm bands. Assume the spectral spacing as per ITU-T recommendation G.692. (A/M 2011)**
   Given data: Mean frequency spacing as per ITU-T is 0.8 nm
   Wavelength = 1565 nm -1525 nm = 40 nm
   Solution:
   Number of independent channel = (40 nm/0.8 nm) = 50 Channels
6. **What is meant by power penalty?**
   When nonlinear effects contribute to signal impairment, an additional amount of power will be needed at the receiver to maintain the same BER. This additional power(dB) is known as the power penalty.
7. **Define – Network**
   Network is defined as to establish connections between these stations; one interconnects them by transmission paths to form a network.
8. **What is meant by topology?**
   The topology is the logical manner in which nodes are linked together by information transmission channels to form a network.
9. **What are the drawbacks of broadcast and select networks for wide area network applications? (M/J 2012)**
   The drawbacks of broadcast and select networks for wide area network applications are:
   More wavelengths are needed as the number of nodes in the network grows
   Without the use of optical booster amplifiers splitting losses occurs

10. Define – WDM (A/M2011)

In fiber-optic communications, wavelength-division multiplexing (WDM) is a technology which multiplexes a number of optical carrier signals onto a single optical fiber by using different wavelengths (i.e. colors) of laser light. This technique enables bidirectional communications over one strand of fiber, as well as multiplication of capacity.

11. **What are the advantages of WDM? (N/D2007)**

The advantages of WDM are

- Various optical channels can support different transmission formats
- Increase in the capacity of optical fiber compared to point-to-point link

12. What is the purpose of rise-time budget analysis? (A/M2008)

Rise-time budget ensures that the link is able to operate for a given data rate at specified BER. All the components in the link must operate fast enough to meet the band width or rise time requirements.

13. The specifications of the light sources are converted to equivalent rise time in rise time budget. Why?

A rise time budget is a convenient method to determine the dispersion limitation of an optical link. This is particularly useful for digital systems. For this purpose, the specifications of the light sources (both the fiber and the photo detector) are converted to equivalent rise time. The overall system rise time is given in terms of the light source rise time, fiber dispersion time and the photo detector rise time.

14. What is EDFA?(A/M2008, 2019), (M/J2012)NOV /DEC2018

An erbium-doped fiber amplifier (EDFA) is a device that amplifies an optical fiber signal. A trace impurity in the form of a trivalent erbium ion is inserted into the optical fiber's silica core to alter its optical properties and permit signal amplification.

15. What are the two different types of WDM? (DEC2016)

The two different types of WDM are

a. Unidirectional WDM
b. Bidirectional WDM

15. Define – Crosstalk

Crosstalk is defined as the feed through one of the channel signals into another channel.

16. Give the important features of time-slotted optical TDM network.

The important features of time slotted optical TDM network are

c. To provide backbone to interconnect high speed networks
d. To transfer quickly very large data blocks
e. To switch large aggregations of traffic
f. To provide both high- rate.

17. How the speckle pattern can form?

The speckle patterns are formed by the interference of the modes from a coherent source when the coherence time of the source is greater than the intermodal dispersion time within the fiber.

18. Define – Full- Width Half- Maximum(FWHM)

The FWHM is a pulse defined as the full width at its half-maximum power level.

19. **What are the advantages of using soliton signals through fiber? (M/J2009)**

The advantages of using soliton signals through fiber are, it is very narrow, high-intensity optical pulses that retain their shape through the interaction of balancing pulse dispersion with the nonlinear properties of an optical fiber

20. **What is chirping? (N/D2009)**

The d.c. modulation of a single longitudinal mode semiconductor laser can cause a dynamic shift of the peak wavelength emitted from the device This phenomenon, which results in dynamic line width broadening under the direct modulation of the injection current, is referred to as frequency chirping.

21. What is the best way to minimize chirping?

It is to choose the LASER emission wavelength close to the zero-dispersion of the wavelength of the fiber.

22. What do you mean by bidirectional WDM?

A single WDM which operates as both multiplexing and demultiplexing device is said to be bidirectional WDM.

23. What are the basic performances of the WDM?

The basic performances of WDM are

- ✓ Insertion loss
- ✓ Channel width
- ✓ Cross talk

25. Distinguish between fundamental and higher order soliton. (N/D2007)

The optical pulse that does not change in shape is called fundamental solitons.

The pulses that undergo periodic shape changes are called higher order solitons.

26. What are the two different types of WDM? (MAY 2016)

The two different types of WDM are

g. Unidirectional WDM

h. Bidirectional WDM

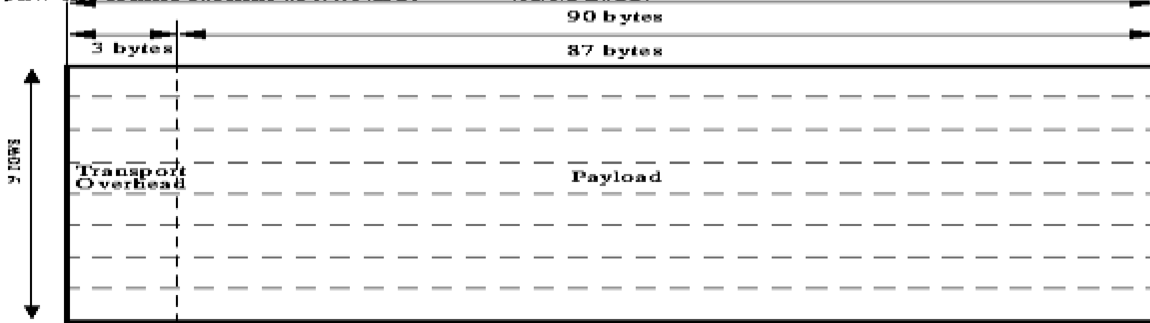
27. What is DWDM?

Dense Wavelength Division Multiplexing (DWDM) is an optical technology used to increased bandwidth over existing fiber-optic bones. It works by combining and transmitting multiple signals simultaneously at different wavelengths on the same fibers.

28. What is SONET/SDH?[Apr 2018]

Synchronous Optical NETworking (SONET) or Synchronous Digital Hierarchy (SDH) is a standardized protocol that transfers multiple digital bit streams over optical fiber using lasers or highly coherent light from light emitting diodes. At low transmission rates data can also be transferred via an electrical interface.

29. Draw the frame format of SONET. (A/M 2011)



30. **What are the drawbacks of broadcast and select network for wavelength multiplexing?**

The problems that arise in broadcast and select networks are:

- More wavelengths are needed as the number of nodes in the network grows.
- Without the width spread use of optical booster amplifiers, due to this splitting loss is high.

31. What is optical CDMA? (Nov-Dec 2015)

Optical CDMA is a multiple access technique in which each user is assigned an unique optical code. When a receiver is placed anywhere on the network with a bar code that matches a transmitter that signal line is decoded and extracted from the network.The codes are orthogonal to each other.

32. Distinguish SONET and SDH. (Nov-Dec 2015)

SONET	SDH
It means synchronous optical network developed by ANSI. Basic signaling unit is OC-I (51.84Mbps) SONET uses the term section, line and path.	It means synchronous digital hierarchy developed by ITU Basic signaling unit is STM-1 (155.52 Mbps) SDH uses the term path, multiplex section and regenerator section.

33. Name two popular architectures of SONET/SDH network.(Nov-Dec 2016) (R)

The two popular architectures of SONET/SDH networks are:

- UPSR - Unidirectional Path Switched Ring, two-fiber.
- BLSR – Bidirectional Line Switched Ring, two-fiber or four-fiber.

34. Obtain the transmission bit rate of the basic SONET frame in Mbps.(Nov-Dec 2013) (E)

STS-1 frame rate = (810 bytes/frame)*(8000 frames/sec)

= 51.840 Mbps.

35. Illustrate inter-channel cross talk that occurs in a WDM system.(Nov- Dec 2013) (A)

Inter-channel crosstalk arises when an interfacing signal comes from a neighboring channel that operates at a different wavelength. This nominally occurs when a wavelength selecting device imperfectly rejects or isolates the signals from other near-by wavelength channels.

36. What is a broadcast and select network?(May-June 2013) (R)

In broadcast and select networks, a node sends its transmission to the star coupler on the available wavelength using a laser which produces an optical information stream. The information stream from multiple sources is optically combined by the star and the signal and the signal power of each stream is equally spilt and forwarded to all the nodes on their receiver fiber.

37. What is SONET?(Apr-May 2015) (R)

SONET means synchronous optical network which is developed by ANSI, standardized protocol that transfer multiple digital bit stream synchronously over optical fiber using laser.

38. What were the problems associated with PDH networks?(Nov-Dec 2012) (AZ)

1. PDH- Plesiochronous Digital Hierarchy
2. It is difficult to "pick out" (drop) a low bit rate stream out of a high bit rate stream it is completely demultiplexing stream**.**
3. Expensive and compromises network reliability.

39. Enumerate the various SONET/SDH layers. ?(Nov-Dec 2012)

The various SONET/SDH layers are,

- Photonic layer
- Section layer
- Line layer
- Path layer.

40. What is DWDM?

Dense Wavelength Division Multiplexing (DWDM) is an optical technology used to increased bandwidth over existing fiber-optic bones. It works by combining and transmitting multiple signals simultaneously at different wavelengths on the same fibers.

41.What are solitons? (N/D2010) (DEC2016)

Solitons are nonlinear optical pulses that have the potential to support very high optical transmission rates of many terabits per second over long distances.

42. Give the important features of time-slotted optical TDM network.

The important features of time slotted optical TDM network are

- ✓ To provide backbone to interconnect high speed networks
- ✓ To transfer quickly very large data blocks
- ✓ To switch large aggregations of traffic
- ✓ To provide both high- rate.

43. How the speckle pattern can form?

The speckle patterns are formed by the interference of the modes from a coherent source when the coherence time of the source is greater than the intermodal dispersion time within the fiber.

44.What do you mean by bidirectional WDM?

A single WDM which operates as both multiplexing and demultiplexing device is said to be bidirectional WDM.

44. Define – Full- Width Half- Maximum(FWHM)

The FWHM is a pulse defined as the full width at its half-maximum power level.

45. What are the types of broadcast and select network?

The types of broadcast and select network are

- ✓ Single – hop networks
- ✓ Multi – hop networks

46. What is meant by cross- phase modulation (XPM)?

Cross- phase modulation, which converts power fluctuations in particular wavelength channel to phase fluctuations in the copropating channels.

48. Define self-healing rings.

The SONET/SDH rings are called **self-healing rings,** since the traffic flowing along a certain path can automatically be switched to an alternate path.

49. Mention the architectures for SONET/SDH networks.

- ✓ Two fiber unidirectional path switched ring.
- ✓ Two or four fiber bidirectional path switched ring.

50. Define single –hop network.

Single –hop network refers to networks where information transmitted in the form of light reaches its destination without being converted to an electrical form at any intermediate point.

og

PART B & C

1. With neat diagram, explain the elements of SONET infrastructure. (16) (May 2007) (MAY 2016)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. NO.472

2. Explain the principle and operation of Erbium doped fiber amplifiers with neat diagrams. (10) (Jan 2010)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 514- 516.

3. Describe the principle and performance of DT-WDMA protocol. (8)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 477- 482

4. Explain the architecture of SONET and discuss the nonlinear effects on network performance (Nov 2011)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. NO.472

5. Explain the principle of solitons and discuss the soliton parameters with necessary expressions (May 2012)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No . 506.

6. Write short notes on optical CDMA,WDM and EDFA performance (May 2012)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 426.

7. Describe the non-linear effects on network performance in detail.(8) (Nov 2012, 2018) [DEC 2016] MAY 2019

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 514- 516.

8. Explain the basics of optical CDMA systems. (8) (Nov 2012)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 514- 516.

9. (i) What is a four fiber BLSR ring in a SONET? Explain the reconfiguration of the same during node or fiber failure

(ii) What is broadcast and select multi hop network? Explain. (Nov 2013)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. NO.4 72

10. (i) Explain the following requirements for the design of an optically amplified WDM link: 1) Link band width

2) Optical power requirements for a specific BER.

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 477- 482

(ii) Write a note on solitons. (Nov 2013)

11. Expalin SONET layers and frame structure with diagram (May 2014) [DEC 2016,2018]

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. NO.4 72

12. Discuss in detail about the effect of noise on system performance.(Nov-Dec 2016)

13. Discuss the performance improvement of WDM and EDFA systems.(Nov- Dec 2015,Apr-May 2015, 2019, Nov-Dec 2014, NOV/DEC2018)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 426.

14. Discuss the non-linear effects on optical network performance.(Apr-May 2015, 2019, Nov-Dec 2012)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 514- 516.

15. Explain i)Optical CDMA ii)Optical Wavelength Routing Network.(Nov-Dec 2012)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 514- 516.

16. Discuss about Ultra High Capacity Networks. (Apr-May 2015,Nov-Dec 2014)

Refer Book: Optical fiber Communications - Gerd Kaiser - Pg. No. 514- 516.

UNIT –I

Introduction to Optical Fibers:

Historical Development

- Fiber optics deals with study of propagation of light through transparent dielectric wageguides. The fiber optics are used for transmission of data from point to point location. Fiber optic systems currently used most extensively as the transmission line between terrestrial hardwired systems.

- The carrier frequencies used in conventional systems had the limitations in handlinmg the volume and rate of the data transmission. The greater the carrier frequency larger the available bandwith and information carrying capacity.

First generation

- The first generation of lightwave systems uses GaAs semiconductor laser and operating region was near 0.8 μm. Other specifications of this generation are as under:

  i) Bit rate : 45 Mb/s

  ii) Repeater spacing : 10 km

Second generation

i) Bit rate : 100 Mb/s to 1.7 Gb/s

ii) Repeater spacing : 50 km

iii) Operation wavelength : 1.3 μm

iv) Semiconductor : In GaAsP

Third generation

i) Bit rate : 10 Gb/s

ii) Repeater spacing : 100 km

iii) Operating wavelength : 1.55 μm

Fourth generation

Fourth generation uses WDM technique.

Bit rate : 10 Tb/s

Repeater spacing : > 10,000 km

Operating wavelength : 1.45 to 1.62 μm

Fifth generation

Fifth generation uses Roman amplification technique and optical solitiors.

Bit rate : 40 - 160 Gb/s

Repeater spacing : 24000 km - 35000 km

Operating wavelength : 1.53 to 1.57 μm

Need of fiber optic communication

- Fiber optic communication system has emerged as most important communication system. Compared to traditional system because of following requirements :
  1. In long haul transmission system there is need of low loss transmission medium
  2. There is need of compact and least weight transmitters and receivers.
  3. There is need of increase dspan of transmission.
  4. There is need of increased bit rate-distrance product.
- A fiber optic communication system fulfills these requirements, hence most widely acception.

General Optical Fiber Communication System

- Basic block diagram of optical fiber communication system consists of following important blocks.
  1. Transmitter
  2. Information channel
  3. Receiver.

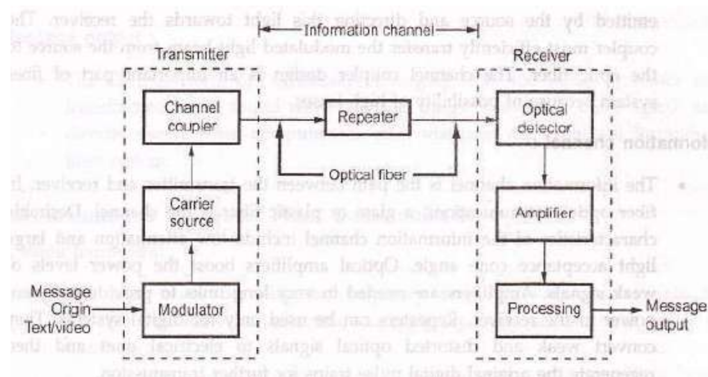


Fig. 1.2.1 shows block diagram of OFC system.

Message origin :

- Generally message origin is from a transducer that converts a non-electrical message into an electrical signal. Common examples include microphones

for converting sound waves into currents and video (TV) cameras for converting images into current. For data transfer between computers, the message is already in electrical form.

Modulator :

- The modulator has two main functions.

  1) It converts the electrical message into the proper format.

  2) It impresses this signal onto the wave generated by the carrier source.

  Two distinct categories of modulation are used i.e. analog modulation and digital modulation.

Carrier source :

- Carrier source generates the wave on which the information is transmitted. This wave is called the carrier. For fiber optic system, a laser diode (LD) or a light emitting diode (LED) is used. They can be called as optic oscillators, they provide stable, single frequency waves with sufficient power for long distance propagation.

Channel coupler :

- Coupler feeds the power into the information channel. For an atmospheric optic system, the channel coupler is a lens used for collimating the light emitted by the source and directing this light towards the receiver. The coupler must efficiently transfer the modulated light beam from the source to the optic fiber. The channel coupler design is an important part of fiber system because of possibility of high losses.

Information channel :

- The information channel is the path between the transmitter and receiver. In fiber optic communications, a glass or plastic fiber is the channel. Desirable characteristics of the information channel include low attenuation and large light acceptance cone angle. Optical amplifiers boost the power levels of weak signals. Amplifiers are needed in very long links to provide sufficient power to the receiver. Repeaters can be used only for digital systems. They convert weak and distorted optical signals to electrical ones and then regenerate the original disgital pulse trains for further transmission.

- Another important property of the information channel is the propagation time of the waes travelling along it. A signal propagating along a fiber normally contains a range of optic frequencies and divides its power along several ray paths. This results in a distortion of the propagating signal. In a digital system, this distortion appears as a spreading and deforming of the pulses. The spreading is so great that adjacent pulses begin to overlap and become unrecognizable as separate bits of information.

Optical detector :

- The information being transmitted is detector. In the fiber system the optic wave is converted into an electric current by a photodetector. The current developed by the

- detector is proportional to the power in the incident optic wave. Detector output current contains the transmitted information. This detector output is then filtered to remove the constant bias and thn amplified.

- The important properties of photodetectors are small size, economy, long life, low power consumption, high sensitivity to optic signals and fast response to quick variations in the optic power.

Signal processing :

- Signal processing includes filtering, amplification. Proper filtering maximizes the ratio of signal to unwanted power. For a digital system decision circuit is an additional block. The bit error rate (BER) should be very small for quality communications.

Message output :

- The electrical form of the message emerging from the signal processor are transformed into a soud wave or visual image. Sometimes these signals are directly usable when computers or other machines are connected through a fiber system.

Advantages of Optical Fiber Communications

1. Wide bandwidth

- The light wave occupies the frequency range between 2 x 1012 Hz to 3.7 x 1012 Hz. Thus the information carrying capability of fiber optic cables is much higher.

2. Low losses

- Fiber optic cables offers bery less signal attenuation over long distances. Typically it is less than 1 dB/km. This enables longer distance between repeaters.

3. Immune to cross talk

- Fiber optic cables has very high immunity to electricaland magnetic field. Since fiber optic cables are non-conductors of electricity hence they do not produce magnetic field. Thus fiber optic cables are immune to cross talk between cables cause dby magnetic induction.

4. Interference immune

- Fiber optic cable sar eimmune to conductive and radiative interferences caused by electrical noise sources such as lighting, electric motors, fluorescent lights.

5. Light weight

- As fiber cables are made of silica glass or plastic which is much lighter than copper or aluminium cables. Light weight fiber cables are cheaper to transport.

6. Small size

- The diameter of fiber is much smaller compared to other cables, therefore fiber calbe is small in size, requires less storage space.

7. More strength

- Fiber cables are stronger and rugged hence can support more weight.

8. Security

- Fiber cables are more secure than other cables. It is almost impossible to tap into a fiber cable as they do n ot radiate signals.

No ground loops exist between optical fibers hence they are more secure.

9. Long distance transmission

- Becaujse of less attenuation transmission at a longer distance is possible.

10. Environment immune

- Fiber calbes are more immune to environmental extremes. They can operate over a large temperature varioations. Also they are not affected by corrosive liquids and gases.

11. Sage and easy installation

- Fiber cables are safer and easier to install and maintain. They are non-conductors hence there is no shock hazards as no current or voltage is associated with them. Their small size and light weight feature makes installation easier.

12. Less cost

- Cost of fiber optic system is less compated to any other system.

Disadvantages of Optical Fiber Communicaitons

1. High initial cost

- The intial cost of installation or setting up cost is very high compared to all other system.

2. Maintenance and repaiding cost

- The maintenance and repaiding of fiber optic systems is not only difficult but expensive also.

3. Jointing and test procedures

- Since optical fibers are of very small size. The fiber joining process is very constly and requires skilled manpower.

4. Tensile stress

- Optical fibers are more susceptible to buckling, bending and tensile stress than copper cables. This leades to restricted practice to use optical fiber technology to premises and floor backbones with a few interfaces to the copper cables.

5. Short links

- Eventhough optical fiber calbes are inexpensive, it is still not cost effective to replace every small conventional connector (e.g. between computers and peripherals), as the price of optoelectronic transducers are very high.

6. Fiber losses

- The amount of optical fiber available to the photodetector at the end of fiber length depends on various fiber losses such as scattering, dispersion, attenuation and reflection.

Applications of Optical Fiber Communicaitons

- Applications of optical fiber communications include telecommunications, data communications, video control and protection switching, sensors and power applications.

1. Telephone networks

- Optical waveguide has low attenuation, high transmission bandwidth compated to copper lines, therefore numbers of long haul co-axial trunks l;links between telephone exchanges are being replaced by optical fiber links.

2. Urban broadband service networks

- Optical waveguide provides much larger bandwidth than co-axial calbe, also the number of repeaters required is reduced considerably.

- Modern suburban communications involves videotext, videoconferencing videotelephony, switched broadband communication network. All these can

be supplied over a single fiber optic link. Fiber optic calbes is the solution to many of today's high speed, high bandwidth data communication problems and will continue to play a large role in future telecom and data-com networks.

Optical Fiber Waveguides

- In free space light ravels as its maximum possible speed i.e. 3×10^8 m/s or 186×10^3 miles/sec. When light travels through a material it exnibits certain behavior explaned by laws of reflection, refraction.

Electromagnetic Spectrum

- The radio waves and light are electromagnetic waves. The rate at which they alternate in
  polarity is called their frequency (f) measured in hertz (Hz). The speed of electromagnetic wave (c) in free space is approximately 3×10^8 m/sec. The distance travelled during each cycle is called as wavelength (λ)

$$\text{Wavelength } (\lambda) = \frac{\text{Speed of light}}{\text{Frequency}} = \frac{c}{f}$$

- In fiber optics, it is more convenient to use the wavelength of light instead of the frequency with light frequencies, wavlengfth is often stated in microns or nanometers.

1 micron (μ) = 1

Micrometre (1×10^{-6}) 1

nano (n) = 10^{-9} metre

- Fiber optics uses visible and infrared light. Infrared light covers a fairly wide range of wavelengths and is generally used for all fiber optic communications. Visible light is normally used for very short range transmission using a plastic fiber.

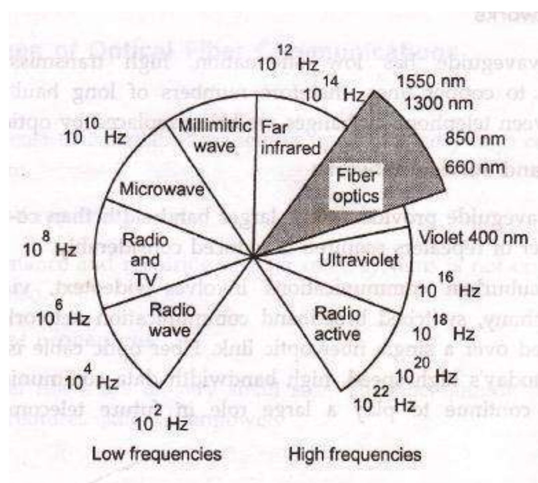


Fig. 1.6.1 shows electromagnetic frequency spectrum

Ray Transmission Theory

- Before studying how the light actually propagates through the fiber, laws governing the nature of light m ust be studied. These was called as **laws of optics (Ray theory).** There is conception that light always travels at the same speed. This fact is simply not true. The speed of light depends upon the material or medium through which it is moving. In free space light travels at its maximum possible speed i.e. 3 x 108 m/s or 186 x 103 miles/sec. When light travels through a material it exhibits certain behavior explained by laws of reflection, refraction.

Reflection

- The law of reflection states that, when a light ray is incident upon a reflective surface at some incident angle ϕ1 from imaginary perpendicular normal, the ray will be reflected from the surface at some angle ϕ2 from normal which is equal to the angle of incidence.

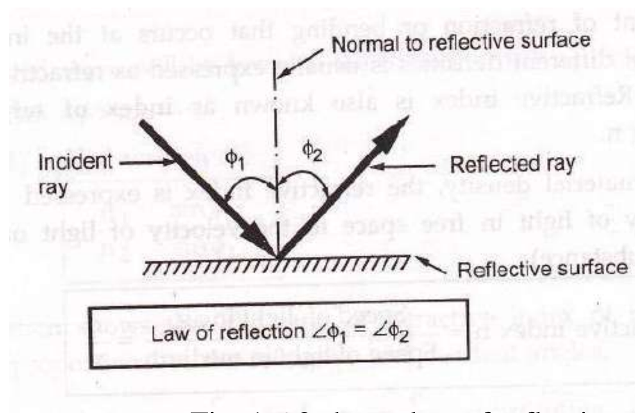


Fig. 1.6.2 shows law of reflection.

Refraction

- Refraction occurs when light ray passes from one medium to another i.e. the light ray changes its direction at interface. Refraction occurs whenever density of medium changes. E.g. refraction occurs at air and water interface, the straw in a glass of water will appear as it is bent.

The refraction can also observed at air and glass interface.

- When wave passes through less dense medium to more dense medium, the wave is refracted (bent) towards the normal. Fig. 1.6.3 shows the refraction phenomena.

- The refraction (bending) takes place because light travels at different speed in different mediums. The speed of light in free space is higher than in water or glass.

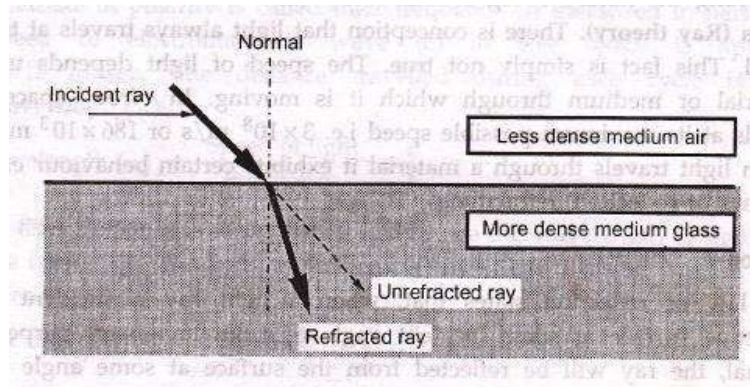


Fig.1.6.3 Refraction

Refractive Index

- The amount of refraction or bending that occurs at the interface of two materials of different densities is usually expressed as refractive index of two materials. Refractive index is also known as **index of refraction** and is denoted by n.
- Based on material density, the refractive index is expressed as the ratio of the velocity of light in free space to the velocity of light of the dielectric material (substance).

$$\text{Refractive index } n = \frac{\text{Speed of light in air}}{\text{Speed of light in medium}} = \frac{c}{v}$$

The refractive index for vacuum and air os 1.0 for water it is 1.3 and for glass refractive index is 1.5.

Snell's Law

- Snell's law states how light ray reacts when it meets the interface of two media having different indexes of refraction.
- Let the two medias have refractive indexes n1 and n2 where $n_1 > n_2$.

$\phi 1$ and $\phi 2$ be the angles of incidence and angle of refraction respectively. Then according to Snell's law, a relationship exists between the refractive index of both materials given by

$$n1 \sin\phi 1 = n2 \sin\phi 2 \qquad \dots (1.6.1)$$

- A refractive index model for Snell's law is shown in Fig. 1.6.4.
- The refracted wave will be towards the normal when $n_1 < n_2$ and will away from it when $n_1 > n_2$.

Equation (1.6.1) can be written as,

$$\frac{n_1}{n_2} = \frac{\sin \phi_2}{\sin \phi_1}$$

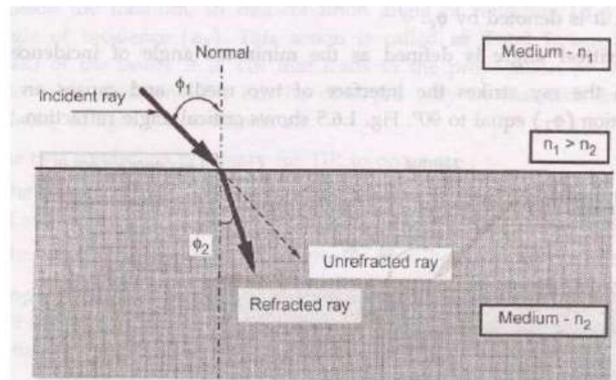


Fig 1.6.4 Refractive model for Snells Law

- This equation shows that the ratio of refractive index of two mediums is inversely proportional to the refractive and incident angles.

As refractive index $n_1 = \frac{c}{v_1}$ and $n_2 = \frac{c}{v_2}$ substituting these values in equation (1.6.2)

$$\frac{c / v_1}{c / v_2} = \frac{\sin \phi_2}{\sin \phi_1}$$

$$\frac{v_2}{v_1} = \frac{\sin \phi_2}{\sin \phi_1}$$

Critical Angle

- When the angle of incidence (ϕ_1) is profressively increased, there will be progressive increase of refractive angle (ϕ_2). At some condition (ϕ_1) the refractive angle (ϕ_2) becomes
  90o to the normal. When this happens the refracted light ray travels along the interface. The angle of incidence (ϕ_1) at the point at which the refractive angle (ϕ_1) becomes 90° is called the critical angle. It is denoted by ϕ_c.
- The **critical angle** is defined as the minimum angle of incidence (ϕ_1) at which the ray strikes the interface of two media and causes an agnle of refraction (ϕ_2) equal to 90°. Fig 1.6.5 shows critical angle refraction.

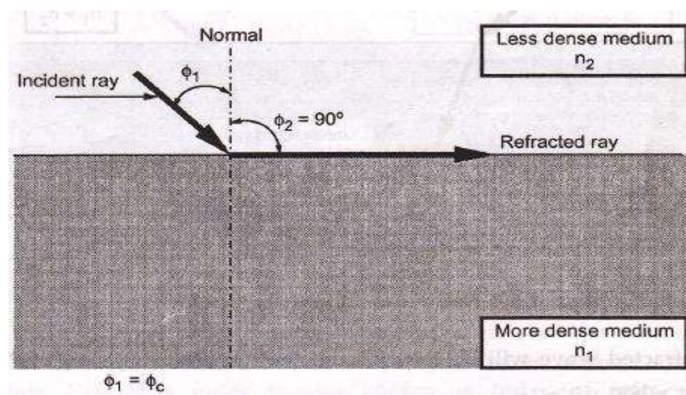


Fig.1.6.5 Critical Angle

Hence at critical angle $\phi_1 = \phi_c$ and $\phi_2 = 90^O$

Using Snell's law : $n_1 \sin \phi_1 = n_2 \sin \phi_2$

$$\sin \phi_c = \frac{n_2}{n_1} \sin 90^o$$

$\because$ $\sin 90^o = 1$

Therefore, $\sin \phi_c = \frac{n_2}{n_1}$

$$\text{Critical angle } \phi_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) \quad \ldots (1.6.3)$$

- The actual value of critical angle is dependent upon combination of materials present on each side of boundary.

Total Internal Refleciton (TIR)

- When the incident angle is increase dbeyond the critical angle, the light ray does not pass through the interface into the other medium. This gives the effect of mirror exist at the interface with no possibility of light escaping outside the medium. In this condition angle of reflection (ϕ_2) is equal to angle of incidence (ϕ_1). This action is called as **Total Internal Reflection (TIR)** of the beam. It is TIR that leads to the propagation of waves within fiber-cable medium. TIR can be observed only in materials in which the velocity of light is less than in air.

The refractive index of first medium must be greater than the refractive index of second one.

1. The angle of incidence must be greater than (or equal to) the critical angle.

Example 1.6.1 : A light ray is incident from medium-1 to medium-2. If the refractive indices of medium-1 and medium-2 are 1.5 and 1.36 respectively then determine the angle of refraction for an angle of incidence of 30°.

Solution : Medium-1 n1 = 1.5

Medium-2 n2 = 1.36

Angle of incidence $\phi_1 = 30^o$.

Angle of incident ϕ_2 = ?

$$\text{Snell's law}: n_1 \sin\phi_1 = n_2 \sin\phi_2$$

$$1.5 \sin 30^o = 1.36 \sin\phi_2$$

$$\sin\phi_2 = \frac{1.5}{1.36}\sin 30^o$$

$$\sin\phi_2 = 0.55147$$

$$\therefore \qquad \phi_2 = 33.46^o$$

Angle of refraction 33.46^o from normal. **... Ans.**

Example 1.6.2 : A light ray is incident from glass to air. Calculate the critical angle (ϕ_c).

Solution : Refractive index of glass $n_1 = 1.50$

Refrative indes of air $n_2 = 1.00$

$$\text{Snell's law}: n_1 \sin\phi_1 = n_2 \sin\phi_2$$

$$\sin\phi_1 = \frac{n_2}{n_1}\sin\phi_2$$

$$\therefore \qquad \sin\phi_1 = \frac{n_2}{n_1}\sin 90^o$$

Example 1.6.3 : Calculate the NA, acceptance angle and critical angle of the fiber having n_1 (Core refractive index) = 1.50 and refractive index of cladding = 1.45.

Soluiton : $n_1 = 1.50$, $n_2 = 1.45$

$$\Delta = \frac{(n_1 - n_2)}{(n_1)} = \frac{1.50 - 1.45}{1.50} = 0.033$$

$$\text{Numerical aperture, NA} = n_1 \sqrt{2\Delta}$$

$$NA = 1.50\sqrt{2 \times 0.033}$$

$$NA = 0.387$$

$$\text{Acceptance angle } \phi_0 = \sin^{-1} NA \qquad \phi_0 = \sin^{-1} 0.387$$

$$\text{Critical angle } \phi_c = \sin^{-1}\frac{n_2}{n_1} \qquad \phi_c = \sin^{-1}\frac{1.45}{1.50}$$

$$\phi_0 = 22.78^o$$

Optical Fiver as Waveguide

- An optical fiber is a cylindrical dielectric waveguide capable of conveying electromagnetic waves at optical frequencies. The electromagnetic energy is in the form of the light and propagates along the axis of the fiber. The structural of the fiver determines the transmission characteristics.
- The propagation of light along the waveguide is decided by the modes of the waveguides, here mode means path. Each mode has distict pattern of electric and magnetic field distributions along the fiber length. Only few modes can satisfy the homogeneous wave

  equation in the fiver also the boundary condition a waveguide surfaces. When there is only one path for light to follow then it is called as single mode propagation. When there is more than one path then it is called as multimode propagation.

Single fiber structure

- A single fiber structure is shown in Fig. 1.6.6. It consists of a solid dielectric cylinder with radius ‘a’. This cylinder is called as **core** of fiber. The core is surrounded by dielectric, called **cladding**. The index of refraction of core (glass fiber) is slightly greater than the index of refraction of cladding.

If refractive index of core (glass fiver) = n_1

and refractive index of cladding = n_2

then $n_1 > n_2$.

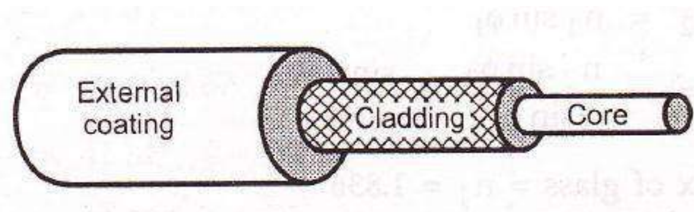


Fig.1.6.6. Single optical Fibre Structure

Propagation in Optical Fiber

- To understand the general nature of light wave propagation in optical fiber. We first consider the construction of optical fiber. The innermost is the glass core of very thin diameter with a slight lower refractive index n2. The light wave can propagate along such a optical fiber. A single mode propagation is illustrated in Fig. 1.6.7 along with standard size of fiber.

Single mode fibers are capable of carrying only one signal of a specific wavelength.

- In multimode propagation the light propagates along the fiber in zigzag fashion, provided it can undergo total internal reflection (TIR) at the core cladding boundaries.
- Total internal reflection at the fiber wall can occur only if two conditions are satisfied.

Condition 1:

The index of refraction of glass fiber must be slightly greater than the index of refraction of material surrounding the fiber (cladding).

If refractive index of glass fiber = n_1

and refractive index of cladding = n_2

then $n_1 > n_2$.

Condition 2 :

The angle of incidence (ϕ_1) of light ray must be greater than critical angle (ϕ_c).

- A light beam is focused at one end of cable. The light enters the fibers at different angles.
  Fig. 1.6.8 shows the conditions exist at the launching end of optic fiber. The light source is surrounded by air and the refractive index of air is n_0 = 1. Let the incident ray makes an angle ϕ_0 with fiber axis. The ray enters into glass fiber at point P making refracted angle ϕ_1 to the fiber axis, the ray is then propagated diagonally down the core and reflect from the core wall at point Q. When the light ray reflects off the inner surface, the angle of incidence is equal to the angle of reflection, which is greater than critical angle.
- In order for a ray of light to propagate down the cable, it must strike the core cladding interface at an angle that is greater than critical angle (ϕ_c).

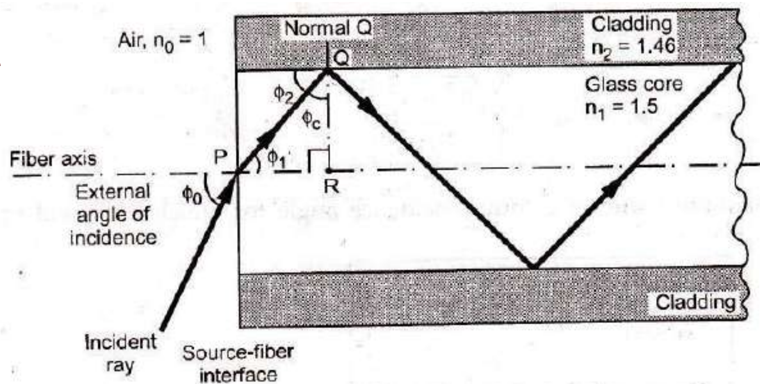


Fig. 1.6.8 Ray propagation by TIR

Acceptance Angle

Applying Snell's law to external incidence angle.

$$n_0 \sin \phi_0 = n_1 \sin \phi_1$$

But $\phi_1 = (90 - \phi_c)$

$$\sin \phi_1 = \text{sing } (90 - \phi_c) = \cos \phi_c$$

Substituting $\sin \phi_1$ in above equation.

$$n_0 \sin \phi_0 = n_1 \cos \phi_c$$

$$\sin \phi_c = \frac{n_1}{n_0} \cos \phi_c$$

Applying Pythagorean theorem to ΔPQR.

$$\cos \phi = \frac{\sqrt{n_1^2 - n_2^2}}{}$$

The maximum value of external incidence angle for which light will propagate in the fiber.

$$\phi_{0(max)} = \sin^{-1}\left[\frac{\sqrt{n_1^2 - n_2^2}}{n_0}\right]$$

When the light rays enters the fivers from an air medium $n_0 = 1$. Then above equation reduces to,

$$\phi_{0(max)} = \sin^{-1}\left(\sqrt{n_1^2 - n_2^2}\right)$$

The angle ϕ_0 is called as **acceptance angle** and $\phi_{0(max)}$ defines the maximum angle in which the light ray may incident on fiber to propagate down the fiber.

Acceptance Cone

- Rotating the acceptance angle $\phi_{0(max)}$ around the fiber axis, a cone shaped pattern is obtained, it is called as **acceptance cone** of the fiber input. Fig 1.6.10 shows formation of acceptance cone of a fiber cable.

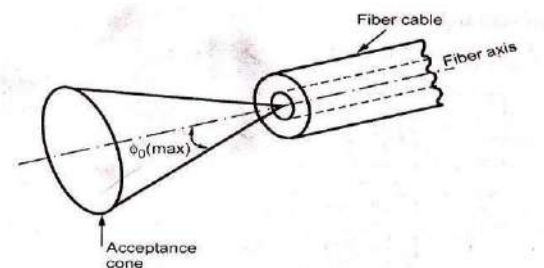


FIG: 1.6.10 shows formation of acceptance cone of a fiber cable.

- The Cone of acceptance is the angle within which the light is accepted into the core and is able to travel along the fiber. The launching of light wave becomes easier for large acceptance come.
- The angle is measured from the axis of the positive cone so the total angle of convergence is actually twice the stated value.

Numerical Aperture (NA)

- The **numerical aperture** (NA) of a fiber is a figure of merit which represents its light gathering capability. Larger the numerical aperture, the greater the

amount of light accepted by fiber. The acceptance angle also determines how much light is able to be enter the fiber and hence there is relation between the numerical aperture and the cone of acceptance.

Numerical aperture (NA) = sin $\phi_{0(max)}$

$$NA = \frac{\sqrt{n_1^2 - n_2^2}}{n_0}$$

For air $n_o = 1$

$$\therefore \quad NA = \sqrt{n_1^2 - n_2^2}$$

$$NA = \sqrt{n_{core}^2 - n_{cladding}^2} \quad \ldots (1.6.4)$$

Hence acceptance angle = $\sin^{-1}$ NA

By the formula of NA note that the numerical aperture is effectively dependent only on refractive indices of core and cladding material. NA is not a function of fiber dimension.

- The index difference (Δ) and the numerical aperture (NA) are related to the core and cladding indices:

$$\Delta = \frac{(n_1 - n_2)}{n_1} \qquad \Delta = \frac{NA^2}{2n_1^2}$$

Also $$NA = \sqrt{n_1^2 - n_2^2}$$

Example 1.6.5 : Calculate the numerical aperture and acceptance angle for a fiber cable of which $n_{core} = 1.5$ and $n_{cladding} = 1.48$. The launching takes place from air.

Solution :

$$NA = (n_1^2 - n_2^2)^{1/2}$$

$$NA = \sqrt{n_{core}^2 - n_{cladding}^2}$$

$$NA = n_1 (2\Delta)^{1/2}$$

$$NA = \sqrt{1.5^2 - 1.48^2}$$

NA = **0.244** **…Ans.**

Acceptance angle =

$$\sin^{-1}\sqrt{n_{core}^2 - n_{cladding}^2} = \sin^{-1} NA$$

Acceptance angle = $\sin^{-1}$ 0.244

$\phi_0 = \mathbf{14.12^0}$ **…Ans.**

Types of Rays

- If the rays are launched within core of acceptance can be successfully propagated along the fiber. But the exact path of the ray is determined by the position and angle of ray at which it strikes the core.
  There exists three different types of rays.
  i) Skew rays ii) Meridional rays iii) Axial rays.

- **The skew rays** does not pass through the center, as show in Fig. 1.6.11 (a). The skew rays reflects off from the core cladding boundaries and again bounces around the outside of the core. It takes somewhat similar shape of spiral of helical path.

Fig:1.6.11 Different Ray Propagation

- The **meridional** ray enters the core and passes through its axis. When the core surface is parallel, it will always be reflected to pass through the enter. The meridional ray is shown in fig. 1.6.11 (b).
- The **axial ray** travels along the axis of the fiber and stays at the axis all the time. It is shown in fig. 1.6.11 (c).

Modes of Fiber

- Fiber cables cal also be classified as per their mode. Light rays propagate as an electromagnetic wave along the fiber. The two components, the electric field and the magnetic field form patterns across the fiber. These patterns are called **modes** of transmission. The **mode** of a fiber refers to the number of paths for the light rays within the cable. According to modes optic fibers can be classified into two types.
  i) Single mode fiber ii) Multimode fiber.

- Multimode fiber was the first fiber type to be manufactured and commercialized. The term multimode simply refers to the fact that numerous modes (light rays) are carried simultaneously through the waveguide. Multimode fiber has a much larger diameter, compared to single mode fiber, this allows large number of modes.
- Single mode fiber allows propagation to light ray by only one path. Single mode fibers are best at retaining the fidelity of each light pulse over longer distance also they do not exhibit dispersion caused by multiple modes.

Thus more information can be transmitted per unit of time.

This gives single mode fiber higher bandwidth compared to multimode fiber.

- Some disadvantages of single mode fiber are smaller core diameter makes coupling light into the core more difficult. Precision required for single mode connectors and splices are more demanding.

Fiber Profiles

- A fiber is characterized by its profile and by its core and cladding diameters.
- One way of classifying the fiber cables is according to the index profile at fiber. The **index profile** is a graphical representation of value of refractive index across the core diameter.
- There are two basic types of index profiles.

i) Step index fiber. ii) Graded index fiber.

Fig. 1.6.12 shows the index profiles of fibers.

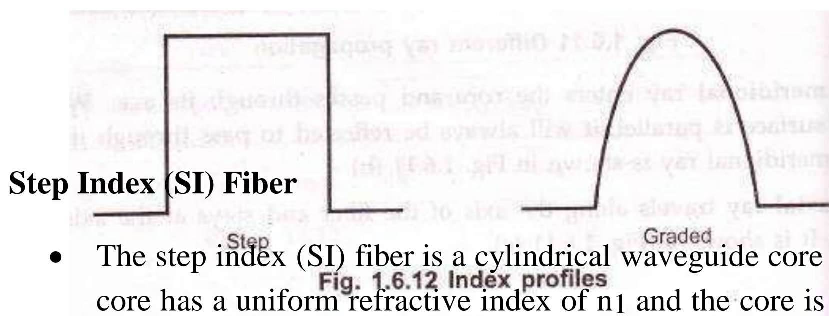


Fig. 1.6.12 Index profiles

Step Index (SI) Fiber

- The step index (SI) fiber is a cylindrical waveguide core with central or inner core has a uniform refractive index of n_1 and the core is surrounded by outer cladding with uniform refractive index of n_2. The cladding refractive index (n_2) is less than the core refractive index (n_1). But there is an abrupt change in the refractive index at the core cladding interface. Refractive index profile of step indexed optical fiber is shown in Fig. 1.6.13. The refractive index is plotted on horizontal axis and radial distance from the core is plotted on vertical axis.

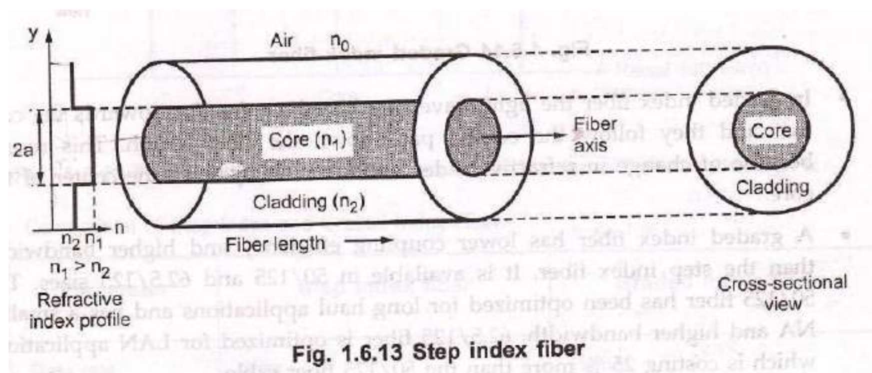


Fig. 1.6.13 Step index fiber

- The propagation of light wave within the core of step index fiber takes the path of meridional ray i.e. ray follows a zig-zag path of straight line segments.
  The core typically has diameter of 50-80 μm and the cladding has a diameter of 125 μm.
- The refractive index profile is defined as –

$$n(r) = \begin{cases} n_1 & \text{when } r < a \text{ (core)} \\ n_2 & \text{when } r \geq a \text{ (cladding)} \end{cases}$$

Graded Index (GRIN) Fiber

- The graded index fiber has a core made from many layers of glass.
- In the **graded index (GRIN)** fiber the refractive index is not uniform within the core, it is highest at the center and decreases smoothly and continuously with distance towards the cladding. The refractive index profile across the core takes the parabolic nature. Fig. 1.6.14 shows refractive index profile of graded index fiber.

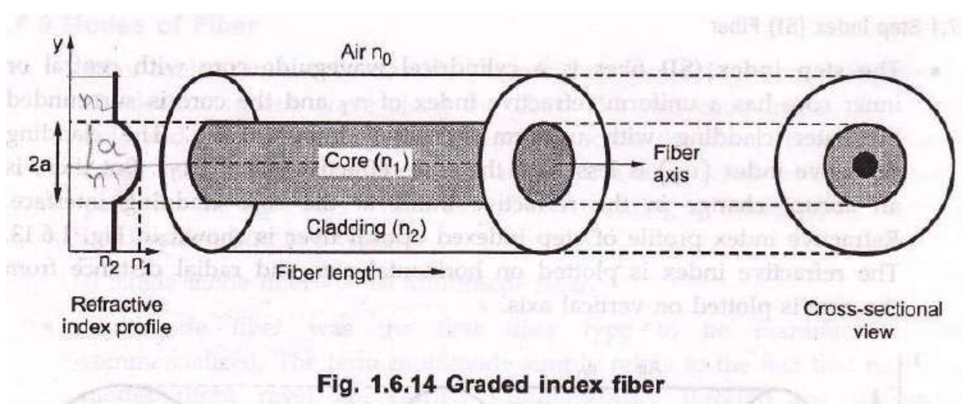


Fig. 1.6.14 Graded index fiber

- In graded index fiber the light waves are bent by refraction towards the core axis and they follow the curved path down the fiber length. This results because of change in refractive index as moved away from the center of the core.
- A graded index fiber has lower coupling efficiency and higher bandwidth than the step index fiber. It is available in 50/125 and 62.5/125 sizes. The 50/125 fiber has been optimized for long haul applications and has a smaller NA and higher bandwidth. 62.5/125 fiber is optimized for LAN applications which is costing 25% more than the 50/125 fiber cable.
- The refractive index variation in the core is giver by relationship

$$n(r) = \begin{cases} n_1\left(1 - 2\Delta\left(\frac{r}{a}\right)^{\alpha}\right) & \text{when } r < a \text{ (core)} \\ n_1(1 - 2\Delta)^{\frac{1}{2}} \approx n_2 & \text{when } r \geq a \text{ (cladding)} \end{cases}$$

where,

r = Radial distance from fiber axis

a = Core radius

n_1= Refractive index of core

n_2 = Refractive index of cladding

α = Shape of index profile.

- Profile parameter α determines the characteristic refractive index profile of fiber core. The range of refractive index as variation of α is shown in Fig. 1.6.1

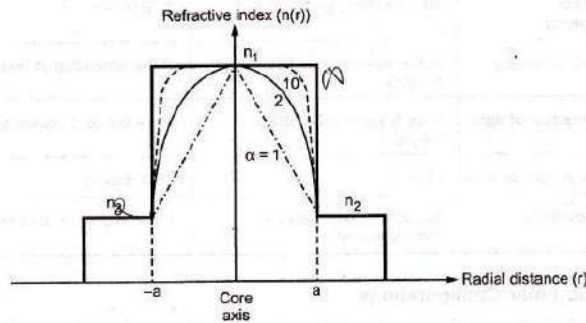


Fig. 1.6.15 Possible fiber refractive index profiles for different values of α

Comparison of Step Index and Graded Index Fiber

Sr. No.	Parameter	Step index fiber	Graded index fiber
1.	Data rate	Slow.	Higher
2.	Coupling efficiency	Coupling efficiency with fiber is higher.	Lower coupling efficiency.
3.	Ray path	By total internal reflection.	Light travelled oscillatory fashion.
4.	Index variation		
5.	Numerical aperture	NA remains same.	Changes continuously distance from fiber axis.
6.	Material used	Normally plastic or glass is preferred.	Only glass is preferred.
7.	Bandwidth efficiency	10 – 20 MHz/km	1 GHz/km
8.	Pulse spreading	Pulse spreading by fiber length is more.	Pulse spreading is less
9.	Attenuation	Less typically 0.34	More 0.6 to 1 dB/km at 1.3

	of light	dB/km at 1.3 μm.	μm.
10	Typical light source	LED.	LED, Lasers.
	Applications	Subscriber local network communicat ion.	networks.

Optic Fiber Configurations

- Depending on the refractive index profile of fiber and modes of fiber there exist three types of optical fiber configurations. These optic-fiber configurations are -

  i) Single mode step index fiber.
  ii) Multimode step index fiber.
  iii) Multimode graded index fiber.

Single mode Step index Fiber

- In single mode step index fiber has a central core that is sufficiently small so that there is essentially only one path for light ray through the cable. The light ray is propagated in the fiber through reflection. Typical core sizes are 2 to 15 μm. Single mode fiber is also known as fundamental or monomode fiber.

Fig. 1.6.16 shows single mode fiber.

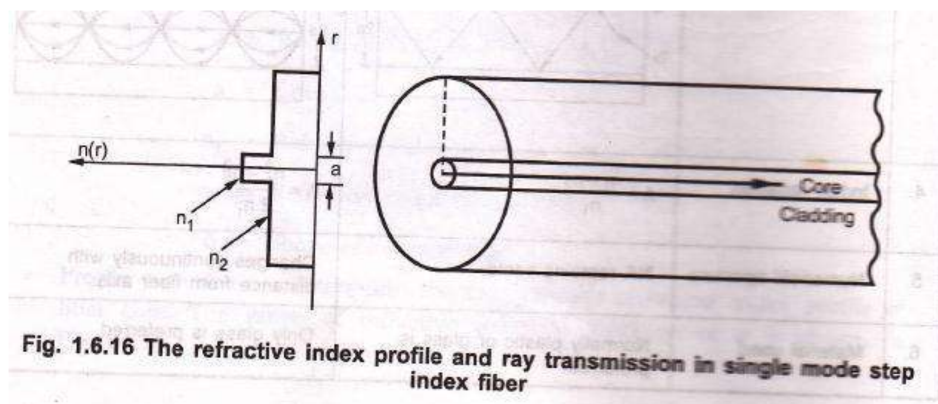


Fig. 1.6.16 The refractive index profile and ray transmission in single mode step index fiber

- Single mode fiber will permit only one mode to propagate and does not suffer from mode delay differences. These are primarily developed for the 1300 nm window but they can be also be used effectively with time division multiplex (TDM) and wavelength division multiplex (WDM) systems operating in 1550 nm wavelength region.
- The core fiber of a single mode fiber is very narrow compared to the wavelength of light being used. Therefore, only a single path exists through the cable core through which light can travel. Usually, 20 percent of the light in a single mode cable actually

travels down the cladding and the effective diameter of the cable is a blend of single mode core and degree to which the cladding carries light. This is referred to as the ‘mode field diameter’, which is larger than physical diameter of the core depending on the refractive indices of the core and cladding.

- The disadvantage of this type of cable is that because of extremely small size interconnection of cables and interfacing with source is difficult. Another disadvantage of single mode fibers is that as the refractive index of glass decreases with optical wavelength, the light velocity will also be wavelength dependent. Thus the light from an optical transmitter will have definite spectral width.

Multimode step Index Fiber

- **Multimode step index fiber** is more widely used type. It is easy to manufacture. Its core diameter is 50 to 1000 μm i.e. large aperture and allows more light to enter the cable. The light rays are propagated down the core in zig-zag manner. There are many many paths that a light ray may follow during the propagation.
- The light ray is propagated using the principle of total internal reflection (TIR). Since the core index of refraction is higher than the cladding index of refraction, the light enters at less than critical angle is guided along the fiber.

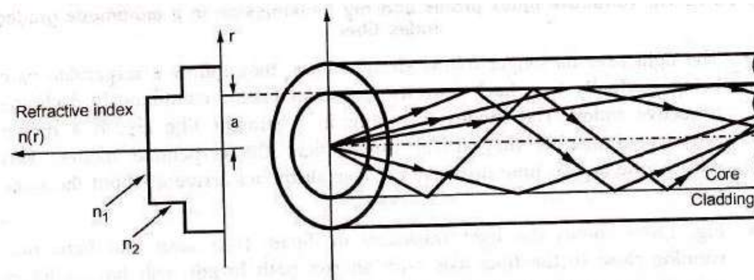


Fig. 1.6.17 TIR in multimode step index fiber

- Light rays passing through the fiber are continuously reflected off the glass cladding towards the centre of the core at different angles and lengths, limiting overall bandwidth.
- The disadvantage of multimode step index fibers is that the different optical lengths caused by various angles at which light is propagated relative to the core, causes the

  transmission bandwidth to be fairly small. Because of these limitations, multimode step index fiber is typically only used in applications requiring distances of less than 1 km.

Multimode Graded Index Fiber

- The core size of **multimode graded index fiber** cable is varying from 50 to 100 μm range. The light ray is propagated through the refraction. The light ray enters the fiber at

  many different angles. As the light propagates across the core toward the center it is intersecting a less dense to more dense medium. Therefore the light rays are being constantly being refracted and ray is bending continuously. This cable is mostly used for long distance communication.

Fig 1.6.18 shows multimode graded index fiber.

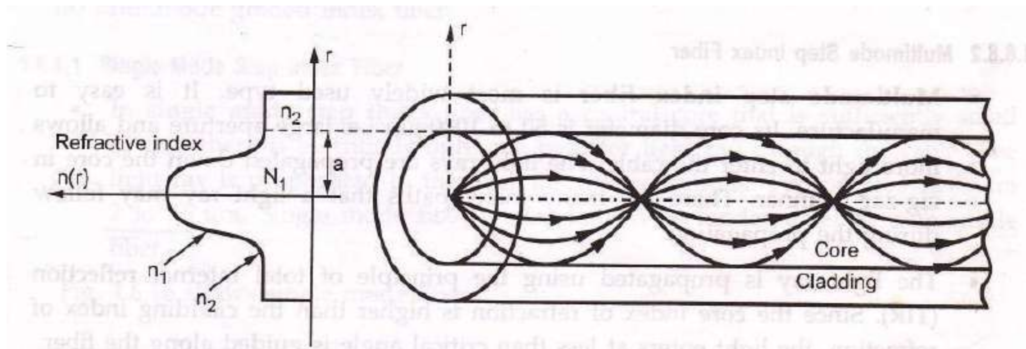


Fig. 1.6.18 The refractive index profile and ray transmission in a multimode graded index fiber

- The light rays no longer follow straight lines, they follow a serpentine path being gradually bent back towards the center by the continuously declining refractive index. The modes travelling in a straight line are in a higher refractive index so they travel slower than the serpentine modes. This reduces the arrival time disparity because all modes arrive at about the same time.
- Fig 1.6.19 shows the light trajectory in detail. It is seen that light rays running close to the fiber axis with shorter path length, will have a lower velocity because they pass through a region with a high refractive index.

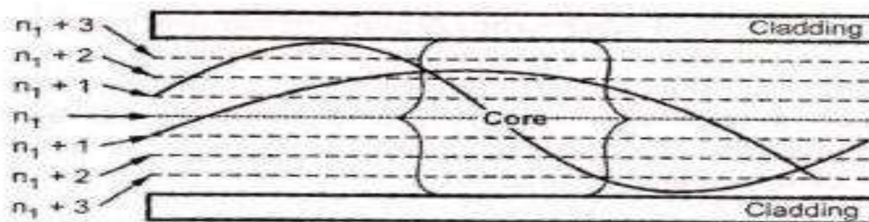


Fig. 1.6.19 Light trajectories in a graded index fiber

Rays on core edges offers reduced refractive index, hence travel more faster than axial rays and cause the light components to take same amount of time to travel the length of fiber, thus minimizing dispersion losses. Each path at a different angle is termed as
'transmission mode' and the NA of graded index fiber is defined as the maximum value of acceptance angle at the fiber axis.

- Typical attenuation coefficients of graded index fibers at 850 nm are 2.5 to 3 dB/km, while at 1300 nm they are 1.0 to 1.5 dB/km.
- The main advantages of graded index fiber are:

1. Reduced refractive index at the centre of core.
2. Comparatively cheap to produce.

Standard fibers

Sr. No.	Fiber type	Cladding Diameter (μm)	Core diameter (μm)		Applications
1.	Single mode (8/125)	125	8	0.1% to 0.2%	1. Long distance 2. High data rate
2.	Multimode (50/125)	125	50	1% to 2%	1. Short distance 2. Low data rate
3.	Multimode (62.5/125)	125	62.5	1% to 2%	LAN
4.	Multimode (100/140)	140	100	1% to 2%	LAN

Mode Theory for Cylindrical Waveguide

- To analyze the optical fiber propagation mechanism within a fiber, Maxwell equations are to solve subject to the cylindrical boundary conditions at core-cladding interface. The core-cladding boundary conditions lead to coupling of electric and magnetic field components resulting in hybrid modes. Hence the analysis of optical waveguide is more complex than metallic hollow waveguide analysis.
- Depending on the large E-field, the hybrid modes are HE or EH modes. The two lowest order does are HE_{11} and TE_{01}.

Overview of Modes

- The order states the number of field zeros across the guide. The electric fields are not completely confined within the core i.e. they do not go to zero at core-cladding interface and extends into the cladding. The low order mode confines the electric field near the axis of the fiber core and there is less penetration into the cladding. While the high order mode distribute the field towards the edge of the core fiber and penetrations into the cladding. Therefore cladding modes also appear resulting in power loss.
- In leaky modes the fields are confined partially in the fiber core attenuated as they propagate along the fiber length due to radiation and tunnel effect.

- Therefore in order to mode remain guided, the propagation factor β must satisfy the condition

$$n_2 k < \beta < n_1 k$$

where, n_1 = Refractive index of fiber core

n_2 = Refractive index of cladding

k = Propagation constant = 2π / λ

- The cladding is used to prevent scattering loss that results from core material discontinuities. Cladding also improves the mechanical strength of fiber core and reduces
  surface contamination. Plastic cladding is commonly used. Materials used for fabrication of optical fibers are silicon dioxide (SiO_2), boric oxide-silica.

Summary of Key Modal Concepts

- Normalized frequency variable, V is defined as

$$V = \frac{2\pi a (n_1^2 - n_2^2)^{1/2}}{\lambda} \quad \ldots (1.7.1)$$

where, a = Core radius

λ = Free space wavelength

$$V = \frac{2\pi a}{\lambda} NA$$

Since $(n_1^2 - n_2^2)^{1/2}$ = NA ... (1.7.2)

- The total number of modes in a multimode fiber is given by

$$M = \frac{1}{2}\left(\frac{2\pi a}{\lambda}\right)^2 (n_1^2 - n_2^2)$$

$$M = \frac{1}{2}\left[\frac{2\pi\, a}{\lambda} NA\right]^2 = \frac{[V]^2}{2}$$

$$M = \frac{1}{2}\left[\frac{\pi d}{\lambda} . NA\right]^2$$

Example 1.7.1 : Calculate the number of modes of an optical fiber having diameter of 50 μm, n_1 = 1.48, n_2 = 1.46 and λ = 0.82 μm.

Solution : d = 50 μm

n_1 = 1.48

n_2 = 1.46

$\lambda = 0.82$
μm

$$NA = (n_1^2 - n_2^2)^{1/2}$$

$$NA = (1.48^2 - 1.46^2)^{1/2}$$

$$NA = 0.243$$

Number of modes are given by,

$$M = \frac{1}{2}\left[\frac{\pi d}{\lambda}.NA\right]^2$$

$$M = \frac{1}{2}\left[\frac{\pi (50 X 10^{-6})}{0.82 X 10^{-6}} X 0.243\right]^2$$

$M = 1083$ …Ans.

Example 1.7.2 : A fiber has normalized frequency V = 26.6 and the operating wavelength is 1300nm. If the radius of the fiber core is 25 μm. Compute the numerical aperture.

Solution : $V = 26.6$
$\lambda = 1300$ nm $= 1300 X 10^{-9}$ m
$a = 25$ μm $= 25 X 10^{-6}$ m

$$V = \frac{2\pi a}{\lambda} NA$$

$$NA = V.\frac{\lambda}{2\pi a}$$

$$NA = 26.6 \frac{1300 X 10^{-9}}{2\pi X 25 X 10^{-6}}$$

NA = **0.220** **… Ans.**

Example 1.7.3 : A multimode step index fiber with a core diameter of 80 μm and a relative index difference of 1.5 % is operating at a wavelength of 0.85 μm. If the core refractive index is 1.48, estimate the normalized frequency for the fiber and number of guided modes.

[July/Aug.-2008, 6 Marks]

Solution : Given : MM step index fiber, 2 a = 80 μm

∴ Core radians a = 40 μm

Relative index difference, = 1.5% = 0.015

Wavelength, λ = 0.85μm

Core refractive index, n1 = 1.48

Normalized frequency, V = ?

Number of modes, M = ?

Numerical aperture

$$NA = n_1 (2\Delta)^{1/2}$$

$$= 1.48\ (2 \text{ X } 0.015)^{1/2}$$

$$= 0.2563$$

Normalized frequency is given by,

$$V = \frac{2\pi a}{\lambda} NA$$

$$V = \frac{2\pi \text{ X } 40}{0.85} \text{ x } 0.2563$$

$$V = 75.78$$

… Ans.

Number of modes is given by,

$$M = \frac{V^2}{2}$$

$$M = \frac{(75.78)^2}{2} = \mathbf{2871.50}$$

Ans

Example 1.7.4 : A step index multimode fiber with a numerical aperture of a 0.20 supports approximately 1000 modes at an 850 nm wavelength.

i) What is the diameter of its core?
ii) How many modes does the fiber support at 1320 nm?
iii) How many modes does the fiber support at 1550 nm? **[Jan./Feb.-2007, 10 Marks]**

Solution : i) Number of modes is given by,

$$M = \frac{1}{2}\left[\frac{\pi a}{\lambda}.NA\right]^2$$

a = **60.49 µm … Ans.**

ii)

$$M = \frac{1}{2}\left[\frac{\pi X 60.49 X 10^{-6}}{1320 X 10^{-9}} X 0.20\right]^2$$

M = $(14.39)^2$ = **207.07**

iii)

$$M = \frac{1}{2}\left[\frac{\pi X 6.49 X 10^{-6}}{1320 X 10^{-9}} X 0.20\right]^2$$

M = **300.63**

Wave Propagation

Maxwell's Equations

Maxwell's equation for non-conducting medium:

∇ X E = - ∂B /

∇ X H = - ∂D /

∇. D = 0

∇. B 0

where,

E and H are electric and magnetic field vectors.

- The relation between flux densities and filed vectors:

  $D = \varepsilon_0 E + P$

  $B = \mu_0 H + M$

where,

ε_0 is vacuum permittivity.

μ_0 is vacuum permeability.

P is induced electric polarization.

M is induced magnetic polarization (M = 0, for non-magnetic silica glass)

- P and E are related by:

$$P(r,t) = \varepsilon_0 \int_{-\infty}^{\infty} X(r, t - t') E(r, t') dt'$$

Where,

X is linear susceptibility.

- Wave equation:

$$\nabla \times \nabla \times E = \frac{-1}{c^2} \frac{\partial^2 E}{\partial E^2} - \mu_0 \frac{\partial^2 P}{\partial t^2}$$

Fourier transform of E (r, t)

$$\tilde{E}(r, \omega) = \int_{-\infty}^{\infty} E(r, t) e^{i\omega t} dt$$

$$\nabla \times \nabla \times \tilde{E} = -\varepsilon(r, \omega) \frac{\omega^2}{c^2} \tilde{E}$$

where,

$$\varepsilon = \left(n + \frac{i\alpha c}{2\omega}\right)^2$$

n is refractive index.

α is absorption coefficient.

$$n = \sqrt{(1 + R_e \chi)}$$

$$\alpha = \left(\frac{\omega}{nc}\right) I_m \chi$$

- Both n and α are frequency dependent. The frequency dependence of n is called as chromatic dispersion or material dispersion.
- For step index fiber,

$$\nabla \times \nabla \times \tilde{E} = \nabla(\nabla . \tilde{E}) - \nabla^2 . \tilde{E} = -\nabla^2 \tilde{E}$$

Fiber Modes

Optical mode : An optical mode is a specific solution of the wave equation that satisfies boundary conditions. There are three types of fiber modes.

a) Guided modes
b) Leaky modes
c) Radiation modes

- For fiber optic communication system guided mode is sued for signal transmission.

Considering a step index fiber with core radius 'a'.

The cylindrical co-ordinates ρ, ϕ and can be used to represent boundary conditions.

$$\frac{\partial^2 E_z}{\partial \rho^2} + \frac{1}{\rho} \cdot \frac{\partial E_z}{\partial \rho} + \frac{1}{\rho^2} \cdot \frac{\partial^2 E_z}{\partial \phi^2} + \frac{\partial^2 E_z}{\partial z^2} + n^2 k_0^2 E_z = 0$$

- The refractive index 'n' has values

$$n = \begin{cases} n_1; & \rho \leq a \\ n_2; & \rho > a \end{cases}$$

- The general solutions for boundary condition of optical field under guided mode is

infinite at $\rho = 0$ and decay to zero at $\rho = \infty$. Using Maxwell's equation in the core region.

$$E_\rho = \frac{i}{p^2}\left(\beta \frac{\partial E_z}{\partial \rho} + \mu_0 \frac{\omega}{\rho} \cdot \frac{\partial H_z}{\partial \phi}\right)$$

- The **cut-off condition** is defined as –

$$V = k_0 a \sqrt{(n_1^2 - n_2^2)}$$

$$V = \left(\frac{2\pi}{\lambda}\right) a\, n_1 \sqrt{2\Delta}$$

It is also called as **normalized frequency**.

Graded Index Fiber Structure

- The Refractive index of graded index fiber decreases continuously towards its radius from the fiber axis and that for cladding is constant.
- The refractive index variation in the core is usually designed by using power law relationship.

$$n(r) = \begin{cases} n_1\left[1 - 2\Delta\left(\frac{r}{a}\right)^{\alpha}\right]^{\frac{1}{2}}, & when\ 0 \leq r \leq a \\ n_1(1-2\Delta)^{\frac{1}{2}} \approx n_1(1-\Delta) = n_2, & when\ r \geq a \end{cases}$$

Where, r = Radial distance from fiber axis

a = Core radius

n_1 = Refractive index core

n_2 Refractive index of cladding and

α = The shape of the index profile

- For graded index fiber, the index difference is given by,

- In graded index fiber the incident light will propagate when local numerical aperture at distance r from axis, NA is axial numerical aperture NA(0). The local numerical aperture is given as,

$$NA(r) = \begin{cases} [n^2(r) - n_2^2]^{\frac{1}{2}} \approx NA(0)\sqrt{1 - \left(\frac{r}{a}\right)^{\alpha}}, & for\ r \leq a \\ 0, & for\ r > a \end{cases}$$

- The axial numerical aperture NA(0) is given as,

$$NA(0) = [n^2(0) - n_2^2]^{1/2}$$

$$NA(0) = [n_1^2 - n_2^2]^{1/2}$$

$$NA(0) = n_1\sqrt{2\Delta} \approx n_1(2\Delta)^{1/2}$$

Hence Na for graded index decreases to zero as it moves from fiber axis to core-cladding boundary.

- The variation of NA for different values of α is shown in Fig. 1.7.1.

- The number of modes for graded index fiber in given as,

$$M = \frac{\alpha}{\alpha + 2} a^2 k^2 n_1^2 \Delta$$

…

Single Mode Fibers

- Propagation in single mode fiber is advantageous because signal dispersion due to delay differences amongst various modes in multimode is avoided. Multimode step index fibers cannot be used for single mode propagation due to difficulties in maintaining single mode operation. Therefore for the

transmission of single mode the fiber is designed to allow propagation in one mode only, while all other modes are attenuated by leakage or absorption.

- For single mode operation, only fundamental LP_{01} mode many exist. The single mode propagation of LP_{01} mode in step index fibers is possible over the range.

$$0 \leq V < 2405$$

- The normalized frequency for the fiber can be adjusted within the range by reducing core radius and refractive index difference < 1%. In order to obtain single mode operation with maximum V number (2.4), the single mode fiber must have smaller core diameter than the equivalent multimode step index fiber. But smaller core diameter has problem of launching light into the fiber, jointing fibers and reduced relative index difference.

- Graded index fibers can also be sued for single mode operation with some special fiber design. The cut-off value of normalized frequency Vc in single mode operation for a graded index fiber is given by,

$$V_c = 2.405\left(1 + \frac{2}{\propto}\right)^{\frac{1}{2}}$$

Example 1.8.1 : A multimode step index optical fiber with relative refractive index difference 1.5% and core refractive index 1.48 is to be used for single mode operation. If the operating wavelength is 0.85µm calculate the maximum core diameter.

Solution : Given :

$n_1 = 1.48$
$\Delta = 1.5\ \% = 0.015$
$\lambda = 0.85\ \mu m = 0.85 \times 10^{-6}$ m

Maximum V value for a fiber which gives single mode operations is 2.4.

Normalized frequency (V number) and core diameter is related by expression,

$$V = \frac{2\pi}{\lambda} a\ (NA)$$

$$\boxed{V = \frac{2\pi}{\lambda} a\ n_1 (2\Delta)^{\frac{1}{2}}}$$

$$a = \frac{V\lambda}{2\pi\ n_1 (2\Delta)^{\frac{1}{2}}}$$

$$a = \frac{2.4 \times (0.85 \times 10^{-6})}{2\pi \times (1.48) \times (0.03)^{\frac{1}{2}}}$$

a = **1.3 µm** **… Ans.**

Maximum core diameter for single mode operation is 2.6 µm.

Example 1.8.2 : A GRIN fiber with parabolic refractive index profile core has a refractive index at the core axis of 1.5 and relative index difference at 1%. Calculate maximum possible core diameter that allows single mode operations at $\lambda = 1.3$ μm.

Solution : Given :

$$n_1 = 1.5$$

$$\Delta = 1\ \% = 0.01$$

$$\lambda = 1.3\ \mu m = 1.3 \times 10^{-6} m$$

for a GRIN

Maximum value of normalized frequency for single mode operation is given by,

$$V = 2.4\left(1 + \frac{2}{\alpha}\right)^{\frac{1}{2}}$$

Maximum core radius is given by expression,

$$a = \frac{V\lambda}{2\pi\, n_1 (2\Delta)^{\frac{1}{2}}}$$

$$a = \frac{24\sqrt{2} \times 1.3 \times 10^{-6}}{2\pi \times 1.5 \times (0.02)^{\frac{1}{2}}}$$

$a = 3.3$ μm **... Ans.**

∴ Maximum core diameter which allows single mode operation is 6.6 μm.

Cut-off Wavelength

- One important transmission parameter for single mode fiber us cut-off wavelength for the first higher order mode as it distinguishes the single mode and multim0de regions.
- The effective cut-off wavelength λ_C is defined as the largest wavelength at which higher order $(L_{P_{11}})$ mode power relative to the fundamental mode $(L_{P_{01}})$ power is reduced to 0.1 dB. The range of cut-off wavelength recommended to avoid modal noise and dispersion problems is : 1100 to 1280 nm (1.1 to 1.28μm) for single mode fiber at 1.3 μm.
- The cut-off wavelength λ_C can be computed from expression of normalized frequency.

$$V = \frac{2\pi}{\lambda} a\, n_1 (2\Delta)^{\frac{1}{2}} \Rightarrow \lambda = \frac{2\pi a n_1}{V}(2\Delta)^{\frac{1}{2}} \quad \text{.... (1.8.1)}$$

∴ .... (1.8.2)

$$\lambda = \frac{2\pi a n_1}{V}(2\Delta)^{\frac{1}{2}}$$

where,

V_C is cut-off normalized frequency.

- λ_C is the wavelength above which a particular fiber becomes single moded. For same fiber dividing λ_C by λ we get the relation as:

$$\frac{\lambda_c}{\lambda} = \frac{V}{V_C}$$

$$\lambda = \frac{V\lambda}{V_C} \qquad \ldots (1.8.3)$$

But for step index fiver $V_C = 2.405$ then

$$\lambda_c = \frac{V\lambda}{2.405}$$

Example 1.8.3 : Estimate cut-off wavelength for step index fiber in single mode operation. The core refractive index is 1.46 and core radius is 4.5 μm. The relative index difference is 0.25 %.

Solutions : Given :

$n_1 = 1.46$

$a = 4.5$ μm

$\Delta = 0.25\ \% = 0.0025$

Cut-off wavelength is given by,

$$\lambda_c = \frac{2\pi a n_1 (2\Delta)^{\frac{1}{2}}}{V_c}$$

For cut-off wavelength, $V_C = 2.405$

$$\lambda_c = \frac{2\pi \text{ x } 4.5 \text{ x } 1.46\ (0.005)^{\frac{1}{2}}}{2.405}$$

$$\lambda_c = 1.214\ \mu m$$

Mode Field Diameter and Spot Size

- The mode filed diameter is fundamental parameter of a single mode fiber. This parameter is determined from mode field distributions of fundamental LP_{01} mode.
- In step index and graded single mode fibers, the field amplitude distribution is

approximated by Gaussian distribution. The **mode Field diameter** (MFD) is distance between opposite 1/e – 0.37 times the near field strength)amplitude) and power is $1/e^2 = 0.135$ times.

- In single mode fiber for fundamental mode, on field amplitude distribution the mode filed diameter is shown in fig. 1.8.1.

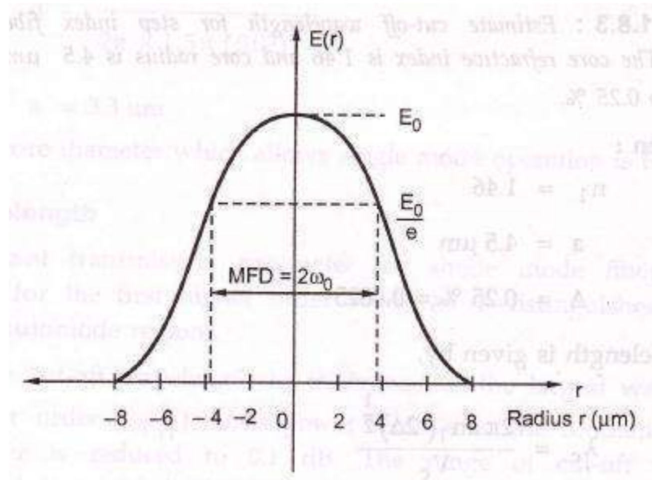


Fig. 1.8.1 Mode field diameter

- The spot size ω_0 is gives as –

$$\omega_0 = \frac{MFD}{2}$$

$$MFD = 2\,\omega_0$$

The parameter takes into account the wavelength dependent filed penetration into the cladding. Fig. 1.8.2 shows mode field diameters variation with λ.

QUESTIONS

1. State and explain the advantages and disadvantages of fiber optic communication systems.
2. State and explain in brief the principle of light propagation.
3. Define following terms with respect to optical laws –
   A) Reflection
   B) Refraction
   C) Refractive index
   D) Snell's law
   E) Critical angle
   F) Total internal reflection (TIR)
4. Explain the important conditions for TIR to exit in fiber.
5. Derive an expression for maximum acceptance angle of a fiber.
6. Explain the acceptance come of a fiber.
7. Define numerical aperture and state its significance also.
8. Explain the different types of rays in fiber optic.
9. Explain the
   A) Step index fiber
   B) Graded index fiber
10. What is mean by mode of a fiber?
11. Write short notes on following –
    A) Single mode step index fiber
    B) Multimode step index fiber
    C) Multimode graded index fiber.

UNIT - 2

SIGNAL DEGRADATION OPTICAL FIBERS.

Introduction

1. One of the important property of optical fiber is signal attenuation. It is also known as fiber loss or signal loss. The signal attenuation of fiber determines the maximum distance between transmitter and receiver. The attenuation also determines the number of repeaters required, maintaining repeater is a costly affair.
2. Another important property of optical fiber is distortion mechanism. As the signal pulse travels along the fiber length it becomes more broader. After sufficient length the broad pulses starts overlapping with adjacent pulses. This creates error in the receiver. Hence the distortion limits the information carrying capacity of fiber.

Attenuation

- Attenuation is a measure of decay of signal strength or loss of light power that occurs as light pulses propagate through the length of the fiber.
- In optical fibers the attenuation is mainly caused by two physical factors absorption and scattering losses. Absorption is because of fiber material and scattering due to structural imperfection within the fiber. Nearly 90 % of total attenuation is caused by Rayleigh scattering only. Microbending of optical fiber also contributes to the attenuation ofsignal.
- The rate at which light is absorbed is dependent on the wavelength of the light and the characteristics of particular glass. Glass is a silicon compound, by adding different additional chemicals to the basic silicon dioxide the optical properties of the glass can be changed.
- The Rayleigh scattering is wavelength dependent and reduces rapidly as the wavelength of the incident radiation increases.
- The attenuation of fiber is governed by the materials from which it is fabricated, the manufacturing process and the refractive index profile chosen. Attenuation loss is measured in dB/km.

Attenuation Units

- As attenuation leads to a loss of power along the fiber, the output power is significantly less than the couples power. Let the couples optical power is p(0) i.e. at origin (z = 0).

Then the power at distance z is given by,

$$P(z) = P(0)e^{-\alpha_p z} \quad \ldots (2.1.1)$$

where, α_p is fiber attenuation constant (per km).

$$\propto_p = \frac{1}{z} \ln \left[\frac{P(0)}{P(z)}\right]$$

$$\boxed{\propto_{dB/km} = 10 . \frac{1}{z} \log \left[\frac{P(0)}{P(z)}\right]}$$

$$\propto_{dB/km} = 4.343 \ \propto_p \ \text{per km}$$

This parameter is known as fiber loss or fiber attenuation.

- Attenuation is also a function of wavelength. Optical fiber wavelength as a function of wavelength is shown in Fig. 2.1.1.

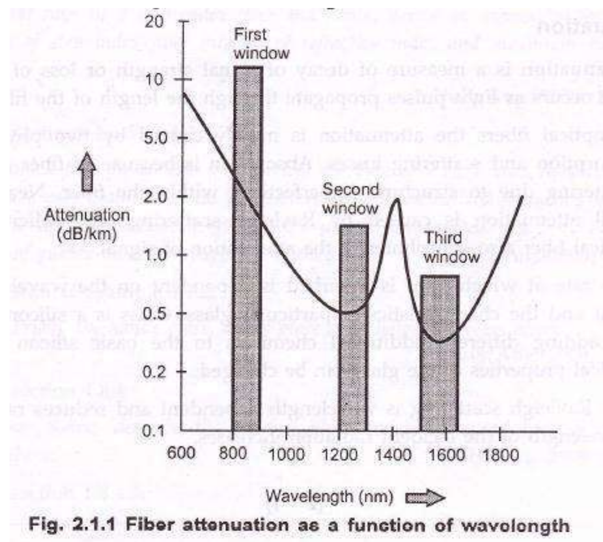


Fig. 2.1.1 Fiber attenuation as a function of wavolength

Example 2.1.1 : A low loss fiber has average loss of 3 dB/km at 900 nm. Compute the length over which –

a) Power decreases by 50 % b) Power decreases by 75 %.

Solution : $\alpha = 3$ dB/km

a) Power decreases by 50 %.

$\Rightarrow \quad \frac{p(0)}{p(z)} = 50\ \% = 0.5$

□ is given by,

$$\alpha = 10 \cdot \frac{1}{z} \log \left[\frac{P(0)}{P(z)}\right]$$

$$3 = 10 \cdot \frac{1}{z} \log [0.5]$$

∴

z = **1 km** **… Ans.**

b) $$\frac{P(0)}{P(z)} = 25\ \% = 0.25$$

Since power decrease by 75 %.

$$3 = 10 \text{ x } \frac{1}{z} \log[0.25]$$

∴ z = **2 km** **… Ans.**

Example 2.1.2 : For a 30 km long fiber attenuation 0.8 dB/km at 1300nm. If a 200 µwatt power is launched into the fiber, find the output power.

Solution : z = 30 km

□ = 0.8 dB/km

P(0) = 200 µW

Attenuation in optical fiber is given by,

$$\alpha = 10 \text{ x } \frac{1}{z} \log \left[\frac{P(0)}{\ }\right]$$

∴ $$P(z) = \left[\frac{200\ \mu W}{P(z)}\right] = 0.7962\ \mu W$$

$$0.8 = 10 \text{ x } \frac{1}{30} \log \left[\frac{200\ \mu W}{P(z)}\right]$$

$$2.4 = 10 \text{ x } \log \left[\frac{200\ \mu W}{P(z)}\right]$$

$$\left[\frac{200\ \mu W}{P(z)}\right] = 10^{2.4}$$

Example 2.1.3 : When mean optical power launched into an 8 km length of fiber is 12 μW, the mean optical power at the fiber output is 3 μW.

Determine –

- Overall signal attenuation in dB.
- The overall signal attenuation for a 10 km optical link using the same fiber with splices at 1 km intervals, each giving an attenuation of 1 dB.

Solution : **Given** : $z = 8$ km

$P(0) = 120$ μW

$P(z) = 3$ μW

1) Overall attenuation is given by,

$$\alpha = 10 \cdot \log\left[\frac{P(0)}{P(z)}\right]$$

$$\alpha = 10 \cdot \log\left[\frac{120}{3}\right]$$

$$\alpha = 16.02 \text{ dB}$$

2) Overall attenuation for 10 km,

Attenuation per km $\alpha_{dB} = \frac{16.02}{z} = \frac{16.02}{8} = 2.00 \text{ dB/km}$

Attenuation in 10 km link = 2.00 x 10 = 20 dB

In 10 km link there will be 9 splices at 1 km interval. Each splice introducing attenuation of 1 dB.

Total attenuation = 20 dB + 9 dB = **29 dB**

Example 2.1.4 : A continuous 12 km long optical fiber link has a loss of 1.5 dB/km.

- What is the minimum optical power level that must be launched into the fiber to maintain as optical power level of 0.3 μW at the receiving end?
- What is the required input power if the fiber has a loss of 2.5 dB/km?

[July/Aug.-2007, 6 Marks]

Solution : Given data : z = 12 km

= 1.5 dB/km

P(0) = 0.3 μW

- Attenuation in optical fiber is given by,

$$\alpha = 10 \text{ x } \frac{1}{z} \log\left(\frac{P(0)}{P(z)}\right)$$

$$1.5 = 10 \text{ x } \frac{1}{12} \log\left(\frac{0.3\ \mu W}{P(z)}\right)$$

$$\log\left(\frac{0.3\ \mu W}{P(z)}\right) = \frac{1.5}{0.833}$$

$$= 1.80$$

$$\left(\frac{0.3\ \mu W}{P(z)}\right) = 10^{1.8}$$

$$P(z) = \left(\frac{0.3\ \mu W}{10^{1.8}}\right) = \frac{0.3}{63.0}$$

$$P(z) = 4.76 \text{ x } 10^{-9} W$$

Optical power output = **$4.76 \text{ x } 10^{-9}$ W** **… Ans.**

ii) Input power = ? P(0)

When α = 2.5 dB/km

$$\boxed{\alpha = 10 \text{ x } \frac{1}{z} \log\left(\frac{P(0)}{P(z)}\right)}$$

$$2.5 = 10 \text{ x } \frac{1}{z} \log\left(\frac{P(0)}{4.76 \text{ x } 10^{-9}}\right)$$

$$\log\left(\frac{P(0)}{4.76 \text{ x } 10^{-9}}\right) = \frac{2.5}{0.833} = 3$$

$$\frac{P(0)}{4.76 \text{ x } 10^{-9}} = 10^3 = 1000$$

$\therefore$ $P(0) = 4.76\ \mu W$

Input power= **4.76 µW** **… Ans.**

Example 2.1.5 : Optical power launched into fiber at transmitter end is 150 µW. The power at the end of 10 km length of the link working in first windows is – 38.2 dBm. Another system of same length working in second window is 47.5 µW. Same length system working in third window has 50 % launched power. Calculate fiber attenuation for each case and mention wavelength of operation. **[Jan./Feb.-2009, 4 Marks]**

Solution : Given data:

$P(0) = 150\ \mu W$

$z = 10\ km$

$$P(z) = -38.2\ dBm \Rightarrow \begin{cases} -38.2 = 10 \log \dfrac{P(z)}{1\ mW} \\ P(z) = 0.151\ \mu W \end{cases}$$

$z = 10\ km$

$$\boxed{\alpha = 10 \times \frac{1}{z} \log \left[\frac{P(0)}{P(z)}\right]}$$

Attenuation in 1st window:

$$\alpha_1 = 10 \times \frac{1}{10} \log \left[\frac{150}{0.151}\right]$$

$$\boldsymbol{\alpha_1 = 2.99\ dB/km}$$ **… Ans.**

Attenuation in 2nd window:

$$\alpha_2 = 10 \times \frac{1}{10} \log \left[\frac{150}{47.5}\right]$$

$$\boldsymbol{\alpha_2 = 0.49\ dB/km}$$ **… Ans.**

Attenuation in 3rd window:

$$\alpha_3 = 10 \times \frac{1}{10} \log \left[\frac{150}{75}\right]$$

$\alpha_3 = 0.30$ dB/km **… Ans.**

Wavelength in 1st window is 850 nm.

Wavelength in 2nd window is 1300 nm.

Wavelength in 3rd window is 1550 nm.

Example 2.1.6 : The input power to an optical fiber is 2 mW while the power measured at the output end is 2 μW. If the fiber attenuation is 0.5 dB/km, calculate the length of the fiber.

[July/Aug.-2006, 6 Marks]

Solution : Given : $P(0) = 2$ mwatt $= 2 \times 10^{-3}$ watt

$P(z) = 2$ μwatt $= 2 \times 10^{-6}$ watt

$\alpha = 0.5$ dB/km

$$\boxed{\alpha = 10 \times \frac{1}{z}\left[\frac{P(0)}{P(z)}\right]}$$

$z = 60$ km **… Ans.**

$$0.5 = 10 \times \frac{1}{z}\log\left[\frac{2 \times 10^{-3}}{2 \times 10^{-6}}\right]$$

$$0.5 = \frac{1}{z} \times 3$$

$$z = \frac{3}{0.05}$$

Absorption

- Absorption loss is related to the material composition and fabrication process of fiber. Absorption loss results in dissipation of some optical power as hear in the fiber cable. Although glass fibers are extremely pure, some impurities still remain as residue after purification. The amount of absorption by these impurities depends on their concentration and light wavelength.

- Absorption is caused by three different mechanisms.
  Absorption by atomtic defects in glass composition.
  Extrinsic absorption by impurity atoms in glass matts.
  Intrinsic absorption by basic constituent atom of fiber.

Absorption by Atomic Defects

- Atomic defects are imperfections in the atomic structure of the fiber materials such as missing molecules, high density clusters of atom groups. These absorption losses are negligible compared with intrinsic and extrinsic losses.
- The absorption effect is most significant when fiber is exposed to ionizing radiation in nuclear reactor, medical therapies, space missions etc. The radiation dames the internal structure of fiber. The damages are proportional to the intensity of ionizing particles. This results in increasing attenuation due to atomic defects and absorbing optical energy. The total dose a material receives is expressed in rad (Si), this is the unit for measuring radiation absorbed in bulk silicon.

$$1 \text{ rad (Si)} = 0.01 \text{ J.kg}$$

The higher the radiation intensity more the attenuation as shown in Fig 2.2.1.

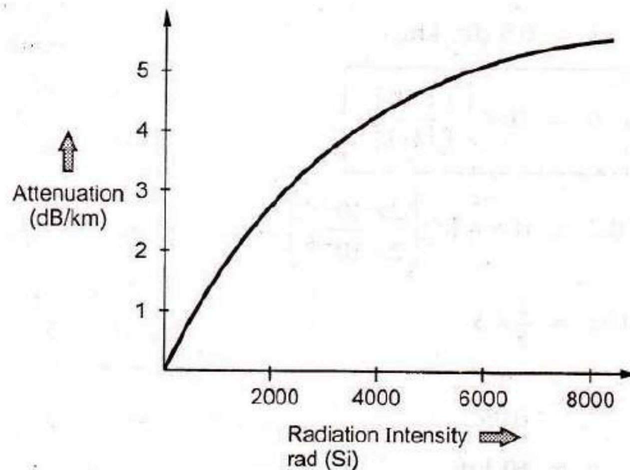


Fig. 2.2.1 Ionizing radiation intensity Vs fiber attenuation

Extrinsic Absorption

- Extrinsic absorption occurs due to electronic transitions between the energy level and because of charge transitions from one ion to another. A major source of attenuation is

  from transition of metal impurity ions such as iron, chromium, cobalt and copper. These losses can be upto 1 to 10 dB/km. The effect of metallic impurities can be reduced by glass refining techniques.
- Another major extrinsic loss is caused by absorption due to **OH (Hydroxil)** ions impurities dissolved in glass. Vibrations occur at wavelengths between 2.7 and 4.2 μm.

The absorption peaks occurs at 1400, 950 and 750 nm. These are first, second and third overtones respectively.

- Fig. 2.2.2 shows absorption spectrum for OH group in silica. Between these absorption peaks there are regions of low attenuation.

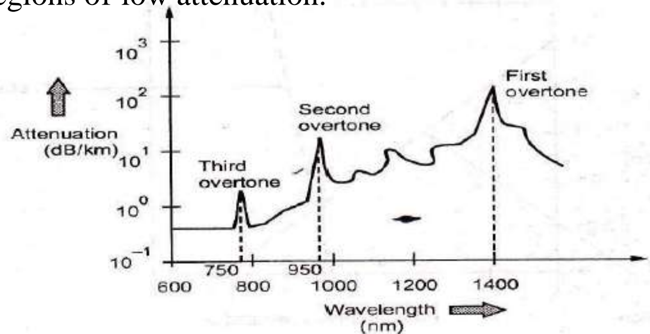


Fig. 2.2.2 Absorption spectra for OH group

Intrinsic Absorption

- Intrinsic absorption occurs when material is in absolutely pure state, no density variation and inhomogenities. Thus intrinsic absorption sets the fundamental lower limit on absorption for any particular material.
- Intrinsic absorption results from electronic absorption bands in UV region and from atomic vibration bands in the near infrared region.
- The electronic absorption bands are associated with the band gaps of amorphous glass materials. Absorption occurs when a photon interacts with an electron in the valene band and excites it to a higher energy level. UV absorption decays exponentially with increasing wavelength (λ).
- In the IR (infrared) region above 1.2 μm the optical waveguide loss is determined by presence of the OH ions and inherent IR absorption of the constituent materials. The inherent IR absorption is due to interaction between the vibrating band and the electromagnetic field of optical signal this results in transfer of energy from field to the band, thereby giving rise to absorption, this absorption is strong because of many bonds present in the fiber.

6. The ultraviolet loss at any wavelength is expressed as,

$$\propto_{uv} = \frac{154.2}{46.6\,x + 60}\ x\ 10^{-2}\ x\ e^{\left(\frac{4.63}{\lambda}\right)} \quad \dots (2.2.1)$$

where, x is mole fraction of GeO_2.

λ is operating wavelength.

α_{uv} is in dB/km.

9. The loss in infrared (IR) region (above 1.2 μm) is given by expression :

$$\propto_{IR} = 7.81 \times 10^{11} \times e^{\left(\frac{-48.48}{\lambda}\right)} \quad \dots (2.2.2)$$

The expression is derived for GeO_2-SiO_2 glass fiber.

Rayleigh Scattering Losses

13. Scattering losses exists in optical fibers because of microscopic variations in the material density and composition. As glass is composed by randomly connected network of molecules and several oxides (e.g. SiO_2, GeO_2 and P_2O_5), these are the major cause of compositional structure fluctuation. These two effects results to variation in refractive index and Rayleigh type scattering of light.
14. **Rayleigh scattering** of light is due to small localized changes in the refractive index of the core and cladding material. There are two causes during the manufacturing of fiber.

---

3. The first is due to slight fluctuation in mixing of ingredients. The random changes because of this are impossible to eliminate completely.
4. The other cause is slight change in density as the silica cools and solidifies. When light ray strikes such zones it gets scattered in all directions. The amount of scatter depends on the size of the discontinuity compared with the wavelength of the light so the shortest wavelength (highest frequency) suffers most scattering. Fig. 2.3.1 shows graphically the relationship between wavelength and Rayleigh scattering loss.

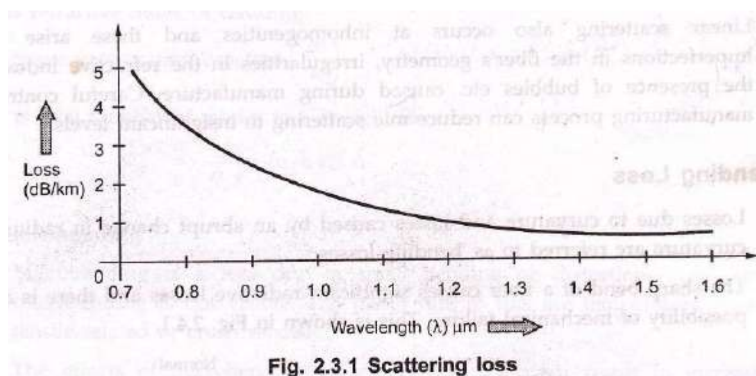


Fig. 2.3.1 Scattering loss

7. Scattering loss for single component glass is given by,

$$\propto_{scat} = \frac{8\pi^3}{3\lambda^4}\,(n^2-1)^2\,k_B\,T_f\,\beta_T \text{ nepers} \quad \dots (2.3.1)$$

where, n = Refractive index

k_B = Boltzmann's constant

β_T = Isothermal compressibility of material

T_f = Temperature at which density fluctuations are frozen into the glass as it solidifies (fictive temperature)

Another form of equation is

$$\propto_{scat} = \frac{8\pi^3}{3\lambda^4}\,n^8\,p^2\,k_B\,T_f\,\beta_T \text{ neper} \propto_{scat} = \frac{8\pi^3}{3\lambda^4}\,(\delta_n^2)^2\,\delta v \quad \dots (2.3.2)$$

where, P = Photoelastic coefficient

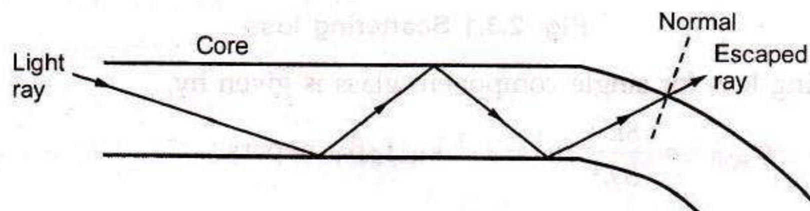
where, δ_n^2 = Mean square refractive index fluctuation

δv = Volume of fiber

- Multimode fibers have higher dopant concentrations and greater compositional fluctuations. The overall losses in this fibers are more as compared to single mode fibers.
  Mie Scattering :
- Linear scattering also occurs at inhomogenities and these arise from imperfections in the fiber's geometry, irregularities in the refractive index and the presence of bubbles etc. caused during manufacture. Careful control of manufacturing process can reduce mie scattering to insignificant levels.

Bending Loss

- Losses due to curvature and losses caused by an abrupt change in radius of curvature are referred to as 'bending losses.'
- The sharp bend of a fiber causes significant radiative losses and there is also possibility of mechanical failure. This is shown in Fig. 2.4.1.



2. As the core bends the normal will follow it and the ray will now find itself on the wrong side of critical angle and will escape. The sharp bends are therefore avoided.
3. The radiation loss from a bent fiber depends on –

   Field strength of certain critical distance x_c from fiber axis where power is lost through radiation.

   The radius of curvature R.
4. The higher order modes are less tightly bound to the fiber core, the higher order modes radiate out of fiber firstly.
5. For multimode fiber, the effective number of modes that can be guided by curved fiber is where, α is graded index profile.

Δ is core – cladding index difference.

n_2 is refractive index of cladding.k is

wave propagation constant $\left(\frac{2\pi}{\lambda}\right)$.

N_∞ is total number of modes in a straight fiber.

$$N_\infty = \frac{\alpha}{\alpha+2}(n_1\, k\, a)^2 \Delta \quad \text{… (2.4.2)}$$

Microbending

- Microbending is a loss due to small bending or distortions. This small microbending is not visible. The losses due to this are temperature related, tensile related or crush related.
- The effects of microbending on multimode fiber can result in increasing attenuation (depending on wavelength) to a series of periodic peaks and troughs on the spectral attenuation curve. These effects can be minimized during installation and testing. Fig. 2.4.2 illustrates microbening.

Macrobending

- The change in spectral attenuation caused by macrobending is different to microbending. Usually there are no peaks and troughs because in a macrobending no light is coupled back into the core from the cladding as can happen in the case of microbends.

- The macrobending losses are cause by large scale bending of fiber. The losses are eliminated when the bends are straightened. The losses can be minimized by not exceeding the long term bend radii. Fig. 2.4.3 illustrates macrobending.

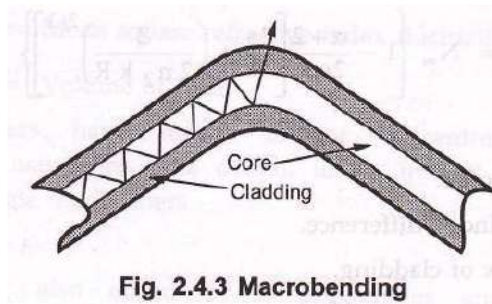


Fig. 2.4.3 Macrobending

Core and Cladding Loss

- Since the core and cladding have different indices of refraction hence they have different attenuation coefficients α1 and α2 respectively.
- For step index fiber, the loss for a mode order (v, m) is given by,

$$\propto_{vm} = \propto_1 \frac{P_{core}}{P} + \propto_2 \frac{P_{cladding}}{P} \quad \ldots (2.5.1)$$

For low-order modes, the expression reduced to

$$\propto_{vm} = \propto_1 + (\propto_2 + \propto_1) \frac{P_{cladding}}{P} \quad \ldots (2.5.2)$$

where, $\frac{P_{core}}{P}$ and $\frac{P_{cladding}}{P}$ are fractional powers.

- For graded index fiber, loss at radial distance is expressed as,

$$\propto (r) = \propto_1 + (\propto_2 - \propto_1) \frac{n^2(0) - n^2(r)}{n^2(0) - n_2^2}$$

… (2.5.3)

The loss for a given mode is expressed by,

$$\alpha_{\text{Graded Index}} = \frac{\int_0^{\infty} \alpha(r)\, P(r)\, r\, dr}{\int_0^{\infty} P(r)\, r\, dr} \qquad \text{… (2.5.4)}$$

where, P(r) is power density of that model at radial distance r.

Signal Distortion in Optical Waveguide

- The pulse get distorted as it travels along the fiber lengths. Pulse spreading in fiber is referred as dispersion. Dispersion is caused by difference in the propagation times of light rays that takes different paths during the propagation. The light pulses travelling down the fiber encounter dispersion effect because of this the pulse spreads out in time domain. Dispersion limits the information bandwidth. The distortion effects can be analyzed by studying the group velocities in guided modes.

Information Capacity Determination

- Dispersion and attenuation of pulse travelling along the fiber is shown in Fig. 2.6.1.

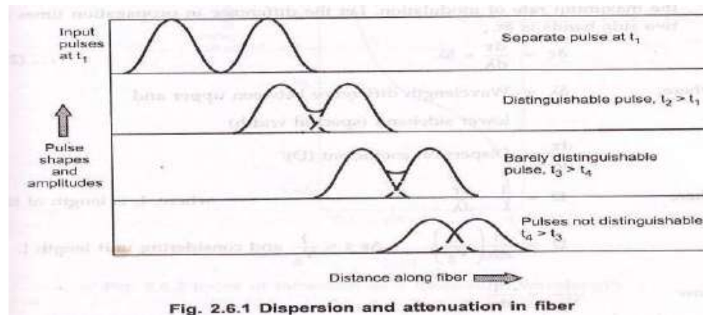


Fig. 2.6.1 Dispersion and attenuation in fiber

- Fig. 2.6.1 shows, after travelling some distance, pulse starts broadening and overlap with the neighbouring pulses. At certain distance the pulses are not even distinguishable and error will occur at receiver. Therefore the information capacity is specified by bandwidth-distance product (MHz . km). For step index bandwidth distance product is 20 MHz . km and for graded index it is 2.5 MHz . km.

Group Delay

- Consider a fiber cable carrying optical signal equally with various modes and each mode contains all the spectral components in the wavelength band. All the spectral components travel independently and they observe different **time delay** and **group delay** in the direction of propagation. The velocity at which the energy in a pulse travels along the fiber is known as **group velocity**. Group velocity is given by,

$$V_g = \frac{\partial w}{\partial \beta} \qquad \text{… (2.6.1)}$$

- Thus different frequency components in a signal will travel at different group velocities and so will arrive at their destination at different times, for digital modulation of carrier, this results in dispersion of pulse, which affects the maximum rate of modulation. Let the difference in propagation times for two side bands is $\delta\tau$.

$$\delta\tau = \frac{d\tau}{d\lambda} \, x \, \delta\lambda \qquad \text{… (2.6.2)}$$

where, $\delta\tau$ = Wavelength difference between upper and lower sideband (spectral width)

$\frac{d\tau}{d\lambda}$ = Dispersion coefficient (D)

Then, $D = \frac{1}{L} \cdot \frac{d\tau}{d\lambda}$ where, L is length of fiber.

$D = \frac{d}{d\lambda}\left(\frac{1}{V_g}\right)$ As $\tau = \frac{1}{V_g}$ and considering unit length L = 1.

Now $\frac{1}{V_g} = \frac{d\beta}{d\omega}$

$$\frac{1}{V_g} = \frac{d\lambda}{d\omega} x \frac{d\beta}{d\lambda}$$

$$\frac{1}{V_g} = \frac{-\lambda^2}{2\pi c} x \frac{d\beta}{d\lambda}$$

$\therefore$

$$D = \frac{d}{d\lambda}\left(\frac{-\lambda^2}{2\pi c} \cdot \frac{d\beta}{d\lambda}\right) \qquad \text{... (2.6.3)}$$

- Dispersion is measured in picoseconds per nanometer per kilometer.

Material Dispersion

- Material dispersion is also called as chromatic dispersion. Material dispersion exists due to change in index of refraction for different wavelengths. A light ray contains components of various wavelengths centered at wavelength λ_{10}. The time delay is

different for different wavelength components. This results in time dispersion of pulse at

the receiving end of fiber. Fig. 2.6.2 shows index of refraction as a function of optical wavelength.

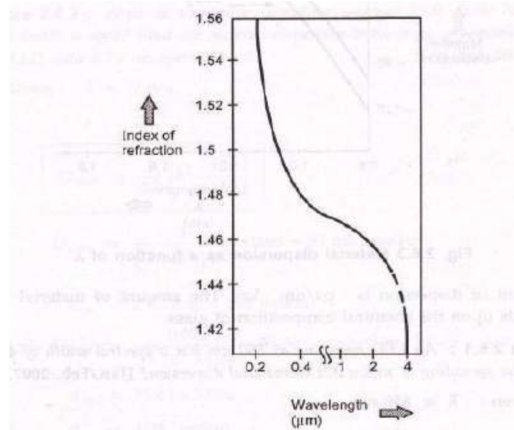


Fig. 2.6.2 Index of refraction as a function of wavelength

2. The material dispersion for unit length (L = 1) is given by

$$D_{mat} = \frac{-\lambda}{c} x \frac{d^2n}{d\lambda^2} \quad \ldots (2.6.4)$$

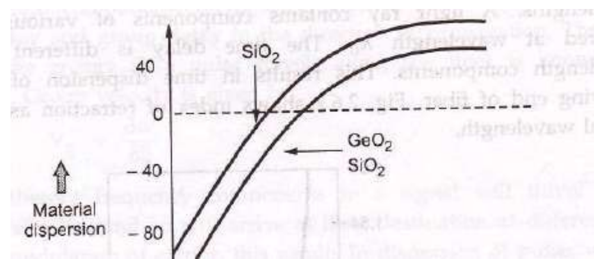
where, c = Light velocity

λ = Center wavelength

$\frac{d^2n}{d\lambda^2}$ = Second derivative of index of refraction w.r.t wavelength

Negative sign shows that the upper sideband signal (lowest wavelength) arrives before the lower sideband (highest wavelength).

- A plot of material dispersion and wavelength is shown in



- The unit of dispersion is : ps/nm . km. The amount of material dispersion depends upon the chemical composition of glass.

Example 2.6.1 : An LED operating at 850 nm has a spectral width of 45 nm. What is the pulse spreading in ns/km due to material dispersion? **[Jan./Feb.-2007, 3 Marks]**

Solution : Given : $\lambda = 850$ nm

$\sigma = 45$ nm

R.M.S pulse broadening due to material dispersion is given by,

$$\sigma_m = \sigma LM$$

Considering length $L = 1$ metre

$$\text{Material dispersion constant } D_{mat} = \frac{-\lambda}{c} \cdot \frac{d^2n}{d\lambda^2}$$

For LED source operating at 850 nm, $\left|\lambda^2 \frac{d^2n}{d\lambda^2}\right| = 0.025$

$$\therefore \quad M = \frac{1}{c\lambda}\left|\lambda^2 \frac{d^2n}{d\lambda^2}\right| = \frac{1}{(3 \times 10^5)(850)} \times 0.025$$

$M = 9.8$ ps/nm/km

$\sigma_m =$ **441 ns/km** **… Ans.**

Example 2.6.2 : What is the pulse spreading when a laser diode having a 2 nm spectral width is used? Find the the material-dispersion-induced pulse spreading at 1550 nm for an LED with a 75 nm spectral width **[Jan./Feb.-2007, 7 Marks]**

Solutions : Given : $\lambda = 2$ nm

$\sigma = 75$

$$D_{mat} = \frac{1}{c\lambda}\left|\lambda^2 . \frac{d^2n}{d\lambda^2}\right|$$

$$D_{mat} = \frac{1}{(3 \times 10^5) \times 2} \times 0.03 = 50 \text{ ps/nm/km}$$

$\sigma m = 2 \times 1 \times 50 =$ **100 ns/km** **… Ans.**

For LED $$D_{mat} = \frac{0.025}{(3 \times 10^5) \times 1550} = 53.76 \text{ ps nm}^{-1}\text{km}^{-1}$$

$\sigma m = 75 \times 1 \times 53.76$

$\sigma m =$ **4.03 ns/km** **… Ans.**

Waveguide Dispersion

- Waveguide dispersion is caused by the difference in the index of refraction between the core and cladding, resulting in a 'drag' effect between the core and cladding portions of the power.
- Waveguide dispersion is significant only in fibers carrying fewer than 5-10 modes. Since multimode optical fibers carry hundreds of modes, they will not have observable waveguide dispersion.
- The group delay (τ_{wg}) arising due to waveguide dispersion.

$$(\tau_{wg}) = \frac{L}{c}\left[n_2 + n_2 \Delta \frac{d\,(kb)}{dk}\right] \quad \text{… (2.6.5)}$$

Where, b = Normalized propagation constant

k = 2π / λ (group velocity)

Normalized frequency V,

$$V = ka(n_1^2 - n_2^2)^{\frac{1}{2}}$$

$$V = k\,a\,n_2\sqrt{2\Delta} \text{ (For small } \Delta)$$

$$\therefore \qquad \tau_{wg} = \frac{L}{c}\left[n_2 + n_2 \Delta \frac{d(Vb)}{dV}\right] \qquad \text{... (2.6.6)}$$

The second term $\frac{d(Vb)}{dV}$ is waveguide dispersion and is mode dependent term..

- As frequency is a function of wavelength, the group velocity of the energy varies with frequency. The produces additional losses (waveguide dispersion). The propagation constant (b) varies with wavelength, the causes of which are independent of material dispersion.

Chromatic Dispersion

- The combination of material dispersion and waveguide dispersion is called chromatic dispersion. These losses primarily concern the spectral width of transmitter and choice of correct wavelength.
- A graph of effective refractive index against wavelength illustrates the effects of material, chromatic and waveguide dispersion.

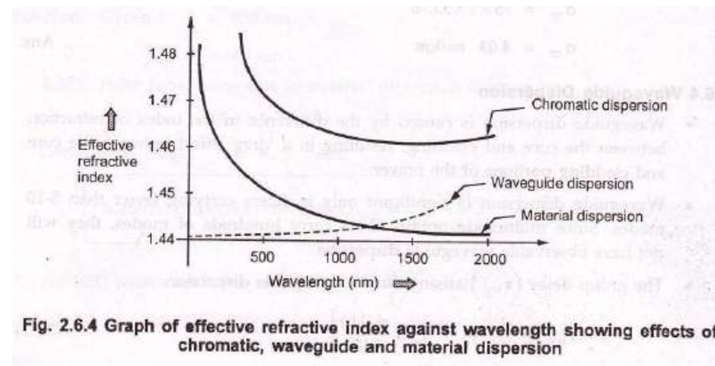


Fig. 2.6.4 Graph of effective refractive index against wavelength showing effects of chromatic, waveguide and material dispersion

- Material dispersion and waveguide dispersion effects vary in vary in opposite senses as the wavelength increased, but at an optimum wavelength around 1300 nm, two effects almost cancel each other and chromatic dispersion is at minimum. Attenuation is therefore also at minimum and makes 1300 nm a highly attractive operating wavelength.

Modal Dispersion

- As only a certain number of modes can propagate down the fiber, each of these modes carries the modulation signal and each one is incident on the boundary at a different angle, they will each have their own individual propagation times. The net effect is spreading of pulse, this form o dispersion is called modal dispersion.
- Modal dispersion takes place in multimode fibers. It is moderately present in graded index fibers and almost eliminated in single mode step index fibers.
- Modal dispersion is given by,

$$\Delta t_{modal} = \frac{n_1 Z}{c}\left(\frac{\Delta}{1-\Delta}\right)$$

where Δt_{modal} = Dispersion

n_1 = Core refractive index

Z = Total fiber length

c = Velocity of light in air

Δ = Fractional refractive index $\left(\frac{n_1 - n_2}{n_1}\right)$

Putting $\Delta = \frac{(NA^2)z}{2n_1 c}$ in above equation

$$\Delta t_{modal} = \frac{(NA^2)Z}{2n_1\ c}$$

- The modal dispersion Δt_{modal} describes the optical pulse spreading due to modal effects optical pulse width can be converted to electrical rise time through the relationship.

$$t_{r\ mod} = 0.44\ (\Delta t_{modal})\pi r^2$$

Signal distortion in Single Mode Fibers

- The pulse spreading σ_{wg} over range of wavelengths can be obtained from derivative of group delay with respect t

$$\sigma_{wg} = \left|\frac{d\tau_{wg}}{d\lambda}\right|\sigma_\lambda$$

where,

$$D_{wg}(\lambda) = \frac{-n_2\Delta}{c\lambda}\left[V\frac{d^2(Vb)}{dV^2}\right] \quad \ldots (2.6.8)$$

- This is the equation for waveguide dispersion for unit length.

Example 2.6.3 : For a single mode fiber $n_2 = 1.48$ and = 0.2 % operating at A = 1320 nm, compute the waveguide dispersion if $V.\frac{d^2(Vb)}{dv^2} = 0.26.$

Solution : $n_2 = 1.48$

0.2

□ = 1320 nm

Waveguide dispersion is given by,

$$D_{wg}(\lambda) = \frac{-n_2\Delta}{c\lambda}\left[V\frac{d^2(Vb)}{dV^2}\right]$$

$$= \frac{-1.48 \text{ x } 0.2}{3 \text{ x } 10^5 \text{ x } 1320}[0.20]$$

i) -1.943 picosec/nm . km.

Higher Order Dispersion

- Higher order dispersive effective effects are governed by dispersion slope S.

$$S = \frac{dD}{d\lambda}$$

where, D is total dispersion

Also, $S = \left(\frac{2\pi c}{\lambda^2}\right)^2 \beta_3 + \left(\frac{4\pi c}{\lambda^3}\right)\beta_2$

where,

β_2 and β_3 are second and third order dispersion parameters.

- Dispersion slope S plays an important role in designing WDM system

Dispersion Induced Limitations

- The extent of pulse broadening depends on the width and the shape of input pulses. The pulse broadening is studied with the help of wave equation.

Basic Propagation Equation

- The basic propagation equation which governs pulse evolution in a single mode fiber is given by,

$$\frac{\partial A}{\partial z} + \beta_1 \frac{\partial A}{\partial t} + \frac{i\beta_2}{2} \cdot \frac{\partial^2 A}{\partial t^2} - \frac{\beta_3}{6} \frac{\partial^3 A}{\partial t^3} = 0$$

where,

β_1, β_2 and β_3 are different dispersion parameters.

Chirped Gaussian Pulses

- A pulse is said to b e chirped if its carrier frequency changes with time.
- For a Gaussian spectrum having spectral width σ_ω, the pulse broadening factor is given by,

$$\frac{\sigma^2}{\sigma_0^2} = \left(1 + \frac{C\beta_2 L}{2\sigma_2^2}\right)^2 + (1 + V_\omega^2)\left(\frac{\beta_2 L}{2\sigma_0^2}\right)^2 + (1 + C + V_\omega^2)^2 \left(\frac{\beta_3 L}{4\sqrt{2\sigma_0^3}}\right)\pi r^2$$

where, $V_\omega = 2\sigma_\omega \sigma_0$

Limitations of Bit Rate

- The limiting bit rate is given by,

$$4B\sigma \leq 1$$

- The condition relating bit rate-distance product (BL) and dispersion (D) is given

$$BL\ |S|\sigma_\lambda^2 \leq \frac{1}{\sqrt{8}}$$

where, S is dispersion slope.

- Limiting bit rate a single mode fibers as a function of fiber length for $\sigma\lambda = 0$, a and 5nm is shown in fig. 2.6.5.

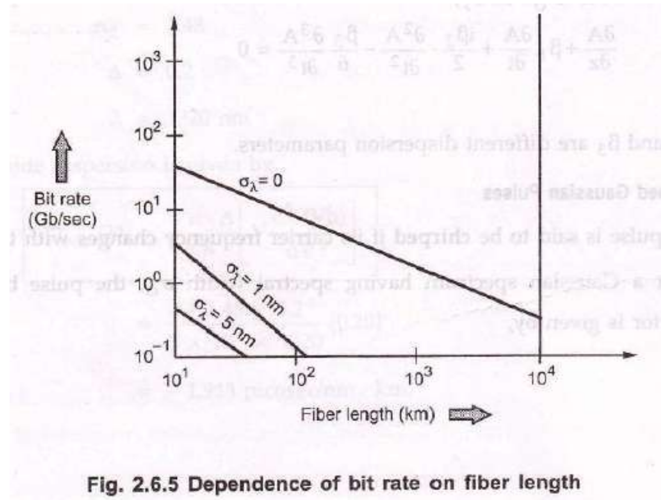


Fig. 2.6.5 Dependence of bit rate on fiber length

Polarization Mode Dispersion (PMD)

- Different frequency component of a pulse acquires different polarization state (such as linear polarization and circular polarization). This results in pulse broadening is kno**w as polarization mode dispersion (PMD).**
- PMD is the limiting factor for optical communication system at high data rates. The effects of PMD must be compensated.

Pulse Broadening in GI Fibers

- The core refractive index varies radially in case of graded index fibers, hence it supports multimode propagation with a low intermodal delay distortion and high data rate over long distance is possible. The higher order modes travelling in outer regions of the core, will travel faster than the lower order modes travelling in high refractive index region. If the index profile is carefully controlled, then the transit times of the individual modes will be identical, so eliminating modal dispersion.
- The r.m.s pulse broadening is given as:

$$\sigma = \left(\sigma^2_{intermodal} + \sigma^2_{intermodal}\right)^{1/2} \qquad \ldots (2.7.1)$$

where,

$\sigma_{intermodal}$ – R.M.S pulse width due to intermodal delay distortion.

$\sigma_{intermodal}$ – R.M.S pulse width resulting from pulse broadening within each mode.

- The intermodal delay and pulse broadening are related by expression given by Personick.

$$\sigma_{intermodal} = \left(\langle\tau_g^2\rangle - \langle\tau_g\rangle^2\right)^{1/2} \qquad \ldots (2.7.2)$$

Where τ_g is group delay.

From this the expression for intermodal pulse broadening is given as:

$$\propto_{intermodal} = \frac{LN_1\Delta}{2c} \cdot \frac{\propto}{\propto+1}\left(\frac{\propto+2}{3\propto+2}\right)^{1/2} \text{x}$$

$$\left[c_1^2 + \frac{4c_1c_2(\propto+1)}{2\propto+1} + \frac{16\Delta^2c_2^2(\propto+1)^2}{(5\propto+2)(3\propto+2)}\right]^{1/2} \qquad \ldots (2.7.3)$$

$$c_1 = \frac{\propto-2-E}{\propto+2} \text{ and } c_2 = \frac{3\propto-2-2c}{2(\propto+2)}$$

- The intramodal pulse broadening is given as :

$$\sigma^2_{intramodal} = \left(\frac{\sigma\lambda}{\lambda}\right)^2 \left\langle\left(\lambda\frac{d\tau g}{d\lambda}\right)^2\right\rangle \qquad \ldots (2.7.4)$$

Where $\sigma\lambda$ is spectral width of optical source.

Solving the expression gives :

$$\sigma^2_{intramodal} = \frac{L}{c} \cdot \frac{\sigma\lambda}{\lambda}\left[\left(-\lambda^2\frac{d^2n_1}{d\lambda^2}\right)^2 - N_1c_1\Delta\right.$$

$$\left.\left(2\lambda^2\frac{d^2n_1}{d\lambda^2} \cdot \frac{\propto}{\propto+1} - N_1c_1\Delta\frac{4\propto^2}{(\propto+2)(3\propto+2)}\right)\right]^{1/2}$$

Mode Coupling

- After certain initial length, the pulse distortion increases less rapidly because of mode coupling. The energy from one mode is coupled to other mods because of:
  - Structural imperfections.
  - Fiber diameter variations.
  - Refractive index variations.
  - Microbends in cable.
- Due to the mode coupling, average propagation delay become less and intermodal distortion reduces.
- Suppose certain initial coupling length = L_C, mode coupling length, over L_C = Z. Additional loss associated with mode coupling = h (dB/ km).

Therefore the excess attenuation resulting from mode coupling = hZ.

The improvement in pulse spreading by mode coupling is given as :

$$hZ\left(\frac{\sigma_c}{\sigma_0}\right) = C$$

where, C is constant independent of all dimensional quantities and refractive indices.

σ_C is pulse broadening under mode coupling.

σ_0 is pulse broadening in absence of mode coupling.

- For long fiber length's the effect of mode coupling on pulse distortion is significant. For a graded index fiber, the effect of distance on pulse broading for various coupling losses are shown

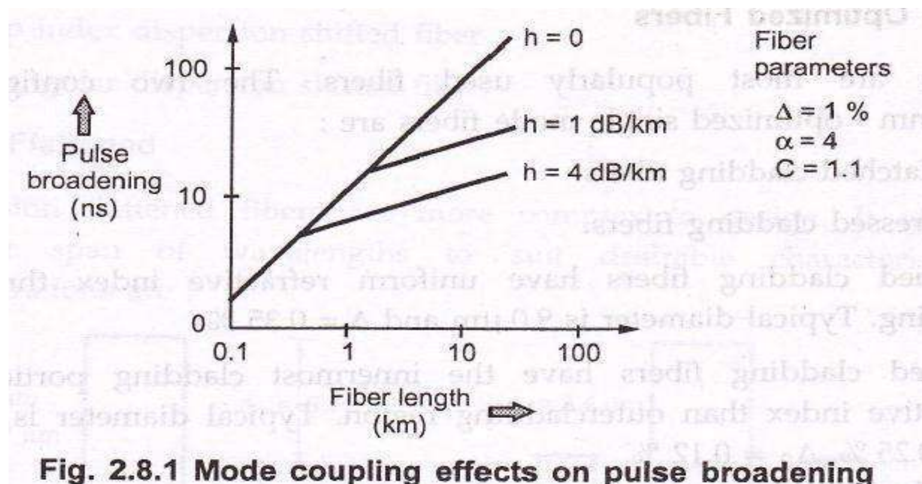


Fig. 2.8.1 Mode coupling effects on pulse broadening

Significant mode coupling occurs of connectors, splices and with other passive components of an optical link.

Design Optimization

- Features of single mode fibers are :
  - Longer life.
  - Low attenuation.
  - Signal transfer quality is good.
  - Modal noise is absent.
  - Largest BW-distance product.
- Basic design – optimization includes the following :
  - Cut-off wavelength.
  - Dispersion.
  - Mode field diameter.
  - Bending loss.
  - Refractive index profile.

Refractive Index Profile

- Dispersion of single mode silica fiber is lowest at 1300 nm while its attenuation is minimum at 1550 nm. For archiving maximum transmission distance the dispersion null should be at the wavelength of minimum attenuation. The waveguide dispersion is easier to control than the material dispersion. Therefore a variety of core-cladding refractive.

  idex configuration fivers. Such as 1300 nm – optimized fibers, dispersion shifted fibers, dispersion – flattened fibers and large effective core area fibers.

1300 nm – Optimized Fibers

These are most popularly used fibers. The two configurations of 1300 nm – optimized single mode fibers are :

Matched cladding fibers.

Dressed cladding fibers.

Matched cladding fibers have uniform refractive index throughout its cladding. Typical diameter is 9.0 μm and $\Delta = 0.35$ %.

Dressed cladding fibers have the innermost cladding portion has low refractive index than outrcladding region. Typical diameter is 8.4 μm and $\Delta_1 = 0.25$ %, $\Delta_2 = 0.12$ %.

Fig 2.9.1 shows both types of fibers.

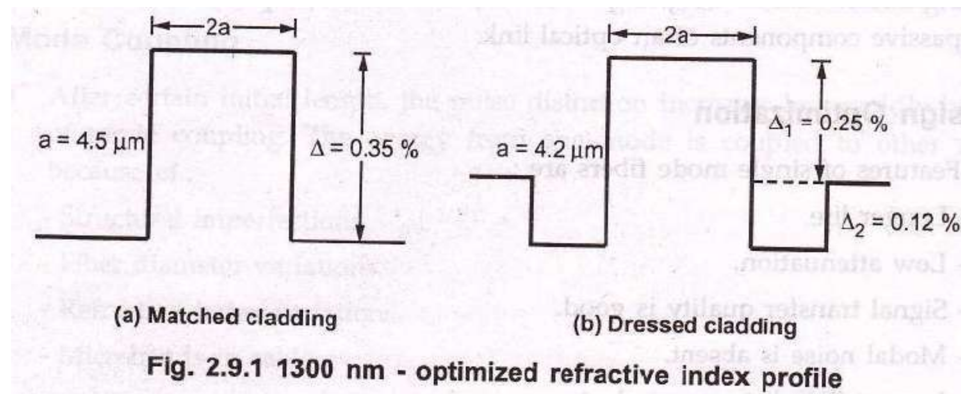


Fig. 2.9.1 1300 nm - optimized refractive index profile

2. Dispersion Shifted Fibers

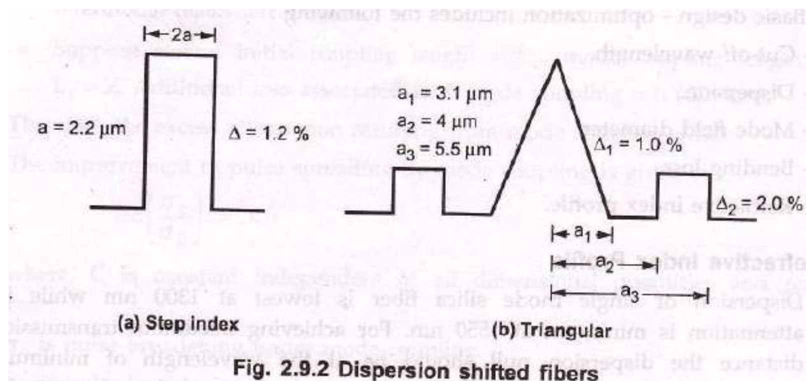


Fig. 2.9.2 Dispersion shifted fibers

3. The addition of wavelength and material dispersion can shift the zero dispersion point of longer wavelength. Two configurations of dispersion shifted fibers are

Step index dispersion shifted fiber.

Triangular dispersion shifted fiber.

□ **Dispersion Flattened**

Dispersion flattened fibers are more complex to design. It offers much broader span of wavelengths to suit desirable characteristics. Two configurations are :

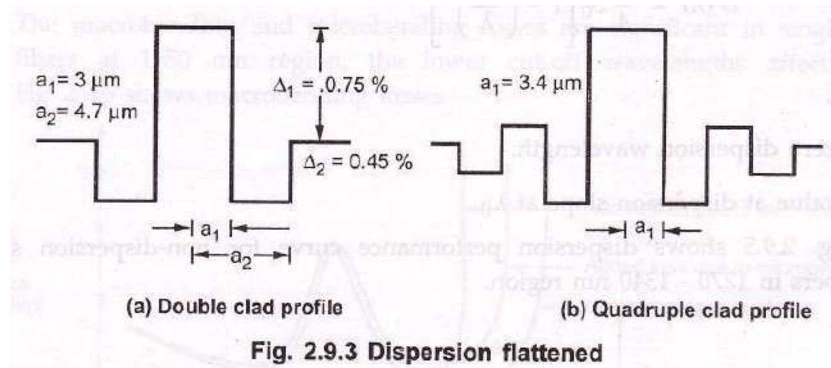


Fig. 2.9.3 Dispersion flattened

- Fig 2.9.4 shows total resultant dispersion.

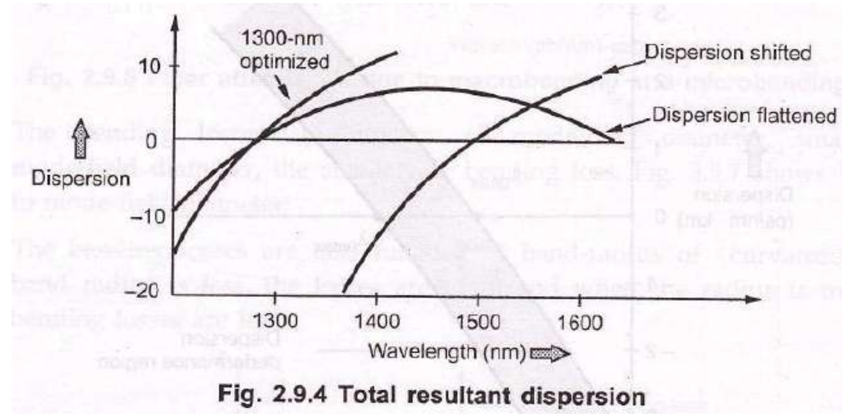


Fig. 2.9.4 Total resultant dispersion

Dispersion Calculations

- The total dispersion consists of material and waveguide dispersions. The resultant intermodal dispersion is given as,

$$D(\lambda) = \frac{d\tau}{d\lambda}$$

where, τ is group delay per unit length of fiber.

- The broadening σ of an optical pulse is given

$$\sigma = D(\lambda) L \sigma \lambda$$

where, $\sigma\lambda$ is half power spectral width of source.

= As the dispersion varies with wavelength and fiber type. Different formulae are used to calculate dispersions for variety of fiber at different wavelength.

= For a non – dispersion shifted fiber between 1270 nm to 1340 nm wavelength, the expression for dispersion is given as :

$$D(\lambda) = \frac{\lambda}{4} S_0 \left[1 - \left(\frac{\lambda_0}{\lambda}\right)^4\right]$$

where,

λ_0 is zero dispersion wavelength.

S_0 is value at dispersion slop at λ_0.

ιιι) Fig 2.9.5 shows dispersion performance curve for non-dispersion shifted fibers in 1270 – 1340 nm region.

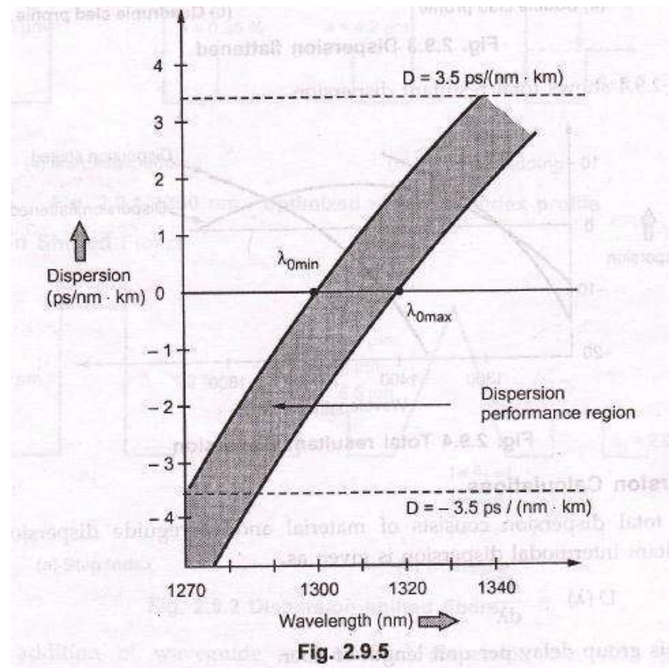


Fig. 2.9.5

- Maximum dispersion specified as 3.5 ps/(nm . km) marked as dotted line in Fig. 2.9.5.

The cut-off frequency of an optical fiber

- The cut-off frequency of an optical fiber is determined not only by the fiber itself (modal dispersion in case of multimode fibers and waveguide dispersion in case of single mode fibers) but also by the amount of material dispersion caused by the spectral width of transmitter.

Bending Loss Limitations

- The macrobending and microbending losses are significant in single mode fibers at 1550 nm region, the lower cut-off wavelengths affects more. Fig. 2.9.6 shows macrobending losses.

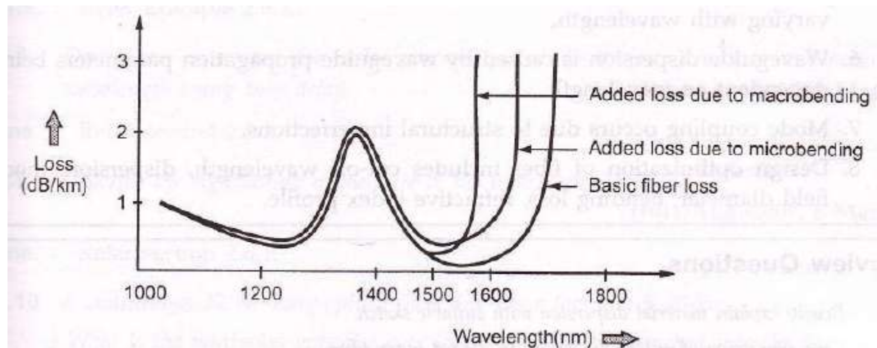


Fig. 2.9.6 Fiber attenuation due to macrobending and microbending

〈 The bending losses are function of mode-filed diameter, smaller the mode-field diameter, the smaller the bending loss. Fig. 2.9.7 shows loss due to mode-field diameter.

〈 The bending losses are also function of bend-radius of curvature. If the bend radius is less, the losses are more and when the radius is more, the bending losses are less.

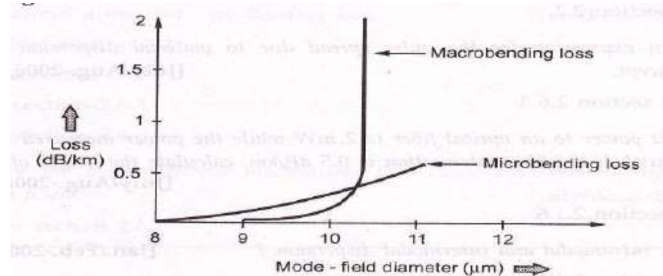


Fig. 2.9.7 Loss due to mode field diameter variation

Recommended Questions:

d) Briefly explain material dispersion with suitable sketch.

e) Give expression of pulse broadening in graded index fiber.

f) State the significance of mode coupling in optic fiber communication.

g) Explain in detail the design optimization of single mode fibers.

h) Elaborate dispersion mechanism in optical fibers.

UNIT - 3

OPTICAL SOURCES AND COUPLING

Optical Sources

3. Optical transmitter coverts electrical input signal into corresponding optical signal. The optical signal is then launched into the fiber. Optical source is the major component in an optical transmitter.
4. Popularly used optical transmitters are Light Emitting Diode (LED) and semiconductor Laser Diodes (LD).

Characteristics of Light Source of Communication

- To be useful in an optical link, a light source needs the following characteristics:
- It must be possible to operate the device continuously at a variety of temperatures for many years.
- It must be possible to modulate the light output over a wide range of modulating frequencies.
- For fiber links, the wavelength of the output should coincide with one of transmission windows for the fiber type used.
- To couple large amount of power into an optical fiber, the emitting area should be small.
- To reduce material dispersion in an optical fiber link, the output spectrum should be narrow.
- The power requirement for its operation must be low.

- The light source must be compatible with the modern solid state devices.
- The optical output power must be directly modulated by varying the input current to the device.
- Better linearity of prevent harmonics and intermodulation distortion.
- High coupling efficiency.
- High optical output power.
- High reliability.
- Low weight and low cost.

Two types of light sources used in fiber optics are light emitting diodes (LEDs) and laser diodes (LDs).

Light Emitting Diodes(LEDs)

p-n Junction

Conventional p-n junction is called as **homojunction** as same semiconductor material is sued on both sides junction. The electron-hole recombination occurs in relatively

layer = 10 μm. As the carriers are not confined to the immediate vicinity of junction, hence high current densities can not be realized.

- The carrier confinement problem can be resolved by sandwiching a thin layer (= 0.1 μm) between p-type and n-type layers. The middle layer may or may not be doped. The carrier confinement occurs due to bandgap discontinuity of the junction. Such a junction is call **heterojunction** and the device is called double **heterostructure.**
- In any optical communication system when the requirements is –
  1. Bit rate f 100-2—Mb/sec.
  2. Optical power in tens of micro watts.

     LEDs are best suitable optical source.

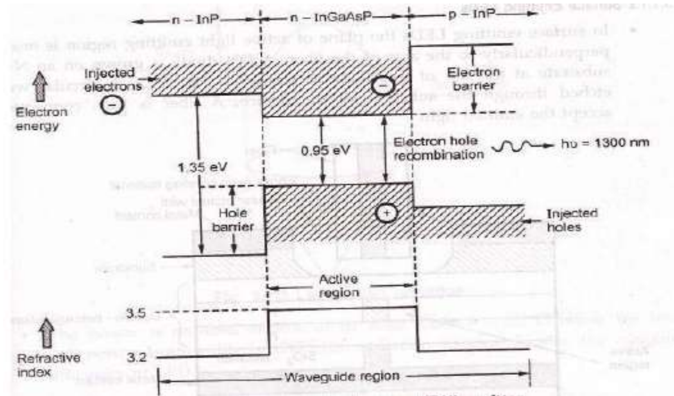
LED Structures

Heterojuncitons

- A heterojunction is an interface between two adjoining single crystal semiconductors with different bandgap.
- Heterojuctions are of two types, Isotype (n-n or p-p) or Antisotype (p-n).

Double Heterojunctions (DH)

In order to achieve efficient confinement of emitted radiation double **heterojunctions** are used in LED structure. A heterojunciton is a junction formed by dissimilar semiconductors. Double heterojunction (DH) is formed by two different semiconductors on each side of active region. Fig. 3.1.1 shows double heterojunction (DH) light emitter.



- The crosshatched regions represent the energy levels of freecharge. Recombination occurs only in active InGaAsP layer. The two materials have different bandgap energies and different refractive indices. The changes in bandgap energies create potential barrier for both holes and electrons. The free charges can recombine only in narrow, well defined active layer side.
- A double heterjuction (DH) structure will confine both hole and electrons to a narrow active layer. Under forward bias, there will be a large number of carriers injected into active region where they are efficiently confined. Carrier recombination occurs in small active region so leading to an efficient device. Antoer advantage DH structure is that the active region has a higher refractive index than the materials on either side, hence light emission occurs in an optical waveguide, which serves to narrow the output beam.

LED configurations

- At present there are two main types of LED used in optical fiber links –

  Surface emitting LED.

  Edge emitting LED.

  Both devices used a DH structure to constrain the carriers and the light to an active layer.

Surface Emitting LEDs

In surface emitting LEDs the plane of active light emitting region is oriented perpendicularly to the axis of the fiber. A DH diode is grown on an N-type substrate at the top of the diode as shown in Fig. 3.1.2. A circular well is etched through the substrate of the device. A fiber is then connected to accept the emitted

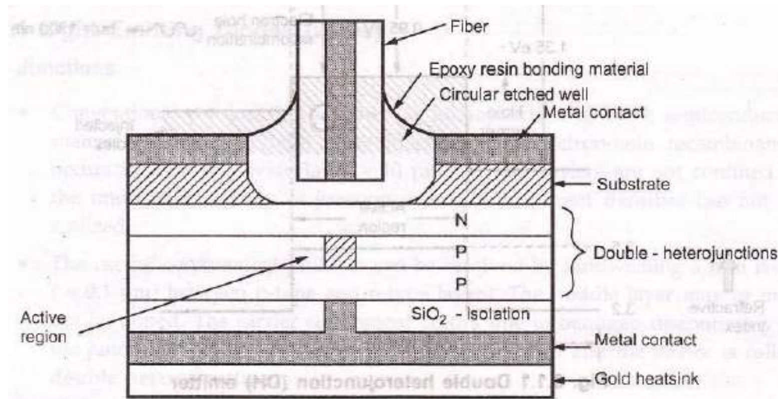


Fig. 3.1.2 Cross-section through a typical surface emitting LED

- At the back of device is a gold heat sink. The current flows through the p-type material and forms the small circular active region resulting in the intense beam of light.
  Diameter of circular active area = 50 μm
  Thickness of circular active area = 2.5 μm
  Current density = 2000 A/cm^2 half-power
  Emission pattern = Isotropic, 120° beamwidth.
- The isotropic emission pattern from surface emitting LED is of Lambartian pattern. In Lambartian pattern, the emitting surface is uniformly bright, but its projected area diminishes as cos θ, where θ is the angle between the viewing direction and the normal to the surface as shown in Fig. 3.1.3. The beam intensity is maximum along the normal.

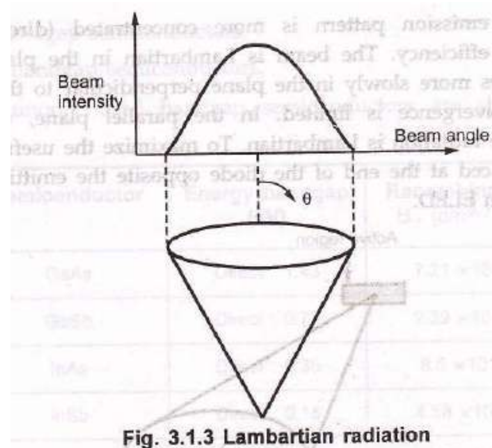


Fig. 3.1.3 Lambartian radiation

- The power is reduced to 50% of its peak when θ = 60°, therefore the total half-power beamwidth is 120°. The radiation pattern decides the coupling efficiency of LED.

Edge Emitting LEDS (ELEDs)

- In order to reduce the losses caused by absorption in the active layer and to make the beam more directional, the light is collected from the edge of the LED. Such a device is known as **edge emitting** LED or ELED.

It consists of an active junction region which is the source of incoherent light and two guiding layers. The refractive index of guiding layers is lower than active region but higher than outer surrounding material. Thus a waveguide channel is form and optical radiation is directed into the fiber. Fig. shows structure of LED

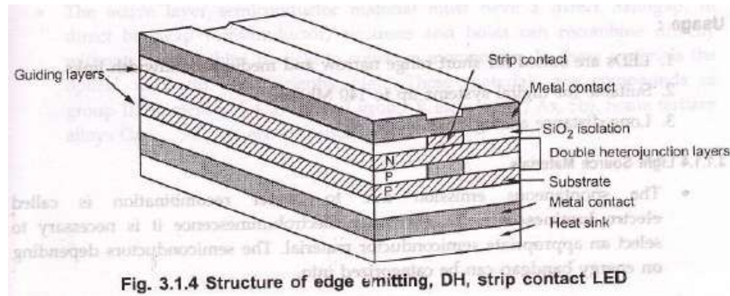


Fig. 3.1.4 Structure of edge emitting, DH, strip contact LED

Edge emitter's emission pattern is more concentrated (directional) providing improved coupling efficiency. The beam is Lambartian in the plane parallel to the junction but diverges more slowly in the plane perpendicular to the junction. In this plane, the beam divergence is limited. In the parallel plane, there is no beam confinement and the radiation is Lambartian. To maximize the useful output power, a reflector may be placed at the end of the diode opposite the emitting edge. Fig. 3.1.5 shows radiation from ELED.

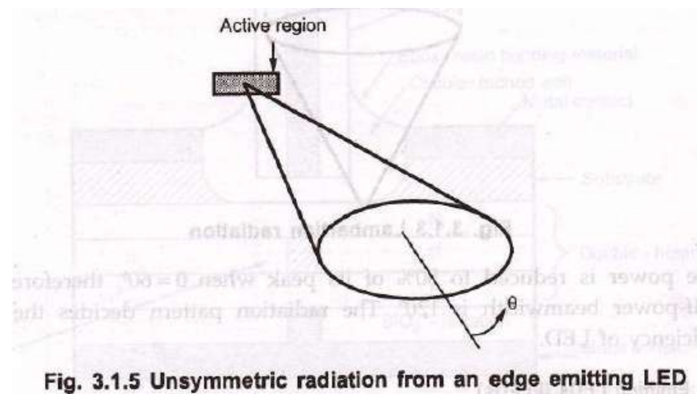


Fig. 3.1.5 Unsymmetric radiation from an edge emitting LED

Features of ELED:

- Linear relationship between optical output and current.
- Spectral width is 25 to 400 nm for $\lambda = 0.8 - 0.9$ μm.
- Modulation bandwidth is much large.

- Not affected by catastrophic gradation mechanisms hence are more reliable.
- ELEDs have better coupling efficiency than surface emitter.
- ELEDs are temperature sensitive.

Usage :

7. LEDs are suited for short range narrow and medium bandwidth links.
8. Suitable for digital systems up to 140 Mb/sec.
9. Long distance analog links

Light Source Materials

10. The spontaneous emission due to carrier recombination is called **electro luminescence**. To encourage electroluminescence it is necessary to select as appropriate semiconductor material. The semiconductors depending on energy bandgap can be categorized into,
    - Direct bandgap semiconductors.
    - Indirect bandgap semiconductors.
11. Some commonly used bandgap semiconductors are shown in following table 3.1.1

Semiconductor	Energy bandgap (eV)	Recombination B_r (cm^3 / sec)
GaAs	Direct : 1.43	7.21×10^{-10}
GaSb	Direct : 0.73	2.39×10^{-10}
InAs	Direct : 0.35	8.5×10^{-11}
InSb	Direct : 0.18	4.58×10^{-11}
Si	Indirect : 1.12	1.79×10^{-15}
Ge	Indirect : 0.67	5.25×10^{-14}
GaP	Indirect : 2.26	5.37×10^{-14}

Table 3.1.1 Semiconductor material for optical sources

15. Direct bandgap semiconductors are most useful for this purpose. In direct bandgap semiconductors the electrons and holes on either side of bandgap have same value of crystal momentum. Hence direct recombination is possible. The recombination occurs within 10^{-8} to 10^{-10} sec.
16. In indirect bandgap semiconductors, the maximum and minimum energies occur at

different values of crystal momentum. The recombination in these semiconductors is quite slow i.e. 10^{-2} and 10^{-3} sec.

The active layer semiconductor material must have a **direct bandgap**. In direct bandgap semiconductor, electrons and holes can recombine directly without need of third particle to conserve momentum. In these materials the optical radiation is sufficiently high. these materials are compounds of group III elements (Al, Ga, In) and group V element (P, As, Sb). Some tertiary allos $Ga_{1-x} Al_x As$ are also used.

5. Emission spectrum of $Ga_{1-x} Al_x As$ LED is shown in Fig. 3.1.6.

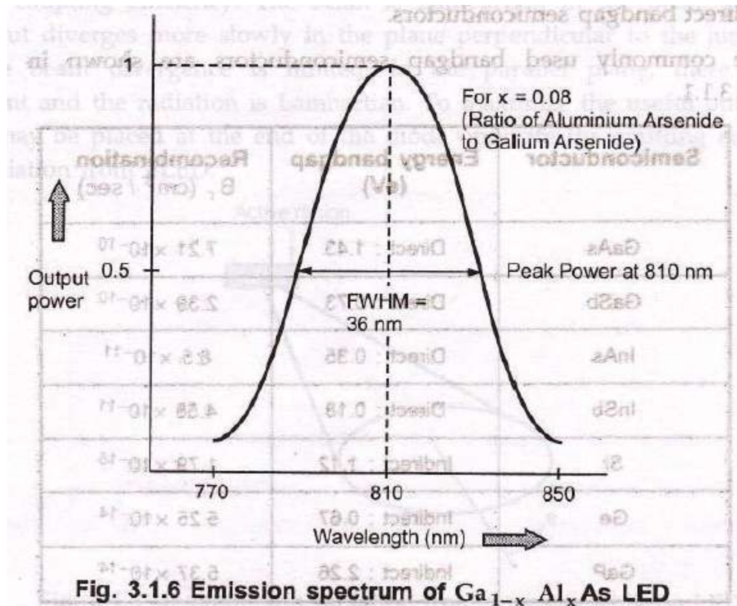


Fig. 3.1.6 Emission spectrum of $Ga_{1-x} Al_x As$ LED

8. The peak output power is obtained at 810 nm. The width of emission spectrum at half power (0.5) is referred as full width half maximum (FWHM) spectral width. For the given LED FWHM is 36 nm.
9. The fundamental quantum mechanical relationship between gap energy E and frequency v is given as –

$$E = hv$$

$$E = h\frac{c}{\lambda}$$

$$\Rightarrow \quad \lambda = \frac{hc}{E}$$

where, energy (E) is in joules and wavelength (λ) is in meters. Expressing the gap energy (E_g) in electron volts and wavelength (λ) in micrometers for this application.

$$\boxed{\lambda(\mu m) = \frac{1.24}{E_g(eV)}}$$

Different materials and alloys have different band gap energies

3. The bandgap energy (E_g) can be controlled by two compositional parameters x and y, within direct bandgap region. The quartenary alloy In_{1-x} Ga_x As_y P_{1-y} is the principal material sued in such LEDs. Two expression relating E_g and x,y are –

$$E_g = 1.424 + 1.266\,x + 0266\,x^2 \quad \ldots 3.1.3$$

$$E_g = 1.35 - 0.72\,y + 0.12\,y^2 \quad \ldots 3.1.4$$

Example 3.1.1 : Compute the emitted wavelength from an optical source having x = 0.07.

Solution : x = 0.07

$$E_g = 1.424 + 1.266\,x + 0266\,x^2$$

$$E_g = 1.424 + (1.266 \times 0.07) + 0266 \times (0.07)^2$$

E_g = 1.513 eV

Now

$$\lambda = \frac{1.24}{E_g}$$

$$\lambda = \frac{1.24}{1.513}$$

$\lambda = 0.819\ \mu m$

$\lambda = \mathbf{0.82}\ \mu m$ **…Ans.**

Example 3.1.2 : For an alloy In0.74 $Ga_{0.26}$ $As_{0.57}$ $P_{0.43}$ to be sued in Led. Find the wavelength emitted by this source.

Solution : Comparing the alloy with the quartenary alloy composition.

In_{1-x} Ga_x As P_{1-y} it is found that

x = 0.26 and y= 0.57

$$E_g = 1.35 - 0.72\,y + 0.12\,y^2$$

Using

$$E_g = 1.35-(0.72 \times 0.57) + 0.12 \times 0.57^2$$

$$E_g = 0.978 \text{ eV}$$

Now

$$\lambda = \frac{1.24}{E_g}$$

$$\lambda = \frac{1.24}{0.978}$$

$$\lambda = 1.2671\ \mu m$$

$$\boldsymbol{\lambda = 1.27\ \mu m}$$ **… Ans.**

Quantum Efficiency and Power

- The internal quantum efficiency (η_{int}) is defined as the ratio of radiative recombination rate to the total recombination rate.

$$\eta_{int} = \frac{R_r}{R_r + R_{nr}} \quad \text{… 3.1.5}$$

Where,

R_r is radiative recombination rate.

R_{nr} is non-radiative recombination rate.

If n are the excess carriers, then radiative life time, $\tau_r = \frac{n}{R_r}$ and

non-radiative life time, $\tau_r = \frac{n}{R_{nr}}$

The internal quantum efficiency is given

6 The recombination time of carriers in active region is τ. It is also known as bulk recombination life time.

$$\frac{1}{\tau} = \quad \eta_{int} = \frac{1}{1 + \frac{R_{nr}}{R_r}} \quad \text{… 3.1.7}$$

$$\eta_{int} = \frac{1}{[illegible]}$$

Therefore internal quantum efficiency is given as –

$$\eta_{int} = \frac{\tau}{\tau_r} \quad \ldots 3.1.8$$

- If the current injected into the LED is I and q is electron charge then total number of recombinations per second is –

$$R_r = R_{nr} = \frac{I}{q}$$ From equation 3.1.5

$$\eta_{int} = \frac{R_r}{I / q}$$

∴

$$R_r = \eta_{int} \times \frac{I}{q} \quad \ldots 3.1.9$$

- Optical power generated internally in LED is given as –

$$P_{int} = R_r . h\,v$$

$$P_{int} = \left(\eta_{int} \times \frac{I}{q}\right) . h\,v$$

$$P_{int} = \left(\eta_{int} \times \frac{I}{q}\right) . h\frac{c}{\lambda}$$

∴

$$P_{int} = \eta_{int} . \frac{hc\,I}{q\lambda} \quad \ldots 3.1.10$$

Not all internally generated photons will available from output of device. The external quantum efficiency is used to calculate the emitted power. The external quantum

efficiency is defined as the ratio of photons emitted from LED to the number of photons generated internally. It is given by equation

$$\eta_{ext} = \frac{1}{n(n+1)^2} \quad \ldots 3.1.11$$

- The optical output power emitted from LED is given as –

$$P = \eta_{ext} . P_{int}$$

$$P = \frac{1}{n\,(n+1)^2} . P_{int}$$

Example 3.1.3 : The radiative and non radiative recombination life times of minority carriers in the active region of a double heterojunction LED are 60 nsec and 90 nsec respectively. Determine the total carrier recombination life time and optical power generated internally if the peak emission wavelength si 870 nm and the drive currect is 40 mA. **[July/Aug.-2006, 6 Marks]**

Solutions : Given : $\lambda = 870$ nm 0.87×10^{-6} m

$\tau_r = 60$ nsec.

$\tau_{nr} = 90$ nsec.

$I = 40$ mA $= 0.04$ Amp.

i) Total carrier recombination life time:

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}}$$

$$\frac{1}{\tau} = \frac{1}{60} + \frac{1}{90}$$

$$\frac{1}{\tau} = \frac{150}{5400}$$

$\therefore$ $\tau =$ **36 nsec.** **… Ans.**

ii) Internal optical power

$$P_{int} = \eta_{int} \cdot \frac{hc\, I}{q\lambda}$$

$$P_{int} = \left(\frac{\tau}{\tau_r}\right)\left(\frac{hc\, I}{q\lambda}\right)$$

$$P_{int} = \left(\frac{30}{60}\right)\left[\frac{(6.625 \times 10^{-34})(3 \times 10^{8}) \times 0.04}{(1.602 \times 10^{-19})(0.87 \times 10^{-6})}\right]$$

$$P_{int} = 34.22 \text{ mW}$$

Example 3.1.4 : A double heterjunciton InGaAsP LED operating at 1310 nm has radiative and non-radiative recombination times of 30 and 100 ns respectively. The current injected is 40 Ma. Calculate –

- Bulk recombination life time.
- Internal quantum efficiency.
- Internal power level.

Solution : λ = 1310 nm = (1.31 x 10^{-6} m)

τ_r = 30 ns

τ_{nr} = 100 ns

I = 40 MA – 0.04 Amp.

- **Bulk Recombination Life time (τ) :**

$$\eta_{int} = \frac{\tau}{\tau_r}$$

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}}$$

$\therefore$ τ = 23.07 **nsec.** **... Ans.**

- **Internal euqntum efficienty (η_{int})**

$$\eta_{int} = \frac{23.07}{30}$$

$$\eta_{int} = \mathbf{0.769}$$

... Ans.

iii) Internal pwer level (P_{int}) :

$$P_{int} = \eta_{int} \cdot \frac{hc\, I}{q\lambda}$$

Advantages and Disadvantages of LED

Advantages of LED

- Simple design.
- Ease of manufacture.
- Simple system integration.
- Low cost.
- High reliability.

Disadvantages of LED

- Refraction of light at semiconductor/air interface.
- The average life time of a radiative recombination is only a few nanoseconds, therefore nodulation BW is limited to only few hundred megahertz.
- Low coupling efficiency.
- Large chromatic dispersion.

Comparison of Surface and Edge Emitting LED

LED type	Maximum modulation frequency (MHz)	Output power (mW)	Fiber coupled power (mW)
Surface emitting	60	< 4	< 0.2
Edge emitting	200	< 7	< 1.0

Injection Laser Diode (ILD)

- The laser is a device which amplifies the light, hence the LASER is an acronym for light amplification by stimulated emission of radiation.

  The operation of the device may be described by the formation of an electromagnetic standing wave within a cavity (optical resonator) which provides an output of monochromatic highly coherent radiation.

Principle :

Material absorb light than emitting. Three different fundamental process occurs between the two energy states of an atom.

Absorption 2) Spontaneous emission 3) Stimulated emission.

- Laser action is the result of three process absorption of energy packets (photons) spontaneous emission, and stimulated emission. (These processes are represented by the simple two-energy-level diagrams).

  Where E_1 is the lower state energy level.

  E_2 is the higher state energy level.

- Quantum theory states that any atom exists only in certain discrete energy state, absorption or emission of light causes them to make a transition from one state to another. The frequency of the absorbed or emitted radiation f is related to the difference in energy E between the two states.

  If E_1 is lower state energy level.

  and E_2 is higher state energy level.

  $E = (E_2 - E_1) = h.f.$

  Where, $h = 6.626 \times 10^{-34}$ J/s (Plank's constant).

- An atom is initially in the lower energy state, when the photon with energy $(E_2 - E_1)$ is incident on the atom it will be excited into the higher energy state E_2 through the absorption of the photon

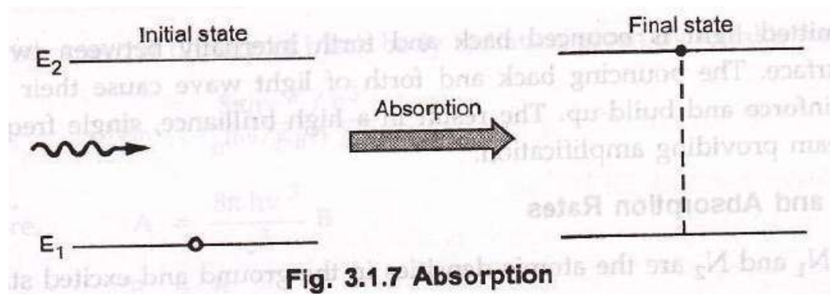


Fig. 3.1.7 Absorption

- When the atom is initially in the higher energy state E_2, it can make a transition to the lower energy state E_1 providing the emission of a photon at a frequency corresponding to $E = h.f.$ The emission process can occur in two ways.

  By spontaneous emission in which the atom returns to the lower energy state in random manner.

  By stimulated emission when a photon having equal energy to the difference between the two states $(E_2 - E_1)$ interacts with the atom causing it to the lower state with the creation of the second photon

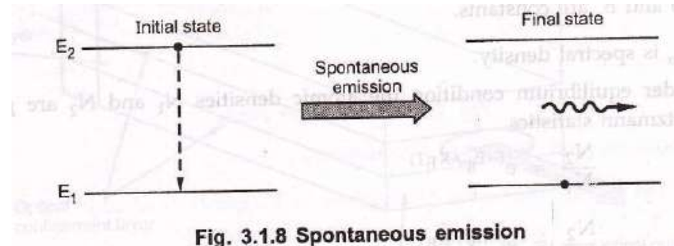


Fig. 3.1.8 Spontaneous emission

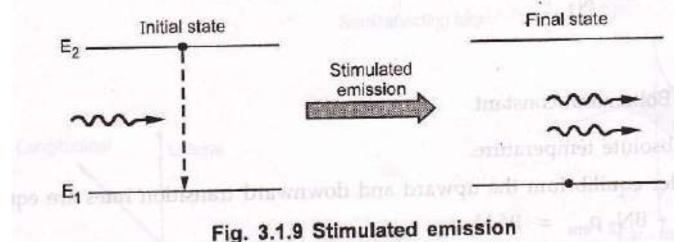


Fig. 3.1.9 Stimulated emission

- Spontaneous emission gives incoherent radiation while stimulated emission gives coherent radiation. Hence the light associated with emitted photon is of same frequency of incident photon, and in same phase with same polarization.
- It means that when an atom is stimulated to emit light energy by an incident wave, the liberated energy can add to the wave in constructive manner. The emitted light is bounced back and forth internally between two reflecting surface. The bouncing back and forth of light wave cause their intensity to reinforce and build-up. The result in a high brilliance, single frequency light beam providing amplification.

Emission and Absorption Rates

3. It N_1 and N_2 are the atomic densities in the ground and excited states.

Rate of spontaneous emission

$$R_{spon} = AN_2 \quad \ldots 3.1.13$$

Rate of stimulated emission

$$R_{stim} = BN_2 \rho_{em} \quad \ldots 3.1.14$$

Rate of absorption

$$R_{abs} = B' N_1 \rho_{em} \quad \ldots 3.1.15$$

where,

A, B and B' are constants.

ρ_{em} is spectral density.

- Under equilibrium condition the atomic densities N_1 and N_2 are given by Boltzmann statistics.

$$\frac{N_2}{N_1} = eg^{(-E_B / K_B T)} \quad \ldots 3.1.16$$

$$\frac{N_2}{N_1} = eg^{(-h_v / K_B T)} \quad \ldots 3.1.17$$

where,

K_B is Boltzmann constant.

T is absolute temperature.

- Under equilibrium the upward and downward transition rates are equal.

  $AN_2 + BN_2 \rho_{em} = B' N_1 \rho_{em}$ … 3.1.18

Spectral density ρ_{em}

… 3.1.19

Comparing spectral density of black body radiation given by Plank's formula,

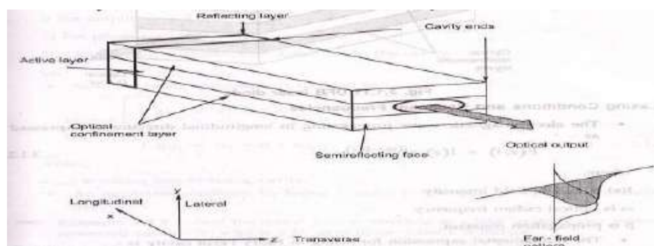
… 3.1.20

Therefore, … 3.1.21

- A and B are called Einstein's coefficient.

Fabry – Perot Resonator

- Lasers are oscillators operating at frequency. The oscillator is formed by a resonant cavity providing a selective feedback. The cavity is normally a Fabry-Perot resonator i.e. two parallel plane mirrors separated by distance L,



Light propagating along the axis of the interferometer is reflected by the mirrors back to the amplifying medium providing optical gain. The dimensions of cavity are 25-500 μm longitudinal 5-15 μm lateral and 0.1-0.2 μm transverse. Fig. 3.1.10 shows Fabry-Perot resonator cavity for a laser diode.

- The two heterojunctions provide carrier and optical confinement in a direction normal to the junction. The current at which lasing starts is the threshold current. Above this current the output power increases sharply.

Distributed Feedback (DFB) Laser

In DFB laster the lasing action is obtained by periodic variations of refractive index along the longitudinal dimension of the diode. Fig. 3.1.11 shows the structure of DFB laser diode

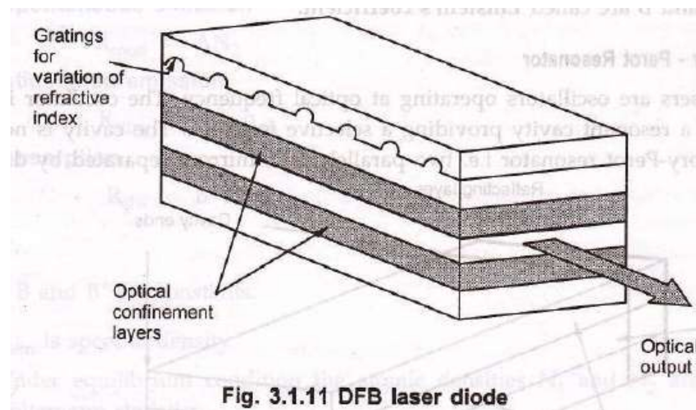


Fig. 3.1.11 DFB laser diode

Lasing conditions and resonant Frequencies

- The electromagnetic wave propagating in longitudinal direction is expressed as –

$$E(z, t) = I(z)\, e^{j(\omega t - \beta z)} \quad \ldots 3.1.23$$

where,

I(z) is optical field intensity.

ω is optical radian frequency.

β is propagation constant.

The fundamental expression for lasing in Fabry-Perot cavity is –

$$I(z) = I(0)e^{[(\Gamma g(h\,v) - \alpha(h\,v)\}z]} \quad \text{… 3.1.24}$$

where,

□ is optical field confinement factor or the fraction of optical power in the active layer.

α is effective absorption coefficient of material.

g is gain coefficient.

h v is photon energy.

z is distance traverses along the lasing cavity.

The condition of lasing threshold is given as –

- For amplitude : I (2L) = I (0)
- For phase : $e^{-j2\beta L} = 1$
- Optical gain at threshold = Total loss in the cavity.

i.e. $\Gamma g_{th} = \alpha_t$

- Now the lasing expression is reduced to –

$$\Gamma g_{th} = a_t = \alpha + \frac{1}{2L} \ln\left(\frac{1}{R_1 R_2}\right) \quad \text{… 3.1.26}$$

$$\Gamma g_{th} = \alpha_t = \alpha + \alpha_{end} \quad \text{… 3.1.27}$$

where,

A_{end} is mirror loss in lasing cavity.

- An important condition for lasing to occur is that gain, g ≥ g th i.e. threshold gain.

Example 3.1.5 : Find the optical gain at threshold of a laser diode having following parametric values – $R_1 = R_2 = 0.32$, $\alpha = 10\text{cm}^{-1}$ and L = 500 μm.

Solution : Optical gain in laser diode is given by –

$$\Gamma g_{th} = 10 + \frac{1}{2 \times (500 \times 10^{-4})} \ln\left(\frac{1}{0.32 \times 0.32}\right)$$

$$\Gamma g_{th} = \mathbf{33.7\ cm^{-1}}$$

… Ans.

Power Current Characteristics

The output optic power versus forward input current characteristics is plotted in Fig. 3.1.12 for a typical laser diode. Below the threshold current (I_{th}) only spontaneous emission is emitted hence there is small increase in optic power with drive current. at threshold when lasing conditions are satisfied. The optical power increases sharply after the lasing threshold because of stimulated emission.

- The lasing threshold optical gain (g_{th}) is related by threshold current density (J_{th}) for stimulated emission by expression –

$$g_{th} = \beta J_{th} \quad \text{… 3.1.28}$$

where, β is constant for device structure.

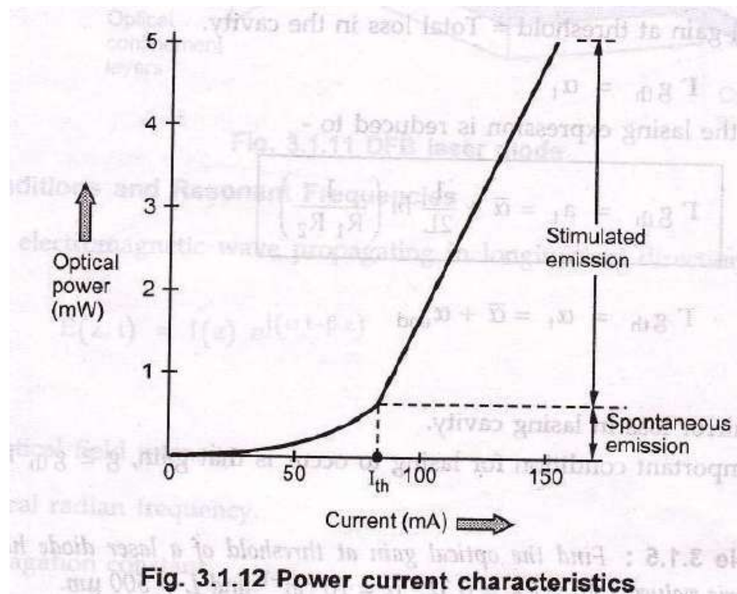


Fig. 3.1.12 Power current characteristics

External Quantum Efficiency

- The external quantum efficiency is defined as the number of photons emitted per electron hole pair recombination above threshold point. The external quantum efficiency η_{ext} is given by –

$$\eta_{ext} = \frac{\eta_i (g_{th} - \alpha)}{g_{th}}$$

… 3.1.29

where,

η_i = Internal quantum efficiency (0.6-0.7).

g_{th} = Threshold gain.

α = Absorption coefficient

- Typical value of η_{ext} for standard semiconductor laser is ranging between 15-20 %.

Resonant Frequencies

- At threshold lasing

$$2\beta L = 2\pi m$$

where, $\beta = \frac{2\pi m}{\lambda}$ (propagation constant)

m is an integer.

∴ $m = 2L.\frac{n}{\lambda}$ ...3.1.30

Since $c = v\lambda$

∴ $\lambda = \frac{c}{v}$

Substituting λ in 3.1.30

$$m = 2L\,\frac{nv}{c}$$

=z… 3.1.31

- Gain in any laser is a function of frequency. For a Gaussian output the gain and frequency are related by expression –

$$g(\lambda) = g(0)e^{\left[-\frac{(\lambda-\lambda_0)^2}{2\sigma^2}\right]} \quad \text{… 3.1.32}$$

where,

g(0) is maximum gain.

λ_0 is center wavelength in spectrum.

- is spectral width of the gain.The frequency spacing between the two successive modes is –

$$\Delta v = \frac{c}{2Ln}$$

$$\Delta \lambda = \frac{\lambda^2}{2Ln} \quad \text{… 3.1.34}$$

Optical Characteristics of LED and Laser

- The output of laser diode depends on the drive current passing through it. At low drive current, the laser operates as an inefficient Led, When drive current crosses threshold value, lasing action beings. Fig. 3.1.13 illustrates graph comparing optical powers of LED operation (due to spontaneous emission) and laser operation (due to stimulated emission).

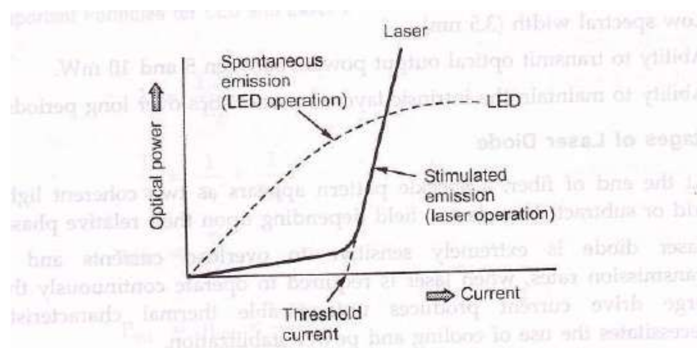


Fig. 3.1.13 Optical characteristics of an LED and laser compared

Spectral and Spatial Distribution of Led and Laser

At low current laser diode acts like normal LED above threshold current, stimulated emission i.e. narrowing of light ray to a few spectral lines instead of broad spectral distribution, exist. This enables the laser to easily couple to single mode fiber and reduces the amount of uncoupled light (i.e. spatial radiation distribution). Fig. 3.1.14 shows spectral and spatial distribution difference between two diodes

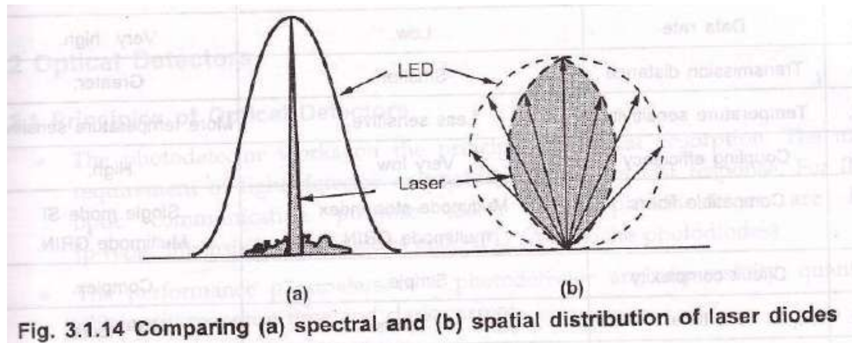


Fig. 3.1.14 Comparing (a) spectral and (b) spatial distribution of laser diodes

Advantages and Disadvantages of Laser Diode

Advantages of Laser Diode

- Simple economic design.
- High optical power.
- Production of light can be precisely controlled.
- Can be used at high temperatures.
- Better modulation capability.
- High coupling efficiency.
- Low spectral width (3.5 nm)
- Ability to transmit optical output powers between 5 and 10 mW.
- Ability to maintain the intrinsic layer characteristics over long periods.

Disadvantages of Laser Diode

- At the end of fiber, a speckle pattern appears as two coherent light beams add or subtract their electric field depending upon their relative phases.
- Laser diode is extremely sensitive to overload currents and at high transmission rates, when laser is required to operate continuously the use of large drive current produces unfavourable thermal characteristics and necessitates the use of cooling and power stabilization.

Comparison of LED and Laser Diode

Sr. No.	Parameter	LED	LD (Laser Diode)
1.	Principle of operation	Spontaneous emission.	Stimulated emission.
2.	Output beam	Non – coherent.	Coherent.
3.	Spectral width	Board spectrum (20 nm – 100 nm)	Much narrower (1-5 nm).

4.	Data rate	Low.	Very high.
5.	Transmission distance	Smaller.	Greater.
6.	Temperature sensitivity	Less sensitive.	More temperature sensitive.
7.	Coupling efficiency	Very low.	High.

8.	Compatible fibers	Multimode step index multimode GRIN.	Single mode Sl Multimode GRIN.
9.	Circuit complexity	Simple	Complex
10.	Life time	10^5 hours.	10^4 hours.
11.	Cost	Low.	High.
12.	Output power	Linearly proportional to drive current.	Proportional to current above threshold.
13.	Current required	Drive current 50 to 100 mA peak.	Threshold current 5 to 40 mA.
14.	Wavelengths available	0.66 to 1.65 μm.	0.78 to 1.65 μm.
15.	Applications	Moderate distance low data rate.	Long distance high data rates.

Important Formulae for LED and Laser

LED

1. $\lambda = \frac{1.24}{E_g}$
2. $\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}}$
3. $\eta_{int} = \frac{\tau}{\tau_r}$
4. $P_{int} = \eta_{int} \text{ x } \frac{hc\,I}{q\lambda}$

LASER

1. $\Gamma g_{th} = \propto + \frac{1}{2L} \ln\left(\frac{1}{R_1 R_2}\right)$
2. $\Delta v = \frac{c}{2\,L\,n}$
3. $\Delta \lambda = \frac{\lambda^2}{2\,L\,n}$

Optical Detectors

Principles of Optical Detectors

- The photodetector works on the principle of optical absorption. The main requirement of light detector or photodector is its fast response. For fiber optic communication purpose most suited photodetectors are PIN (p-type- Instrinsic-n-type) diodes and APD (Avalanche photodiodes)
- The performance parameters of a photodetector are responsivity, quantum efficiency, response time and dark current.

Cut-off Wavelength (λ_c)

- Any particular semiconductor can absorb photon over a limited wavelength range. The highest wavelength is known as cut-off wavelength (λ_c). The cut-off wavelength is determined by bandgap energy E_g of material.

$$\lambda_c = \frac{hc}{E_g} = \frac{1.24}{E_g} \qquad \dots 3.2.1$$

where,

E_g inelectron volts (eV) and

λ_c cut-off wavelength is in μm.

Typical value of λ_c for silicon is 1.06 μm and for germanium it is 1.6 μm.

Quantum Efficiency (η)

- The quantum efficiency is define as the number of electron-hole carrier pair generated per incident photon of energy h v and is given as –

$$\eta = \frac{\text{Number of electron hole pairs generated}}{\text{Number of incident photons}}$$

$$\eta = \frac{I_p / q}{P_{in} / h\nu} \qquad \dots 3.2.2$$

where, I_p is average photocurrent.

P_{in} is average optical power incident on photo detectors.

- Absorption coefficient of material determines the quantum efficiency. Quantum efficiency $\eta < 1$ as all the photons incident will not generate e-h pairs. It is normally expressed in percentage.

Fiber Alignment

5. In any fiber optic communication system, in order to increase fiber length there is need to joint the length of fiber. The interconnection of fiber causes some loss of optical power. Different techniques are used to interconnect fibers. A permanent joint of cable is referred to as **splice** and a temporary joint can be done with the connector.
6. The fraction of energy coupled from one fiber to other proportional to common mode volume M_{common}. The fiber – to – fiber coupling efficiency is given as –

$$\eta_F = \frac{M_{common}}{M_E}$$

where,

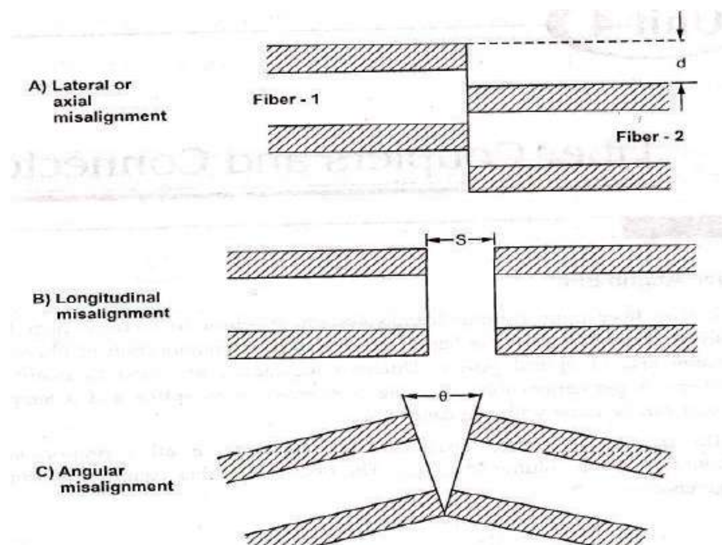
M_E is number of modes in fiber which launches power into next fiber.

- The fiber – to – fiber coupling loss L_F is given as –

$$L_F = -10\log \eta_F$$

Mechanical Misalignment

The diameter of fiber is few micrometer hence the microscopic alignment is required. If the radiation cone of emitting fiber does not match the acceptance cone of receiving fiber, radiation loss takes place. The magnitude of radiation loss depends on the degree of misalignment. Different types of mechanical misalignments are shown in Fig. 4.1.1.



Lateral misalignment

Lateral or axial misalignment occurs when the axes of two fibers are separated by distance 'd'.

Longitudinal misalignment

Longitudinal misalignment occurs when fibers have same axes but their end faces are separated by distance 'S'.

Angular misalignment

Angular misalignment occurs when fiber axes and fiber end faces are no longer parallel. There is an angle 'θ' between fiber end faces.

The axial or lateral misalignment is most common in practice causing considerable power loss. The axial offset reduces the common core area of two fiber end faces as shown in

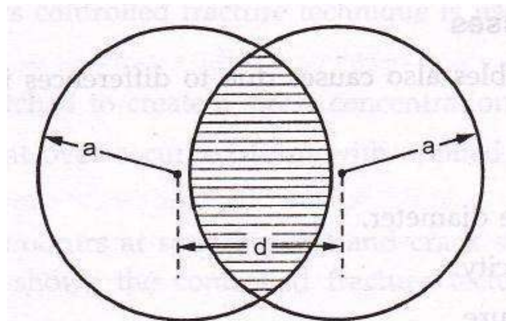


Fig. 4.1.2 Axial offset of fibers

- The optical power coupled is proportional to common area of two fiber cores. The common area is given by expression –

$$A_{common} = 2a^2 \text{ across} \frac{d}{2a} - d\left(a^2 - \frac{d^2}{n}\right)^{1/2} \quad \ldots (4.1.3)$$

where,

a is core radius of fiber.

d is separation of core axes.

- The coupling efficiency for step index fiber is the ratio of common core area to the end-face area.

$$\eta_{step} = \frac{A_{common}}{\pi a^2}$$

$$\eta_{step} = \frac{2}{\pi} \text{ across} \frac{d}{2a} - \frac{d}{\pi a}\left[1 - \left(\frac{d}{2a}\right)^2\right]^{1/2}$$

- For graded index fiber, the total received power for axial misalignment is given by –

.1.5) $$P_T = \frac{2}{\pi} P \left\{ \text{across} \frac{d}{2a} - \left[1 - \left(\frac{d}{2a}\right)^2\right]^{1/2} \frac{d}{6a}\left(5 - \frac{d^2}{2a^2}\right) \right\}$$

where,

P is the power in emitting fiber.

When, d << a, the above expression reduces … (4.1.6)

Fiber Related Losses

- Losses in fiber cables also causes due to differences in geometrical and fiber characteristics.
  These includes,
  1) Variation in core diameter.
  2) Core area ellipticity.
  3) Numerical aperture.
  4) Refractive – index profile.
  5) Core-cladding concentricity.

  The user have less control over these variations since they are related to manufacturing process.

- Coupling loss when emitter fiber radius a_E and receiving fiber radius a_R is not same, is given as –

$$L_F(a) = \begin{cases} -10 \log \left(\frac{a_R}{a_E}\right)^2 & for\ a_R < a_E \\ 0 & for\ a_R \geq a_E \end{cases} \quad (4.1.7)$$

where,

a_E is emitter fiber radius.

a_R is receiver fiber radius.

- Coupling loss when numerical apertures of two fibers are not equal, to expressed as –

$$L_F(NA) = \begin{cases} -10 \log \left[\frac{NA_R\ (0)}{NA_E\ (0)}\right]^2, & for\ NA_R < NA_E \\ 0, & for\ NA_R \geq NA_E \end{cases} \quad (4.1.8)$$

- Coupling loss when core refractive index of two fibers are not same, is expressed as

$$L_F(\propto) = \begin{cases} -10 \log \frac{\propto_R\ (\propto_E + 2)}{\propto_E\ (\propto_R + 2)}, & for\ \propto_R < \propto_E \\ 0, & for\ \propto_R \geq \propto_E \end{cases} \quad (4.1.9)$$

Fiber Splices

- A permanent or semipermanent connection between two individual optical fibers is known as **fiber splice**. And the process of joining two fibers is called as **splicing**.
- Typically, a splice is used outside the buildings and connectors are used to join the cables within the buildings. Splices offer lower attenuation and lower back reflection than connectors and are less expensive.

Types of Splicing

- There are two main types of splicing
  - Fusion splicing.
  - Mechanical splicing / V groove

Fusion Splicing

10. Fusion splicing involves butting two cleaned fiber end faces and heating them until they melt together or fuse.
11. Fusion splicing is normally done with a fusion splicer that controls the alignment of the two fibers to keep losses as low as 0.05 dB.

6. Fiber ends are first prealigned and butted together under a microscope with micromanipulators. The butted joint is heated with electric arc or laser pulse to melt the fiber ends so can be bonded together. Fig. 4.2.1 shows fusion splicing of optical fiber

Mechanical Splicing / V Groove

12. Mechanical splices join two fibers together by clamping them with a structure or by epoxying the fibers together.
13. Mechanical splices may have a slightly higher loss and back reflection. These can be reduced by inserting index matching gel.
14. V groove mechanical splicing provides a temporary joint i.e fibers can be disassembled if required. The fiber ends are butted together in a V – shaped groove as shown in Fig..

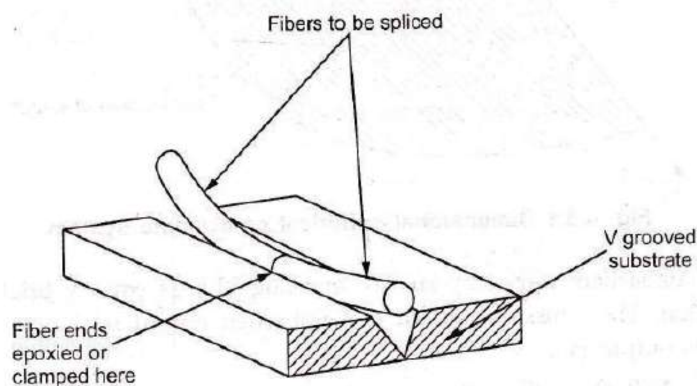


Fig. 4.2.2 V groove optical fiber splicing technique

17. The splice loss depends on fiber size and eccentricity.

Source-to-Fiber Power Launching

Optical output from a source is measured in radiance (B). Radiance is defined as the optical power radiated into a solid angle per unit emitting surface area. Radiance is specified in Watts/cm^2/Steradian. Radiance is important for defining source to fiber coupling efficiency.

Source Output Pattern

10. Spatial radiation pattern of source helps to determine the power accepting capability of fiber.
11. Fig. 4.3.1 shows three dimensional spherical co-ordinate system for characterizing the emission pattern from an optical source. Where the polar axis is normal to the emitting surface and radiance is a function of θ and ϕ.

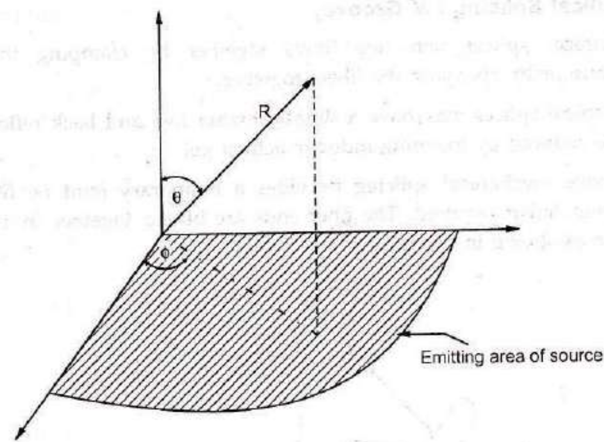


Fig. 4.3.1 Dimensional spherical co-ordinate system

4. The Lambartian output by surface emitting LED is equally bright from any direction. The emission pattern of Lambartian output is shown in Fig. 4.3.2 and its output is –

$$B(\theta,\phi) = B_0 \cos\theta$$

where, B_0 is the radiance along the normal to the radiating surface.

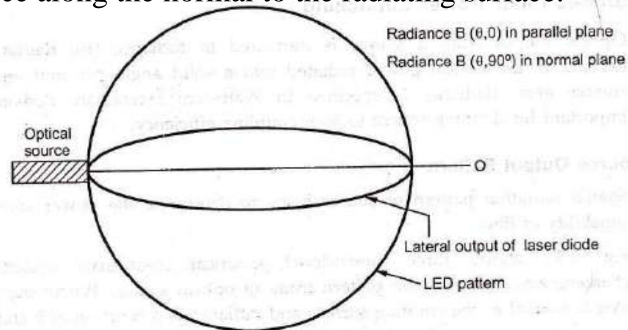


Fig. 4.3.2 Radiance pattern of Lambartian source

- Both radiations in parallel and normal to the emitting plane are approximated by expression –

$$\frac{1}{B(\theta,\phi)} = \frac{\sin^2\phi}{B_0 \cos^T\theta} + \frac{\cos^2\phi}{B_0 \cos^L\theta} \quad \dots (4.3.2)$$

where,

T and L are transverse and lateral power distribution coefficients.

Power Coupling Calculation

- To calculate power coupling into the fiber, consider an optical source launched into the fiber as shown in Fig. 4.3.3.

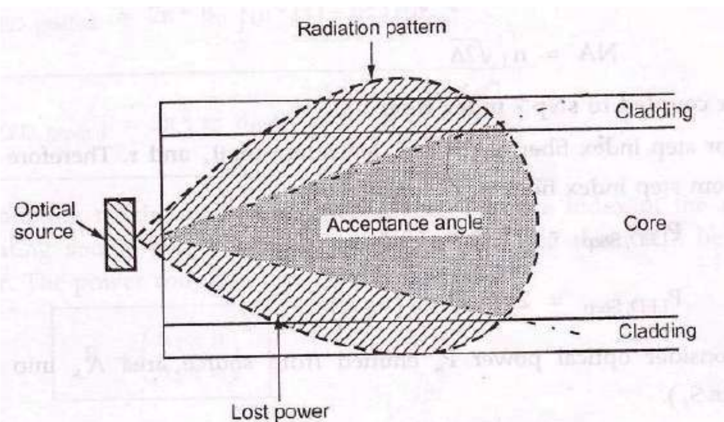


Fig. 4.3.3 Optical source coupled to fiber

Brightness of source is expressed as $B(A_S, \Omega_S)$,

Where, A_S is area of source

Ω_S is solid emission angle of source.

The coupled power P can be calculated as –

$$P = \int_{A_f} d A_s \int_{\Omega_2} d\Omega_s \; B(A_s \Omega_s)$$

$$P = \int_0^r \int_0^{2\pi} \left[\int_0^{2\pi^\phi} \int_0^{0_{max}} B(\theta,\phi) \sin\theta \, d\theta \, d\phi \right] d\theta_s \; r dr \quad \dots (4.3.3)$$

The integral limits are area of source and solid acceptance angle (θ_{0max}).

Here $d\theta_s$ rdr is incremental emitting area.

- Let the radius of surface emitting LED is rs, and for Lambartian emitter, $B(\theta,\phi) = B_0 \cos \theta$, then

$$P = \int_0^{r} \int_0^{2\pi} \left(2\pi\, B_0 \int_0^{0_{max}} \cos\theta \sin\theta \, d\theta\right) d\theta_s \, rdr$$

$$P = B_0 . \pi \int_0^{r_s} \int_0^{2\pi} \sin^2\theta \; 0_{max} \, d\theta_s \, rdr$$

$$P = B_0 . \pi \int_0^{r_s} \int_0^{2\pi} NA^2 \, d\theta_s \, rdr \qquad \dots (4.3.4)$$

Since $$NA = n_1 \sqrt{2\Delta}$$

Power coupled to step – index fiver

- For step index fiber NA is not dependent on θ_s and r. Therefore LED power from step index fiber is,

$$P_{LED}, \text{Step} = \pi^2 r_s^2 B_0 (NA)^2$$

$$P_{LED}, \text{Step} = 2\pi^2 r_s^2 B_0 \,(n_1^2 \,\Delta) \qquad \dots (4.3.5)$$

- Consider optical power P_s emitted from source are A_s into hemisphere(2π S_r).

$$P = A_s \int_0^{2\pi} \int_0^{\pi/2} B(\theta,\phi) \sin\theta \, d\theta \, d\phi$$

$$P = \pi \, r_s^2 2\pi B_o \int_0^{\pi/2} \cos\theta \sin\theta \, d\theta$$

- When source radius r_s <a, the fiber core radius, the LED output power is given from equation (4.3.5).

$$P_s, \text{Step} = P_s (NA)^2 \qquad \dots (4.3.7)$$

- When r_s > a equation (4.3.5) becomes,

$$\boxed{P_{LED}, \text{Step} = \left(\frac{a}{r_s}\right)^2 . P_s \,(NA)^2} \qquad \dots (4.3.8)$$

Power coupled to graded index fiber

- In graded index fiber, the index of refraction varies radially from fiber axis. Numerical aperture for graded index fiber is given by,

$$P_{LED}, Step = 2\pi^2 B_0 \int_0^{r_s} [n^2(r) - n_2^2] r dr$$

Is source radius (r_s) is less than fiber core radius (a) i.e. $r_s < a$, the power coupled from surface emitting LED is given as –

$$P_{LED}, Step = 2\pi^2 r_s^2 B_0 n_1^2 \Delta \left[1 - \frac{2}{\alpha + 2}\left(\frac{r_s}{a}\right)^\alpha\right]$$

- For coupling maximum power to fiber, the refractive index of the medium separating source and fiber must be same, otherwise there will be loss of power. The power couple is reduced by factor,

$$R = \left(\frac{n_1 - n}{n_1 + n}\right)^2$$

where,

n is the refractive index of medium.

n_1 is the refractive index of fiber core. R is the Fresnel reflection or reflectivity

Lensing Schemes for Coupling Improvement

- When When the emitting area of the source is smaller than the core area of fiber, the power coupling efficiency becomes poor. In order to improve the coupling efficiency miniature lens is placed between source and fiber. Microlens magnifies the emitting area of source equal to core area. The power coupled increases by a factor equal to magnification factor of lens.

- Important types of lensing schemes are :
  - Rounded – end fiber.
  - Spherical – surfaced LED and Spherical-ended fiber.
  - Taper ended fiber.
  - Non imaging microsphere.
  - Cylindrical lens,
  - Imaging sphere.

- There are some drawbacks of using lens.

Complexity increases.

Fabrication and handling difficulty.

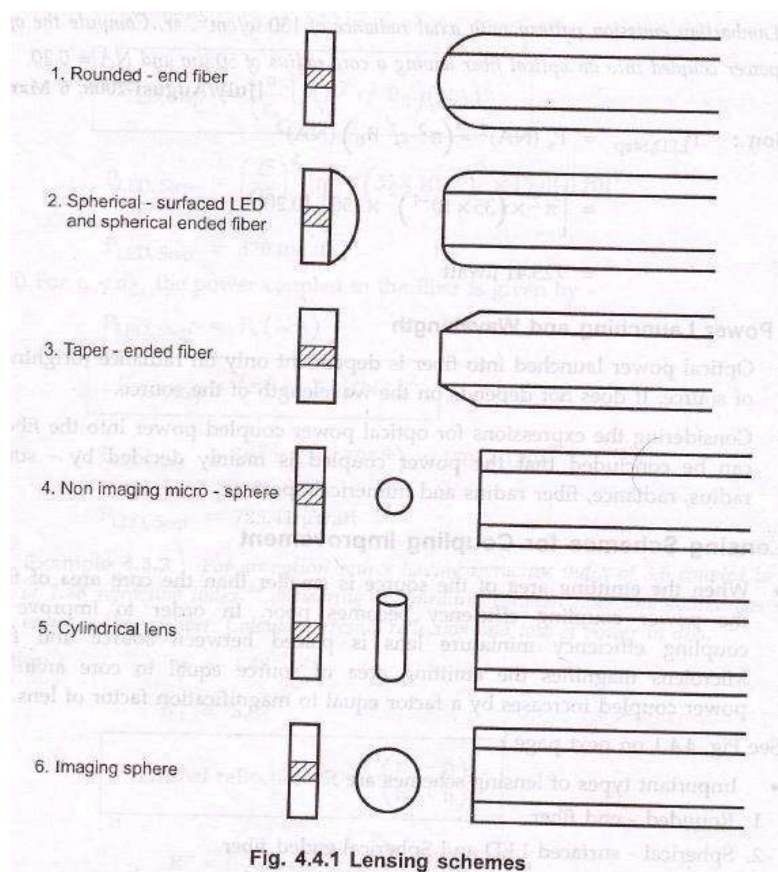


Fig. 4.4.1 Lensing schemes

UNIT-IV

Fiber Optical Receivers

Detector Responsivity ($\Re$)

- The responsivity of a photodetector is the ratio of the current output in amperes to the incident optical power in watts. Responsivity is denoted by $\Re$.

$$\Re = \frac{I_p}{P_{in}} \quad \dots 3.2.3$$

But $$\eta = \frac{I_p - q}{P_{in} - h\nu} = \frac{I_p}{q}\frac{h\nu}{P_{in}}$$

$\therefore$ $$\frac{I_p}{P_{in}} = \frac{\eta q}{h\nu} \quad \dots 3.2.4$$

Therefore $$\Re = \frac{\eta q}{h\nu} = \frac{\eta q^{\lambda}}{h\nu} \qquad \therefore \nu = \frac{c}{\lambda} \quad \dots 3.2.5$$

- Responsivity gives transfer characteristics of detector i.e. photo current per unit incident optical power.
- Typical responsivities of pin photodiodes are –
  Silicon pin photodiode at 900 nm →0.65 A/W.
  Germanium pin photodiode at 1.3 µm →0.45 A/W.
  In GaAs pin photodiode at 1.3 µm →0.9 A/W.

$$\eta = \frac{5.4 \times 10^6}{6 \times 10^6}$$

$$\eta = 0.9 = \mathbf{90\ \%} \quad \textbf{... Ans.}$$

- r photodetectors are sued. As the intensity of optical signal at the receiver is very low, the detector has to meet high performance specifications.
  The conversion efficiency must be high at the operating wavelength.

The speed of response must be high enough to ensure that signal distortion does not occur

The detection process introduce the minimum amount of noise.

It must be possible to operate continuously over a wide range of temperatures for many years.

The detector size must be compatible with the fiber dimensions.

4. At present, these requirements are met by reverse biased p-n photodiodes. In these devices, the semiconductor material absorbs a photon of light, which excites an electron from the valence band to the conduction band (opposite of photon emission). The photo
5. generated electron leaves behind it a hole, and so each photon generates two charge carriers. The increases the material conductivity so call **photoconductivity** resulting in an increase in the diode current. The diode equation is modified as –

$$I_{diode} = (I_d + I_s)\left(e^{V_q/\eta kT} - 1\right) \quad \ldots 3.2.6$$

where,

I_d is dark current i.e. current that flows when no signal is present.

I_s is photo generated current due to incident optical signal.

Fig. 3.2.1 shows a plot of this equation for varying amounts of incident optical power.

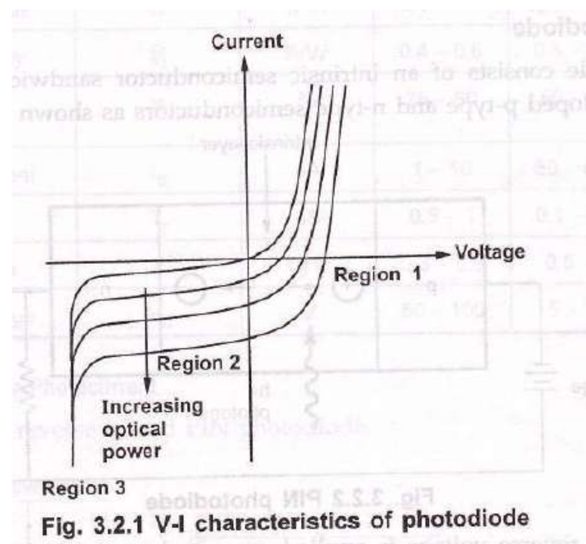


Fig. 3.2.1 V-I characteristics of photodiode

- Three regions can be seen forward bias, reverse bias and avalanche breakdown.

Forward bias, region 1 : A change in incident power causes a change in terminal voltage, it is called as **photovoltaic mode**. If the diode is operated in this mode, the frequency response of the diode is poor and so photovoltaic operation is rarely used in optical links

Reverse bias, region 2 : A change in optical power produces a proportional change in diode current, it is called as **photoconductive mode** of operation which most detectors use. Under these condition, the exponential term in equation 3.2.6 becomes insignificant and the reverse bias current is given by –

$$I_{diode} = -(I_L + I_s)$$

- **Responsivity** of photodiode is defined as the change in reverse bias current per unit change in optical powr, and so efficient detectors need large responsivities.

Avalanche breakdown, region 3 : When biased in this region, a photo generated electron-hole pair causes avalanche breakdown, resulting in large diode for a single incident photon. Avalance photodiodes (APDs) operate in this region APDs exhibit carrier multiplication. They are usually very sensitive detectors. Unfortunately V-I characteristic is very steep in this region and so the bias voltage must be tightly controlled to prevent spontaneous breakdown.

PIN Photodiode

- PIN diode consists of an intrinsic semiconductor sandwiched between two heavily doped p-type and n-type semiconductors as shown in Fig. 3.2.2

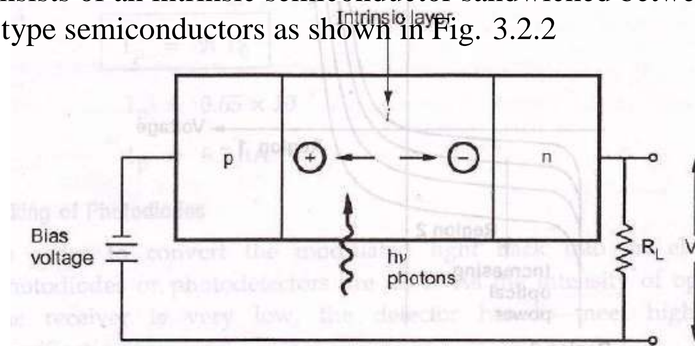


Fig. 3.2.2 PIN photodiode

- Sufficient reverse voltage is applied so as to keep intrinsic region free from carries, so its resistance is high, most of diode voltage appears across it, and the electrical forces are strong within it. The incident photons give up their energy and excite an electron from valance to conduction band. Thus a free electron hole pair is generated, these are

as **photocarriers**. These carriers are collected across the reverse biased junction resulting in rise in current in external circuit called **photocurrent**.

In the absence of light, PIN photodiodes behave electrically just like an ordinary rectifier diode. If forward biased, they conduct large amount of current

PIN detectors can be operated in two modes : **Photovoltaic** and **photoconductive**. In photovoltaic mode, no bias is applied to the detector. In this case the detector works very slow, and output is approximately logarithmic to the input light level. Real world fiber optic receivers never use the photovoltaic mode.

In photoconductive mode, the detector is reverse biased. The output in this case is a current that is very linear with the input light power.

The intrinsic region some what improves the sensitivity of the device. It does not provide internal gain. The combination of different semiconductors operating at different wavelengths allows the selection of material capable of responding to the desired operating wavelength.

Characteristics of common PIN photodiodes

Sr. No.	Parameters	Symbol	Unit	Si	Ge	InGaAs
1.	Wavelength	λ	μ m	0.4 – 1.1	0.8 – 1.8	1.0– 1.7
2.	Reponsivity	$\Re$	A/W	0.4 – 0.6	0.5 – 0.7	0.6– 0.9
3.	Quantum efficiency	H	%	75 -90	50 – 55	60– 70
4.	Darl current	I_d	nA	1 – 10	50 – 500	1 - 20
5.	Rise time	T_r	nS	0.5 – 1	0.1 – 0.5	0.02 – 0.5
6.	Bandwidth	B	GHz	0.3 – 0.6	0.5 – 3	1 – 10
7.	Bias voltage	V_b	V	50 – 100	5 – 10	5 - 6

Depletion Layer Photocurrent

ιϖ) Consider a reverse biased PIN photodiode.

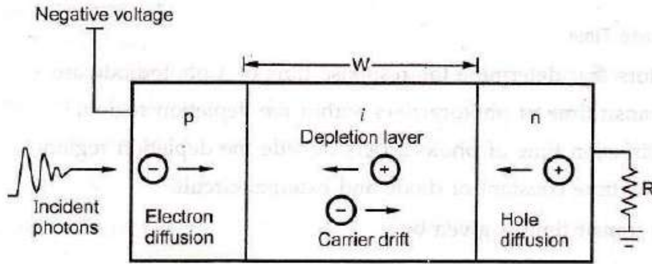


Fig. 3.2.3 Reverse biased PIN diode

- The total current density through depletion layer is –

$$J_{tot} = J_{dr} + J_{diff} \quad \text{… 3.2.7}$$

Where,

J_{dr} is drift current densioty due to carriers generated in depletion region.

J_{diff} is diffusion current density due to carriers generated outside depletion region.

- The drift current density is expressed as –

$$J_{dr} = \frac{I_p}{A}$$
$$J_{dr} = q\,\phi_0\,(1 - e^{-\alpha_s w})$$

where,

A is photodiode area.

ϕ_0 is incident photon flux per unit area.

- The diffusion current density is expressed as –

$$J_{diff} = q\,\phi_0 \frac{\alpha_s L_p}{1+\alpha_s L_p} e^{-\alpha_s w} + qP_{n0}\frac{D_p}{L_p} \quad \text{… 3.2.}$$

where,

D_p is hole diffusion coefficient

P_n is hole concentration in n-type material.

P_{n0} is equilibrium hole density.

Substituting in equation 3.2.7, total current density through reverse biased depletion layer is –

$$J_{tot} = q\,\phi_0 \left[1 - \frac{e^{-\alpha_s W}}{1+\alpha_s L_p}\right] + qP_{n0}\frac{D_p}{L_p}$$

Response Time

- Factors that determine the response time of a photodiode are –
  - Transit time of photocarriers within the depletion region.
  - Diffusion time of photocarriers outside the depletion region.
  - RC time constant of diode and external circuit.
- The transit time is given by –

$$t_d = \frac{w}{v_d}$$

ι) The diffusion process is slow and diffusion times are less than carrier drift time. By considering the photodiode response time the effect of diffusion can be calculated. Fig. 3.2.4 shows the response time of photodiode which is not fully depleted.

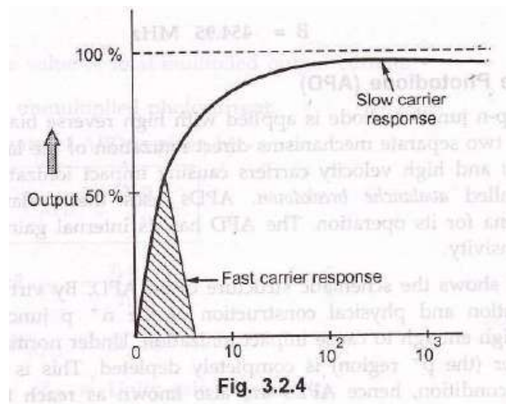


Fig. 3.2.4

- The detector behaves as a simple low pass RC filter having passband of

$$N = \frac{1}{2\pi R_R C_T}$$

where

R_T, is combination input resistance of load and amplifier.

C_T is sum of photodiode and amplifier capacitance.

Example 3.2.5 : Compute the bandwidth of a photodetector having parameters as –

Photodiode capacitance = 3 pF

Amplifier capacitance = 4 pF

Load resistance = 50 Ω

Amplifier input resistance = 1 MΩ

Solution : Sum of photodiode and amplifier capacitance

$$C_T = 3 + 4 = 7 \text{ pF}$$

Combination of load resistance and amplifier and input resistance

$$R_T = 50\Omega \parallel 1\text{ M}\Omega \approx 50\ \Omega$$

$$\text{Bandwidth of photodetector } B = \frac{1}{2\pi R_R C_T}$$

$$B = \frac{1}{2\pi\ 50 \times 7 \times 10^{-12}}$$

$$B = 454.95 \text{ MHz}$$ **… Ans.**

Avalanche Photodiode (APD)

- When a p-n junction diode is applied with high reverse bias breakdown can occur by two separate mechanisms direct ionization of the lattice atoms, zener breakdown and high velocity carriers impact ionization of the lattice atoms called avalanche breakdown. APDs uses the avalanche breakdown phenomena for its operation. The APD has its internal gain which increases its responsivity.
  Fig. 3.2.5 shows the schematic structure of an APD. By virtue of the doping concentration and physical construction of the n^+ p junction, the electric filed is high enough to cause impact ionization. Under normal operating bias, the I-layer (the p^- region) is completely depleted. This is known as **reach through** condition, hence APDs are also known as **reach through APD** or **RAPDs**

- Similar to PIN photodiode, light absorption in APDs is most efficient in I-layer. In this region, the E-field separates the carriers and the electrons drift into the avalanche region where carrier multiplication occurs. If the APD is biased close to breakdown, it will result

in reverse leakage current. Thus APDs are usually biased just below breakdown, with the bias voltage being tightly controlled.

- The multiplication for all carriers generated in the photodiode is given as –

$$M = \frac{I_M}{I_P} \quad \ldots$$

where,

I_M = Average value of total multiplied output current.

I_P = Primary unmultiplied photocurrent.

- Responsivity of APD is given by –

$$\boxed{\Re_{APD} = \frac{\eta\, q}{h\, v} M}$$

$$\Re_{APD} = \frac{\eta\, q\, \lambda}{h\, v} M \qquad \because\ v = \frac{c}{\lambda}$$

$$\Re_{APD} = \Re_0\, M \quad \ldots$$

where, $\Re_0$ = Unity gain responsivity

MSM Photodetector

- Metal-semiconductor-metal (MSM) photodetector uses a sandwiched semiconductor between two metals. The middle semiconductor layer acts as optical absorbing layer. A Schottky barrier is formed at each metal semiconductor interface (junction), which prevents flow of electrons.
- When optical power is incident on it, the electron-hole pairs generated through photo absorption flow towards metal contacts and causes photocurrent.
- MSM photodetectors are manufactured using different combinations of semiconductors such as – GaAs, InGaAs, InP, InAIAs. Each MSM photodetectors had distinct features e.g. responsivity, quantum efficiency, bandwidth etc.

- With InAIAs based MSM photodetector, 92 % quantum efficiency can be obtained at 1.3 μm with low dark current. An inverted MSM photodetector shows high responsivity when illuminated from top.
- A GaAs based device with travelling wave structure gives a bandwidth beyond 500 GHz.

Optical Detector

- With a proper sketch briefly explain the structure of PIN diode.
- Explain the following term relating to PIN photodiode with proper expressions.
  - Cut-off wavelength.
  - Quantum efficiency.
  - Responsivity.
- Explain the structure and principle of working of APD.
- Deduce the expression for total current density for APD.
- How the response time of APD is estimated?
- Give expression for passband of APD detector.
- Compare the performance parameters of PIN and APD.

UNIT-V
DESIGN OF DIGITAL SYSTEMS

Digital Links:

System Design Considerations:

- In optical system design major consideration involves
  1. Transmission characteristics of fiber (attenuation & dispersion).
  2. Information transfer capability of fiber.
  3. Terminal equipment & technology.
  4. Distance of transmission.
- In long-haul communication applications repeaters are inserted at regular intervals as shown in Fig. 6.2.1

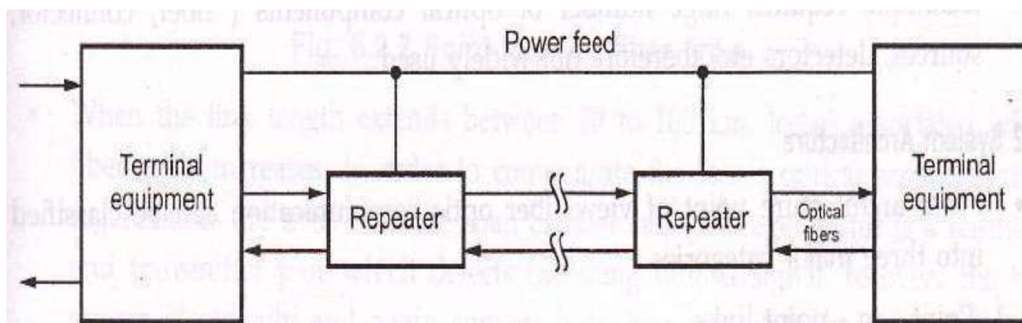


Fig. 6.2.1 Repeaters in long-haul communication system

- Repeater regenerates the original data before it is retransmitted as a digital optical signal. The cost of system and complexity increases because of installation of repeaters.
- An optical communication system should have following basic required specifications –

7. Transmission type (Analog / digital).
8. System fidelity (SNR / BER)
9. Required transmission bandwidth
10. Acceptable repeater spacing
11. Cost of system
12. Reliability
13. Cost of maintenance.

Multiplexing

- Multiplexing of several signals on a single fiber increases information transfer rate of communication link. In Time Division Multiplexing (TDM) pulses from multiple channels are interleaved and transmitted sequentially, it enhance the bandwidth utilization of a single fiber link. In Frequency Division Multiplexing (FDM) the optical

channel bandwidth is divided inot various nonoverlapping frequency bands and each signal is assigned one of these bands of frequencies. By suitable filtering the combined FDM signal can be retrieved.

- When number of optical sources operating at different wavelengths are to be sent on single fiber link Wavelength Division Multiplexing (WDM) is used. At receiver end, the separation or extraction of optical signal is performed by optical filters (interference filters, differaction filters prism filters).
- Another technique called Space Division Multiplexing (SDM) used separate fiber within fiber bundle for each signal channel. SDM provides better optical isolation which eliminates cross-coupling between channels. But this technique requires huge number of optical components (fiber, connector, sources, detectors etc) therefore not widely used.

System Architecture

- From architecture point of view fiber optic communication can be classified into three major categories.
  - Point – to – point links
  - Distributed networks
  - Local area networks.

Point-to-Point Links:

- A point-to-point link comprises of one transmitter and a receiver system. This is the simplest form of optical communication link and it sets the basis for examining complex optical communication links.
- For analyzing the performance of any link following important aspects are to be considered.
  - Distance of transmission
  - Channel data rate
  - Bit-error rate
- All above parameters of transmission link are associated with the characteristics of various devices employed in the link. Important components and their characteristics are listed below.
- When the link length extends between 20 to 100 km, losses associated with fiber cable increases. In order to compensate the losses optical amplifier and regenerators are used over the span of fiber cable. A regenerator is a receiver and transmitter pair which detects incoming optical signal, recovers the bit stream electrically and again convert back into optical from by modulating an optical source. An optical amplifier amplify the optical bit stream without converting it into electrical form.
- The spacing between two repeater or optical amplifier is called as repeater spacing (L). The repeater spacing L depends on bit rate B. The bit rate-distance product (BL) is a measure of system performance for point-to-point links.

Two important analysis for deciding performance of any fiber link are –

- Link power budget / Power budget
- Rise time budget / Bandwidth budget

- The Link power budget analysis is used to determine whether the receiver has sufficient power to achieve the desired signal quality. The power at receiver is the transmitted power minus link losses.

- The components in the link must be switched fast enough and the fiber dispersion must be low enough to meet the bandwidth requirements of the application. Adequate bandwidth for a system can be assured by developing a rise time budget.

System Consideration:

- Before selecting suitable components, the operating wavelength for the system is decided. The operating wavelength selection depends on the distance and attenuation. For shorter distance, the 800-900 nm region is preferred but for longer distance 100 or 1550 nm region is preferred due to lower attenuations and dispersion.
- The next step is selection of photodetector. While selecting a photodetector following factors are considered –
  - Minimum optical power that must fall on photodetector to satisfy BER at specified data rate.
  - Complexity of circuit.
  - Cost of design.
  - Bias requirements.
- Next step in system consideration is choosing a proper optical source, important factors to consider are –
  - Signal dispersion.
  - Data rate.
  - Transmission distance.
  - Cost.
  - Optical power coupling.
  - Circuit complexity.
- The last factor in system consideration is to selection of optical fiber between single mode and multimode fiber with step or graded index fiber. Fiber selection depends on type of optical source and tolerable dispersion. Some important factors for selection of fiber are :
  - Numerical Aperture (NA), as NA increases, the fiber coupled power increases also the dispersion.
  - Attenuation characteristics.
  - Environmental induced losses e.g. due to temperature variation, moisture and dust etc.

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

EC3552-VLSI AND CHIP DESIGN

Semester - 05

Notes

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

Vision

To excel in providing value based education in the field of Electronics and Communication Engineering, keeping in pace with the latest technical developments through commendable research, to raise the intellectual competence to match global standards and to make significant contributions to the society upholding the ethical standards.

Mission

- ✓ To deliver Quality Technical Education, with an equal emphasis on theoretical and practical aspects.
- ✓ To provide state of the art infrastructure for the students and faculty to upgrade their skills and knowledge.
- ✓ To create an open and conducive environment for faculty and students to carry out research and excel in their field of specialization.
- ✓ To focus especially on innovation and development of technologies that is sustainable and inclusive, and thus benefits all sections of the society.
- ✓ To establish a strong Industry Academic Collaboration for teaching and research, that could foster entrepreneurship and innovation in knowledge exchange.
- ✓ To produce quality Engineers who uphold and advance the integrity, honour and dignity of the engineering.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

1. To provide the students with a strong foundation in the required sciences in order to pursue studies in Electronics and Communication Engineering.
2. To gain adequate knowledge to become good professional in electronic and communication engineering associated industries, higher education and research.
3. To develop attitude in lifelong learning, applying and adapting new ideas and technologies as their field evolves.
4. To prepare students to critically analyze existing literature in an area of specialization and ethically develop innovative and research oriented methodologies to solve the problems identified.
5. To inculcate in the students a professional and ethical attitude and an ability to visualize the engineering issues in a broader social context.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO1: Design, develop and analyze electronic systems through application of relevant electronics, mathematics and engineering principles.

PSO2: Design, develop and analyze communication systems through application of fundamentals from communication principles, signal processing, and RF System Design & Electromagnetics.

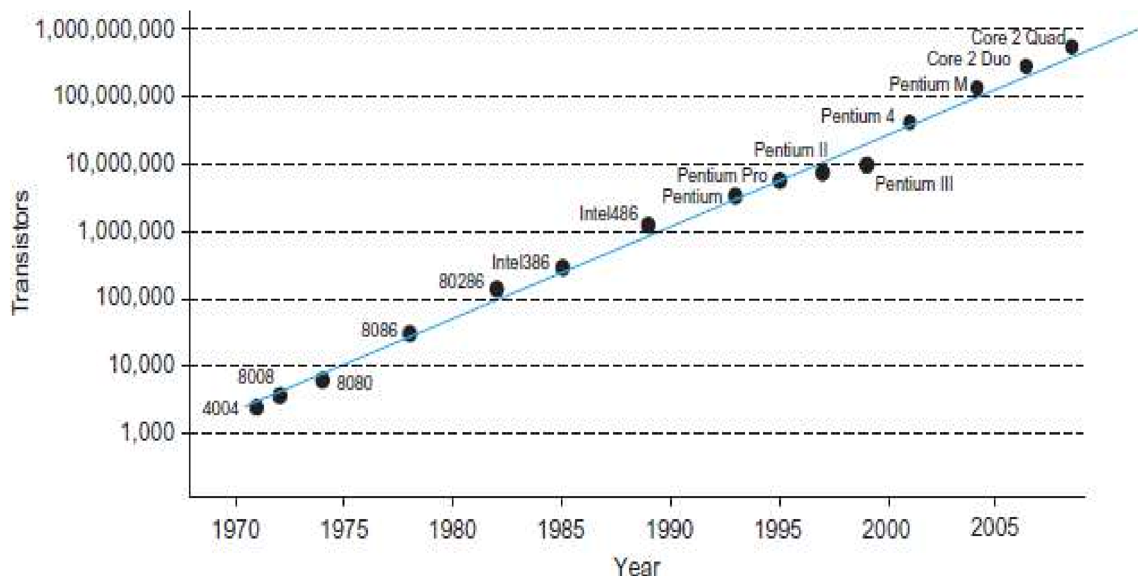
PSO3: Adapt to emerging electronics and communication technologies and develop innovative solutions for existing and newer problems.

UNIT I - MOS TRANSISTOR PRINCIPLES

MOS logic families (NMOS and CMOS), Ideal and Non Ideal IV Characteristics, CMOS devices, MOS(FET) Transistor Characteristic under Static and Dynamic Conditions, Technology Scaling, power consumption

1.1 : INTRODUCTION: (VLSI)

- ✓ In 1958, Jack Kilby built the first integrated circuit flip-flop at Texas Instruments.
- ✓ Bell Labs developed the bipolar junction transistor. Bipolar transistors were more reliable, less noisy and more power-efficient.
- ✓ In 1960s, Metal Oxide Semiconductor Field Effect Transistors (MOSFETs) began to enter in the production.
- ✓ MOSFETs offer the compelling advantage that; they draw almost zero control current while idle.
- ✓ They come in two flavors: nMOS and pMOS, using n-type and p-type silicon respectively.
- ✓ In 1963, Frank Wanlass at Fairchild described the first logic gates using MOSFETs. Fairchild's gates used both nMOS and pMOS transistors, naming as Complementary Metal Oxide Semiconductor (CMOS).
- ✓ Power consumption became a major issue in the 1980s as hundreds of thousands of transistors were integrated onto a single die.
- ✓ CMOS processes were widely adopted and replaced nMOS and bipolar processes for all digital logic applications.
- ✓ In 1965, Gordon Moore observed that plotting the number of transistors that can be most economically manufactured on a chip gives a straight line on a semi logarithmic scale.



- ○ ***Moore's Law* is defined as transistor count doubling every 18 months.**

The level of integration of chips is classified as

- Small Scale Integration (SSI)
- Medium Scale Integration (MSI)
- Large Scale Integration (LSI)

- Very Large Scale Integration (VLSI)
- Ultra Large Scale Integration (ULSI)

Small scale Integration:

- ✓ *Small-Scale Integration* (SSI) circuits have less than 10 gates. Example: 7404 inverter.

Medium scale Integration:

- ✓ *Medium-Scale Integration* (MSI) circuits have up to 1000 gates. Example: 74161 counter.

Large scale Integration:

- ✓ *Large-Scale Integration* (LSI) circuits have up to 10,000 gates. Example: 8-bit microprocessor (8085).

Very large scale Integration:

- ✓ Very large scale Integration (VLSI) with gates counting up to lakhs. Example: 16-bit microprocessor (8086).
- ✓ The feature size of a CMOS manufacturing process refers to the minimum dimension of a transistor that can be reliably built.

Ultra large scale Integration:

- ✓ Ultra Large-Scale Integration (ULSI) is the process of integrating millions of transistors on a single silicon semiconductor microchip.

****

1.2 : MOS Transistor

nMOS and pMOS transistor

Explain the basic concept of nMOS and pMOS transistor with relevant symbol.

- ✓ A Metal-Oxide-Semiconductor (MOS) structure is created by superimposing layers of conducting and insulating materials.
- ✓ CMOS technology provides two types of transistors. They are n-type transistor (nMOS) and p-type transistor (pMOS).
- ✓ As transistor operation is controlled by electric fields, the devices are also called Metal Oxide Semiconductor Field Effect Transistors (MOSFETs).
- ✓ The transistor consists of a stack of the conducting gate, an insulating layer of silicon dioxide (SiO_2) and the silicon wafer, also called as substrate, body or bulk.
- ✓ A pMOS transistor consists of p-type source and drain region with an n-type body.
- ✓ An nMOS transistor consists of n-type source and drain region with a p-type body.

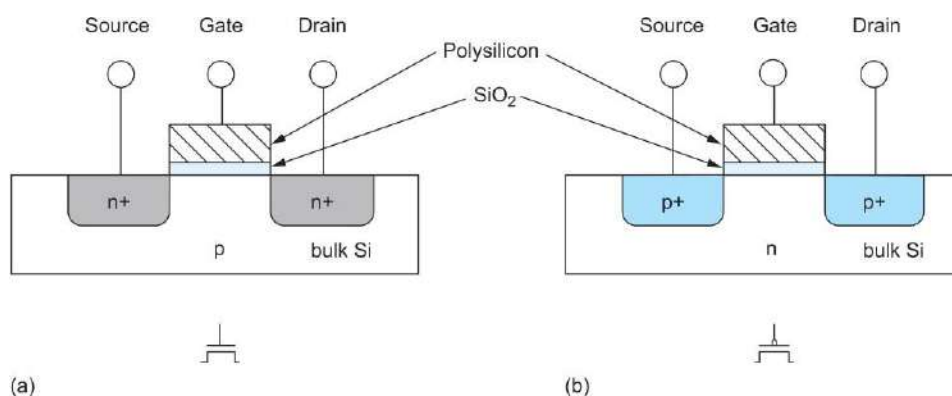


Figure 1: (a) n-MOS transistor (b) p-MOS transistor

nMOS Transistor:

- ✓ In an nMOS transistor, the body is grounded and the p–n junction of the source and drain to body are reverse-biased.
- ✓ As the gate is grounded, no current flows through junction. Hence, the transistor is OFF.
- ✓ If the gate voltage is raised, it creates an electric field, that start to attract free electrons to the underside of the Si–SiO_2 interface.
- ✓ If the voltage is raised more, a thin region under the gate called the channel is inverted.
- ✓ Since a conducting path of electron carriers is formed from source to drain, current starts to flow. So, the transistor is said to be ON.

pMOS Transistor:

- ✓ For a pMOS transistor, the body is held at a positive voltage.
- ✓ When the gate terminal has a positive voltage, the source and drain junctions are reverse-biased and no current flows. So, the transistor is said to OFF.
- ✓ When the gate voltage is lowered, positive charges are attracted to the underside of the Si–SiO2 interface.
- ✓ When a sufficient low gate voltage is applied, the channel inverts and a conducting path of positive carriers is formed from source to drain, which makes the transistor ON.

NOTE:

- ✓ The symbol for the pMOS transistor has a bubble on the gate, indicating that the transistor behavior is opposite to nMOS.
- ✓ When the gate of an nMOS transistor is 1, the transistor is ON. When the gate is 0, the nMOS transistor is OFF.
- ✓ A pMOS transistor is ON when the gate is low(0) and OFF when the gate is high(1).

1.3 : Modes of MOS TRANSISTOR

Explain the accumulation (Enhancement) mode, depletion layer and inversion layer of MOS transistor with diagram.

- ✓ The MOS transistor is a majority-carrier device, in which the current in a conducting channel is controlled by gate voltage.
- ✓ In an nMOS transistor, the majority carriers are electrons.
- ✓ In a pMOS transistor, the majority carriers are holes.
- ✓ Figure 2 shows a simple MOS structure. The top layer of the structure is a good conductor called the gate.
- ✓ Transistor gate is polysilicon, i.e., silicon formed from many small crystals. The middle layer is a very thin insulating film of SiO_2, called the gate oxide. The bottom layer is the doped silicon body.
- ✓ The figure 2 shows a p-type body, in which the carriers are holes. The body is grounded and voltage is applied to the gate.
- ✓ The gate oxide is a good insulator, so almost zero current flows from the gate to the body.

Accumulation (Enhancement) mode:

- ✓ In Figure 2(a), when a negative voltage is applied to the gate, negative charges are formed on the gate.

- ✓ The positively charged holes are attracted to the region under the gate. This is called the accumulation mode.

Depletion mode:

- ✓ In Figure 2(b), when a small positive voltage is applied to the gate, positive charges are formed on the gate.
- ✓ The holes in the body are repelled from the region directly under the gate, resulting in a depletion region forming below the gate.

Inversion layer:

- ✓ In Figure 2(c), when a higher positive potential greater than threshold voltage (V_t) is applied, more positive charges are attracted to the gate.
- ✓ The holes are repelled and some free electrons in the body are attracted to the region under 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.

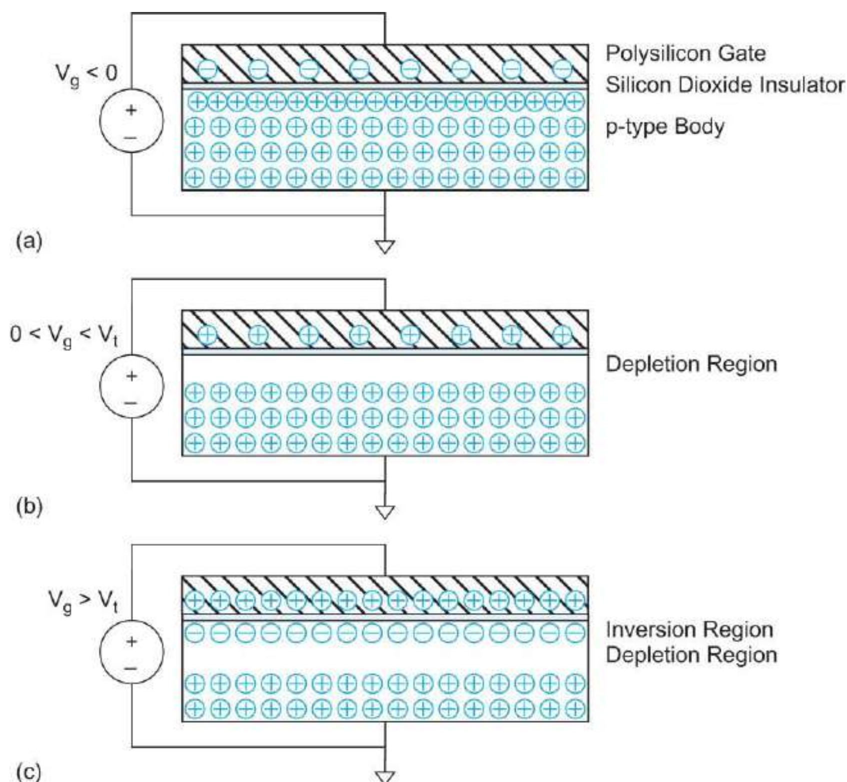


Figure 2: MOS structure demonstrating (a) accumulation, (b) depletion, and (c) inversion layer

1.4 : Operating regions of MOS transistor:

Draw the small signal model of device during cut-off, linear and saturation. (April 2018)
Discuss the cutoff, linear and saturation region operation of MOS transistor. (Nov 2009)

- ✓ The MOS transistor operates in cutoff region, linear region and saturation region.

Cutoff region:

- ✓ In Figure 3(a), the gate-to-source voltage (V_{gs}) is less than the threshold voltage (V_t) and source is grounded.
- ✓ Junctions between the body and the source or drain are reverse biased, so no current flows. Thus, the transistor is said to be OFF and this mode of operation is called cutoff.
- ✓ If $V_{gs} < V_t$, the transistor is cutoff (OFF).

Linear (Active) Region:

- ✓ In Figure 3(b), the gate voltage is greater than the threshold voltage.
- ✓ An inversion region of electrons, called the channel connects the source and drain, creating a conductive path and making the transistor ON.
- ✓ If $V_{gs} > V_t$, the transistor turns ON. If V_{ds} is small, the transistor acts as a linear resistor, in which the current flow is proportional to V_{ds}.
- ✓ The number of carriers and the conductivity increases, with the gate voltage.

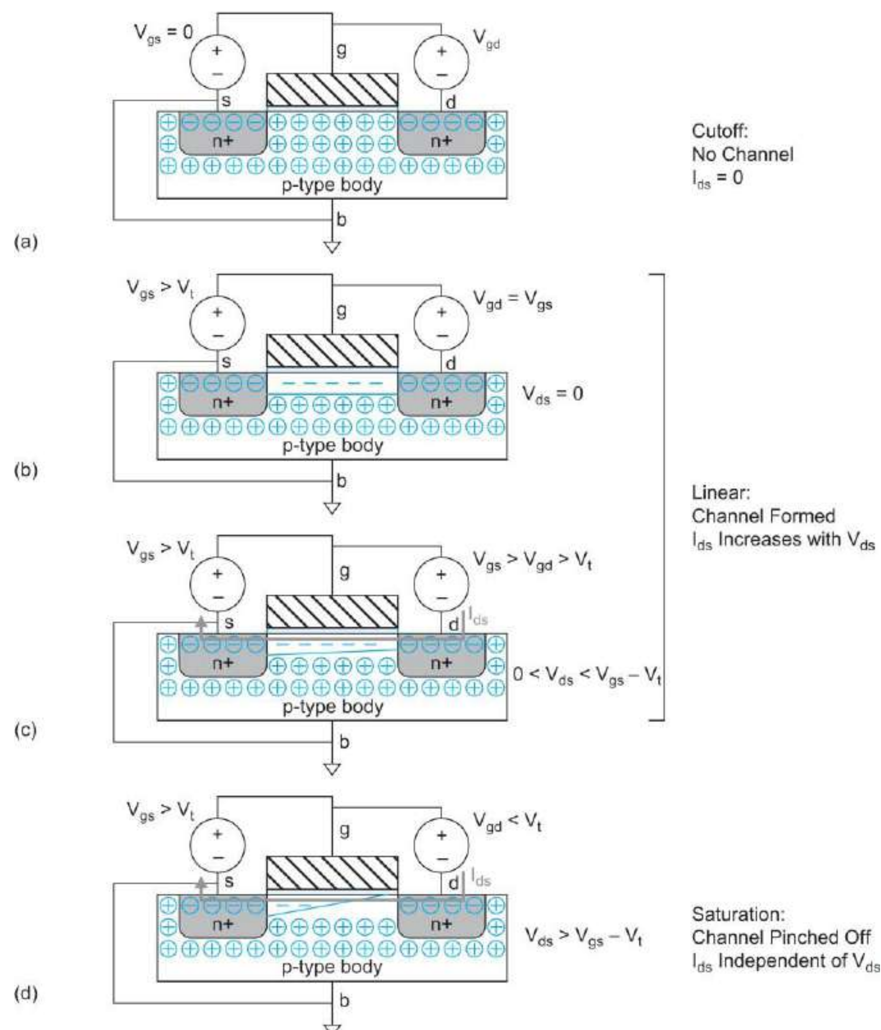


Figure 3: nMOS transistor demonstrating cutoff, linear, and saturation regions of operation

- ✓ The voltage between drain and source is $V_{ds} = V_{gs} - V_{gd}$. If $V_{ds} = 0$ (i.e., $V_{gs} = V_{gd}$), there is no electric field to push current from drain to source.
- ✓ When a small positive voltage V_{ds} is applied to the drain (Figure 3(c)), current I_{ds} flows through the channel from drain to source.
- ✓ This mode of operation is termed as linear, resistive, triode, nonsaturated, or unsaturated.

Saturation region:

- ✓ The current increases with increase in both the drain voltage and gate voltage.

- ✓ If V_{ds} becomes sufficiently large that $V_{gd} < V_t$, the channel is no longer inverted near the drain and becomes pinched off (Figure 3(d)).
- ✓ As electrons reach the end of the channel, they are injected into the depletion region near the drain and accelerated toward the drain.
- ✓ Above this drain voltage, current I_{ds} are controlled only by the gate voltage. This mode is called saturation.
- ✓ If $V_{gs} > V_t$ and V_{ds} is large, the transistor acts as a current source, in which the current flow becomes independent of V_{ds}.

Explain the three different types of modes of operation of pMOS transistor.

- ✓ The pMOS transistor in Figure 4 operates in just the opposite fashion. The n-type body is tied to high potential, junctions of p-type source and drains are normally reverse-biased.
- ✓ When the gate has high potential, no current flows between drain and source.
- ✓ When the gate voltage is lowered by a threshold V_t, holes are attracted to form a p-type channel beneath the gate, allowing current to flow between drain and source.

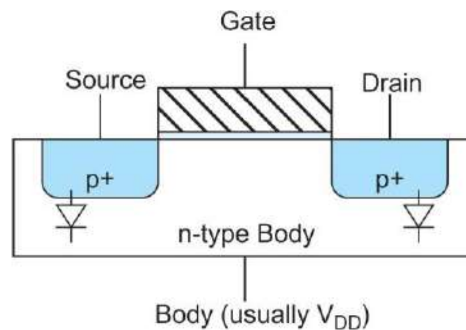


Figure 4: pMOS transistor

****

1.5 : IDEAL I-V CHARACTERISTICS OF MOS TRANSISTOR

- ❖ **Derive an expression for I_{ds} of nMOS in linear and saturated region. (April 2019-6M)**
- ❖ **Derive an expression to show the drain current of MOS for various operating region. Explain one non-ideality for each operating region that changes the drain current. (NOV 2018)**
- ❖ **Explain the dynamic behavior of MOSFET transistor with neat diagram. (April 2018)**
- ❖ **Explain the electrical properties CMOS. (Nov 2017)**
- ❖ **Explain in detail about the ideal I-V characteristics of a NMOS and PMOS device. (MAY 2013)**
- ❖ **Discuss in detail with necessary equations the operation of MOSFET and its current-voltage characteristics. (April/May 2011, May 2016).**
- ❖ **Derive drain current of MOS device in different operating regions. (Nov/Dec 2014)(May/June 2013) (Nov 2012, Nov 2016)**
- ❖ **Explain in detail about the ideal I-V characteristics and non-ideal I-V characteristics of a NMOS and PMOS device. (May/June 2013)**
- ❖ **Derive expressions for the drain-to-source current in the nonsaturated and saturated regions of operation of an nMOS transistor. (Nov 2007, Nov 2008) [April / May – 2023]**

- ✓ MOS transistor has three regions of operation:
  - Cutoff (or) sub threshold region
  - Linear region (or) Non saturation region

- Saturation region

✓ The current through an OFF transistor is zero. When a transistor turns ON ($V_{gs} > V_t$), the gate attract electrons to form a channel.

✓ Current is measured from the amount of charge in the channel.

✓ The charge on each plate of a capacitor is $Q = CV$. Thus, the charge in the channel $Q_{channel}$ is

$$Q_{channel} = C_g (V_{gc} - V_t)$$

where C_g : Capacitance of the gate to the channel

$V_{gc} - V_t$: Amount of voltage attracting charge to the channel.

✓ If the source is at V_s and the drain is at V_d ,

✓ Average channel voltage is $V_c = (V_s + V_d)/2 = V_s + V_{ds} /2$.

✓ Gate and channel voltage V_{gc} is $V_g – V_c = V_{gs} – V_{ds} /2$,

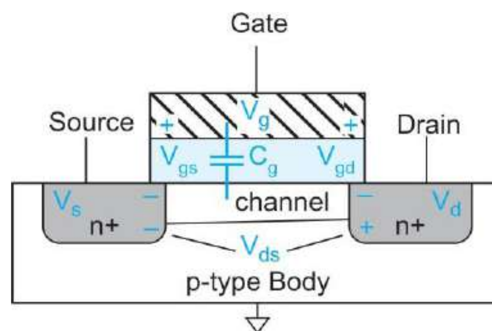


Figure 5: Average gate to channel voltage

✓ If the gate has length L and width W and the oxide thickness is t_{ox}, as shown in Figure 6, Then the capacitance C_g is

$$C_g = \kappa_{ox} \varepsilon_o \frac{WL}{t_{ox}} = \varepsilon_{ox} \frac{WL}{t_{ox}} = c_{ox} WL \quad \text{---------(1)}$$

Where, ε_o is the permittivity of free space, 8.85×10^{-14} F/cm,

Permittivity of SiO2 is $k_{ox} = 3.9$ times as great.

✓ The ε_{ox}/t_{ox} term is called as C_{ox}. Capacitance (C_{ox}) is a per unit area of the gate oxide.

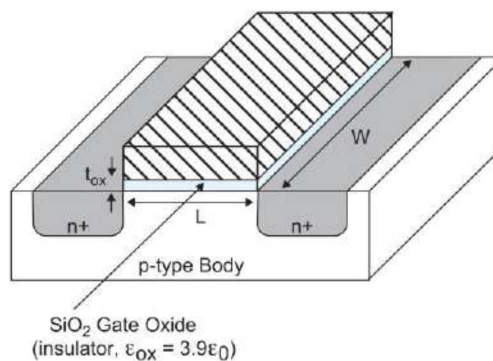


Figure 6: Transistor dimensions

✓ Average velocity (v) of carrier is proportional to the lateral electric field (field between source and drain). The constant of proportionality μ is called the mobility.

$$v = \mu E \text{ --------- (2)}$$

✓ The electric field E is the voltage difference between drain and source (V_{ds}) divided by the channel length (L).

$$E = \frac{V_{ds}}{L} \text{ ---------(3)}$$

✓ The time required for carriers to cross the channel is L divided by v.

- ✓ The current between source and drain is the total amount of charge in the channel divided by the time required to cross.

$$I_{ds} = \frac{Q_{channel}}{L/v}$$
$$= \mu C_{ox} \frac{W}{L}\left(V_{gs} - V_t - V_{ds}/2\right)V_{ds}$$
$$= \beta\left(V_{GT} - V_{ds}/2\right)V_{ds}$$

where

$$\beta = \mu C_{ox}\frac{W}{L};\ V_{GT} = V_{gs} - V_t \quad \text{-------- (4)}$$

- ✓ Equation (4) is called linear or resistive, because when $V_{ds} << V_{GT}$, I_{ds} increases linearly with V_{ds}, like an ideal resistor.
- ✓ k' is the k prime, k' = μC_{ox}.
- ✓ If $V_{ds} > V_{dsat} = V_{GT}$, the channel is no longer inverted in the drain region. Channel is pinched off.
- ✓ Beyond this point (called the drain saturation voltage), increasing the drain voltage has no further effect on current.
- ✓ Substituting $V_{ds} = V_{dsat}$ in Eq (4), we can find an expression for the saturation current (I_{ds}) that is independent of V_{ds}.

$$I_{ds} = \frac{\beta}{2}V_{GT}^2 \quad \text{- ------------------ (5)}$$

- ✓ This expression is valid for $V_{gs} > V_t$ and $V_{ds} > V_{dsat}$.

- ✓ Summarizes the current in the three regions:

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{Cutoff} \\ \beta\left(V_{GT} - V_{ds}/2\right)V_{ds} & V_{ds} < V_{dsat} \quad \text{Linear} \\ \frac{\beta}{2}V_{GT}^2 & V_{ds} > V_{dsat} \quad \text{Saturation} \end{cases}$$

****

1.6 : C – V CHARACTERISTICS OF MOS TRANSISTOR (AC characteristics)

- ❖ **Explain the dynamic behavior of MOSFET transistor with neat diagram. (April 2018)**
- ❖ **Discuss the CV characteristics of the CMOS. (Nov 2012, May 2014, Nov 2015, Nov 2016)**
- ❖ **Explain the electrical properties CMOS. (Nov 2017) [April / May 2023]**

- ✓ Each terminal of an MOS transistor has capacitance to the other terminals. Capacitances are nonlinear and voltage dependent (C-V).

SIMPLE MOS CAPACITANCES MODEL:

- ✓ The gate of an MOS transistor is a good capacitor. Its capacitance is necessary to attract charge to invert the channel, so high gate capacitance is required to obtain high I_{ds}.

- ✓ The gate capacitor can be viewed as a parallel plate capacitor with the gate on top, channel on bottom and the thin oxide dielectric between.
- ✓ The capacitance is $C_g = C_{ox} WL$-----------------(1)

$$C_g = C_{permicron} W \text{---------------(2)}$$

$$\text{Where } C_{permicron} = C_{ox} L = \frac{\varepsilon_{ox}}{t_{ox}} L$$

- ✓ In addition to the gate, the source and drain also have capacitances. These capacitances are called parasitic capacitors.
- ✓ The source and drain capacitances arise from the p–n junctions between the source or drain diffusion and the body. These capacitances are called diffusion capacitance C_{sb} and C_{db}.
- ✓ The depletion region acts as an insulator between the conducting p- and n-type regions, creating capacitance across the junction.
- ✓ The capacitance of junctions depends on the area and perimeter of the source and drain diffusion, the depth of the diffusion, the doping levels and the voltage.
- ✓ As diffusion has both high capacitance and high resistance, it is generally made as small as possible in the layout.

DETAILED MOS GATE CAPACITANCE MODEL:

- ✓ MOS gate places above the channel and may partially overlap the source and drain diffusion areas.
- ✓ The gate capacitance has two components, (i) the intrinsic capacitance C_{gc} (over the channel) and (ii) the overlap capacitances C_{gol} (to the source and drain).
- ✓ The intrinsic capacitance was approximated as a simple parallel plate with capacitance
- ✓ $C_0 = WLC_{ox}$.
- ✓ The intrinsic capacitance has three components representing the different terminals connected to the bottom plate are C_{gb} (gate-to-body), C_{gs} (gate-to-source), and C_{gd} (gate-to-drain).
- ✓ The behavior in three regions (Cutoff, Linear and Saturation) can be approximated as shown in Table 1.

Parameter	Cutoff	Linear	Saturation
C_{gb}	$\leq C_0$	0	0
C_{gs}	0	$C_0/2$	$2/3\ C_0$
C_{gd}	0	$C_0/2$	0
$C_g = C_{gs} + C_{gd} + C_{gb}$	C_0	C_0	$2/3\ C_0$

Table1: Approximation for intrinsic MOS gate capacitance

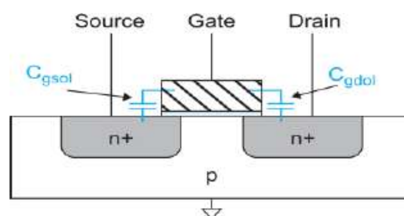


Figure 8: Overlap capacitances

DETAILED MOS DIFFUSION CAPACITANCE MODEL:

- ✓ The capacitance depends on both the area AS and sidewall perimeter PS of the source diffusion region. The area is AS = WD.
- ✓ The perimeter is PS = 2W +2D.

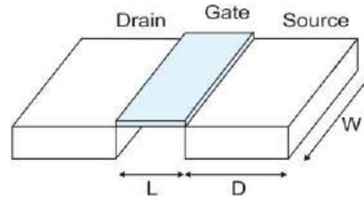


Figure 9: Diffusion region geometry

- ✓ The total source parasitic capacitance is $C_{sb} = AS * C_{jbs} + PS * C_{jbssw}$

  Where, C_{jbs} - Capacitance of the junction between the body and the bottom of the source
  C_{jbssw} - Capacitance of the junction between the body and the side walls of the source
- ✓ In summary, MOS transistor can be viewed as a four-terminal device with capacitances between each terminal pair, as shown in Figure 10.

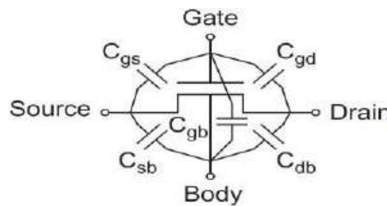


Figure 10: Capacitance of a MOS Transistor

- ✓ The gate capacitance includes an intrinsic component and overlap terms with the source and drain. The source and drain have parasitic diffusion capacitance to the body.

****

MOS(FET) Transistor Characteristic under Static and Dynamic Conditions

1.7 : DC TRANSFER CHARACTERISTICS

- ❖ **Explain the DC transfer characteristic of CMOS inverter.[APRIL-2015, Nov 2015]**
- ❖ **Draw and explain the DC and transfer characteristics of a CMOS inverter with necessary conditions for the different regions of operation. (Nov/Dec 2011) (Nov/Dec 2012) (May/June 2013) (April/May 2012) (May/June 2014) (Nov/Dec 2013) (May 2016, May 2017, Nov 2008) [April/May-2022.] [Nov/Dec-2020., April/May-2021] [April/May 2023]**
- ❖ **Explain the CMOS inverter DC characteristics. (Nov 2007, Nov 2009) [Nov/Dec 2022]**

- ✓ The DC transfer characteristics of a circuit relate the output voltage to the input voltage.

(i) Static CMOS inverter DC Characteristics:
The DC transfer function (V_{out} Vs. V_{in}) for the static CMOS inverter shown in Figure 11.

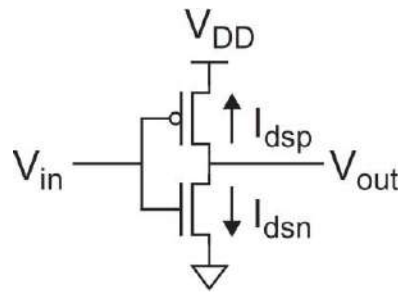


Figure 11: A Static CMOS inverter

- ✓ Table 2, shows various regions of operation for the n and p transistors.
- ✓ In this table, V_{tn} is the threshold voltage of the n-channel device, and V_{tp} is the threshold voltage of the p-channel device. V_{tp} is negative.
- ✓ The equations are given both in terms of V_{gs} /V_{ds} and V_{in} /V_{out}.
- ✓ As the source of the nMOS transistor is grounded, $V_{gsn} = V_{\text{in}}$ and $V_{dsn} = V_{\text{out}}$.
- ✓ As the source of the pMOS transistor is tied to V_{DD}, $V_{gsp} = V_{\text{in}} - V_{DD}$ and $V_{dsp} = V_{\text{out}} - V_{DD}$.

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

Table 2: Relationships between voltages for the three regions of operation of a CMOS inverter

- ✓ Figure 12(a), shows I_{dsn} and I_{dsp} in terms of V_{dsn} and V_{dsp} for various values of V_{gsn} and V_{gsp}.
- ✓ Figure 12(b), shows the same plot of I_{dsn} and $|I_{dsp}|$ in terms of V_{out} for various values of V_{in}.
- ✓ Operating points are plotted on V_{out} vs. V_{in} axes in Figure 12(c) to show the inverter DC transfer characteristics.
- ✓ The supply current $I_{DD} = I_{dsn} = |I_{dsp}|$ is plotted against V_{in} in Figure 13(d) showing that both transistors are momentarily ON as V_{in}.
- ✓ The operation of the CMOS inverter can be divided into five regions as indicated on figure 12(c).

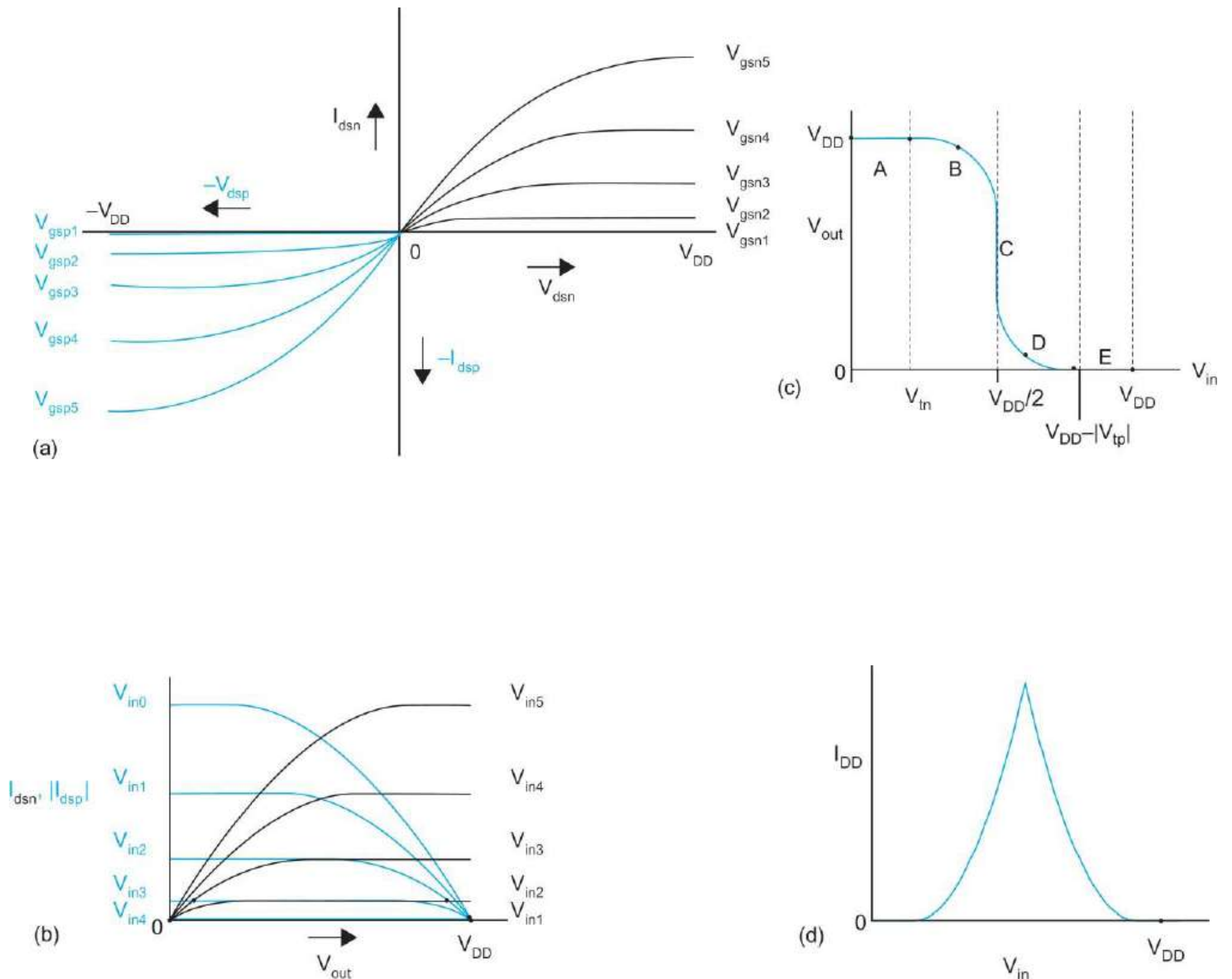


Figure 12: CMOS inverter DC characteristic

- ✓ The state of each transistor in each region is shown in Table 3.

Region	Condition	p-device	n-device	Output
A	$0 \le V_{in} < V_{tn}$	linear	cutoff	$V_{out} = V_{DD}$
B	$V_{tn} \le V_{in} < V_{DD}/2$	linear	saturated	$V_{out} > V_{DD}/2$
C	$V_{in} = V_{DD}/2$	saturated	saturated	V_{out} drops sharply
D	$V_{DD}/2 < V_{in} \le V_{DD} - \lvert V_{tp} \rvert$	saturated	linear	$V_{out} < V_{DD}/2$
E	$V_{in} > V_{DD} - \lvert V_{tp} \rvert$	cutoff	linear	$V_{out} = 0$

Table 3: Summary of CMOS inverter operation.

- ✓ In region *A*, the nMOS transistor is OFF and the pMOS transistor pulls the output to V_{DD}.
- ✓ In region *B*, the nMOS transistor starts to turn ON. It is pulling the output down.
- ✓ In region *C*, both transistors are in saturation.
- ✓ In region *D*, the pMOS transistor is partially ON.
- ✓ In region *E*, PMOS is completely OFF, making the nMOS transistor to pull the output down to GND.

(ii) Beta ratio Effects:

- ✓ For $\beta_p = \beta_n$, the inverter threshold voltage V_{inv} is $V_{DD}/2$.
- ✓ It allows a capacitive load to charge and discharge in equal times by providing equal current source and equal sink capabilities.

- ✓ Inverter with different beta ratios r = β_p / β_n is called skewed inverter.
- ✓ If r > 1, the inverter is HI-skewed. If r < 1, the inverter is LO-skewed. If r = 1, the inverter has normal skew or is unskewed.
- ✓ Figure 13, shows the impact of skewing the beta ratio on the DC transfer characteristics.
- ✓ As the beta ratio is changed, the switching threshold is varied.

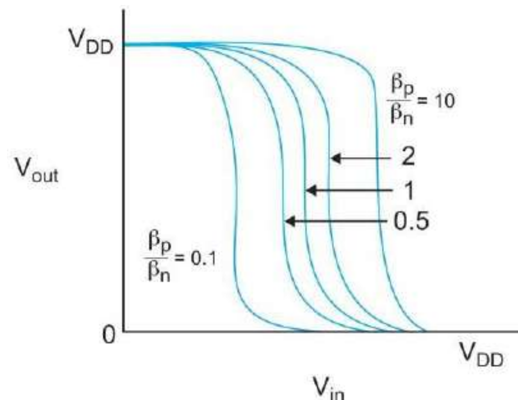


Figure 13: Transfer characteristics of skewed inverters

❖ **Derive the noise margins for a CMOS inverter. (May 2010, Nov 2016)**

(iii) **Noise Margins:**

- ✓ Noise margin (Noise immunity) is related to the DC voltage characteristics.
- ✓ Noise Margin allows determining the allowable noise voltage on the input of a gate, so that the output will not be corrupted.
- ✓ Two parameters of the noise margin are LOW noise margin (NM_L), and the HIGH noise margin (NM_H).

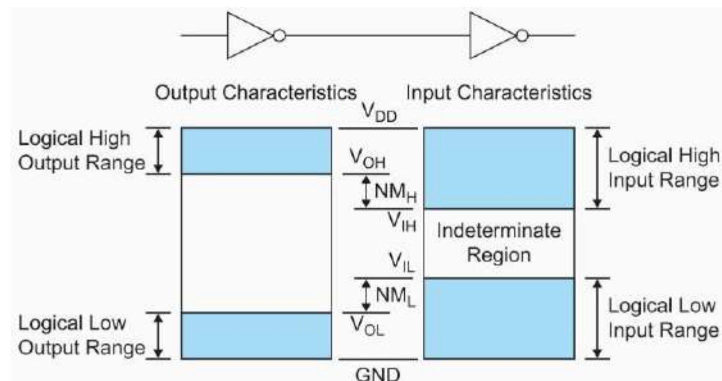


Figure 14: Noise Margin Definitions

- ✓ NM_L is defined as the difference in maximum LOW input voltage V_{IL} and the maximum LOW output voltage V_{OL}. $NM_L = V_{IL} - V_{OL}$
- ✓ The value of NM_H is the difference between the minimum HIGH output voltage V_{OH} and the minimum HIGH input voltage V_{IH}. i.e., $NM_H = V_{OH} - V_{IH}$
- ✓ Inputs between V_{IL} and V_{IH} are said to be in the indeterminate region or forbidden zone.

(iv) Pass Transistor DC Characteristics:

- ✓ The nMOS transistors pass 0's well but 1's poorly. Figure 15(a), shows an nMOS transistor with the gate and drain tied to V_{DD}.
- ✓ Initially at $V_s = 0$. $V_{gs} > V_{tn}$, so the transistor is ON and current flow.

- ✓ Therefore, nMOS transistors attempting to pass a 1 never pull the source above $V_{DD} - V_{tn}$. This loss is called a threshold drop.
- ✓ The pMOS transistors pass 1's well but 0's poorly.
- ✓ If the pMOS source drops below $|V_{tp}|$, the transistor cuts off.
- ✓ Hence, pMOS transistors only pull down to a threshold above GND, as shown in Figure 15(b).

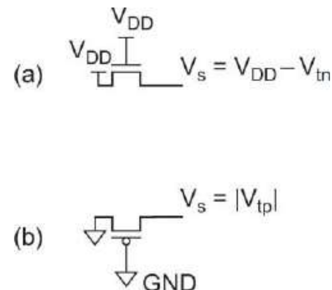


Figure 15: Pass Transistor threshold drops

****

1.8 : NON IDEAL I-V EFFECTS

- ❖ **Explain in detail about the non ideal I-V characteristics of a CMOS device. (MAY 2013)**
- ❖ **Explain channel length modulation and body effect. (Nov 2009, May 2013)**

- ✓ MOS characteristics degrade with temperature. It is useful to have a qualitative understanding of non ideal effects to predict their impact on circuit behavior.

(i) Mobility Degradation and Velocity Saturation:

- ✓ Current is proportional to the lateral electric field $E_{lat} = V_{ds} /L$ between source and drain.
- ✓ A high voltage at the gate of the transistor attracts the carriers to the edge of the channel, causing carriers collision with the oxide interface that slows the carriers. This is called mobility degradation.
- ✓ Carriers approach a maximum velocity (v_{sat}) when high fields are applied. This phenomenon is called velocity saturation.

(ii) Channel Length Modulation:

- ✓ Current I_{ds} is an independent of V_{ds} for a transistor in saturation.
- ✓ The p–n junction between the drain and body forms a depletion region with a width L_d that increases with V_{db} , as shown in Figure 16.
- ✓ The depletion region effectively shortens the channel length to $L_{eff} = L - L_d$
- ✓ To avoid the body voltage into calculations, assume the source voltage is close to the body voltage i.e $V_{db} = V_{ds}$.
- ✓ Hence, increasing V_{ds} decreases the effective channel length.
- ✓ Shorter channel length results in higher current. Thus, I_{ds} increases with V_{ds} in saturation, as shown in Figure 16.

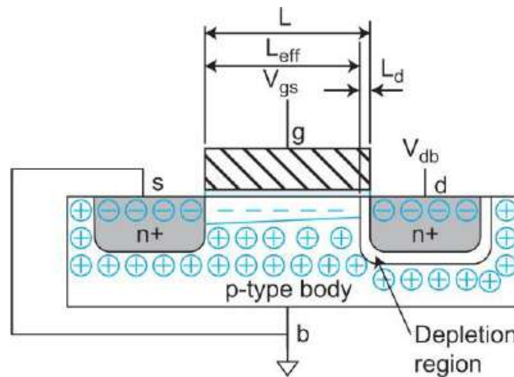


Figure 16: Depletion region shortens effective channel length

- ✓ In Saturation region, I_{ds} is

$$I_{ds} = \frac{\beta}{2} V_{GT}^2 \left(1 + \frac{V_{ds}}{V_A}\right)$$

- ✓ Hence, V_A is proportional to channel length. This channel length modulation model is a gross oversimplification of nonlinear behavior.

(iii) Threshold Voltage (V_t) Effects

❖ **Explain in detail about effect and its effect in MOS device. (May 2016)**

- ✓ Threshold voltage V_t increases with the source voltage, decreases with the body voltage, decreases with the drain voltage and increases with channel length.

Body Effect:

- ✓ 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

$$V_t = V_{t0} + \gamma\left(\sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s}\right)$$

where V_{t0} is the threshold voltage when the source is at the body potential, ϕ_s is the surface potential at threshold and γ is the body effect coefficient.

(iv) Leakage:

- ✓ Even when transistors are OFF, transistors leak small amounts of current.
- ✓ Leakage mechanisms include subthreshold conduction between source and drain, gate leakage from the gate to body and junction leakage from source to body and drain to body.
- ✓ Subthreshold conduction is caused by thermal emission of carriers over the potential barrier set by the threshold.
- ✓ Gate leakage is a quantum-mechanical effect caused by tunneling through the extremely thin gate dielectric.
- ✓ Junction leakage is caused by current through the p-n junction between the source/drain diffusions and the body.

****

1.9 : Device models:

Explain the following: Device models and device characteristics. (MAY 2014)

- ✓ SPICE (Simulation Program with Integrated Circuit Emphasis) provides a wide variety of MOS transistor models with various trade-offs between complexity and accuracy.
- ✓ Level 1 and Level 3 models were 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.

i. Level 1 model:

- ✓ The SPICE Level 1, or Shichman-Hodges Model is closely related to the Shockley model, enhanced with channel length modulation and the body effect.
- ✓ The basic current model is:

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t & \text{cutoff} \\ \mathrm{KP}\dfrac{W_{\text{eff}}}{L_{\text{eff}}}\left(1+\mathrm{LAMBDA}\times V_{ds}\right)\left(V_{gs}-V_t-\dfrac{V_{ds}}{2}\right)V_{ds} & V_{ds} < V_{gs}-V_t & \text{linear} \\ \dfrac{\mathrm{KP}}{2}\dfrac{W_{\text{eff}}}{L_{\text{eff}}}\left(1+\mathrm{LAMBDA}\times V_{ds}\right)\left(V_{gs}-V_t\right)^2 & V_{ds} > V_{gs}-V_t & \text{saturation} \end{cases}$$

- ✓ The parameters from the SPICE model are given in ALL CAPS.
- ✓ β is written instead as KP (W_{eff} /L_{eff}), where KP is a model parameter. W_{eff} and L_{eff} are the effective width and length.
- ✓ The LAMBDA term (LAMBDA =$1/V_A$) models channel length modulation.
- ✓ The threshold voltage is modulated by the source-to-body voltage V_{sb} through the body effect.
- ✓ For non negative V_{sb} , the threshold voltage, V_t is

$$V_t = \mathrm{VTO} + \mathrm{GAMMA}\left(\sqrt{\mathrm{PHI}+V_{sb}} - \sqrt{\mathrm{PHI}}\right)$$

Where, VTO is the "zero-bias" threshold voltage V_{t0},

GAMMA is the body effect coefficient , and PHI is the surface potential.

- ✓ Level 1 model is easy to correlate with hand analysis, but it is too simplistic for modern design.

ii. Level 2 and 3 models

- ✓ The SPICE Level 2 and 3 models add effects of velocity saturation, mobility degradation, subthreshold conduction and drain-induced barrier lowering.
- ✓ The Level 2 model is based on the Grove-Frohman equations.
- ✓ Level 3 model is based on empirical equations that provide similar accuracy, faster simulation times and better convergence.
- ✓ These models are not efficient models for measuring I-V characteristics of modern transistors.

iii. BSIM models

- ✓ The Berkeley Short-Channel IGFET Model (BSIM) is a very elaborate model that is now widely used in circuit simulation.
- ✓ The models are derived from the fundamental device physics but uses many number of parameters to fit the behavior of modern transistors.

- ✓ BSIM versions 1, 2, 3v3, and 4 are implemented as SPICE levels 13, 39, 49, and 54, respectively.
- ✓ BSIM is quite good for digital circuit simulation.
- ✓ Features of the model are
  - Continuous and differentiable I-V characteristics across subthreshold, linear and saturation regions for good convergence.
  - Sensitivity of parameters such as V_t to transistor length and width.
  - Detailed threshold voltage model including body effect and drain-induced barrier Lowering (DIBL).
  - Velocity saturation, mobility degradation and other short-channel effects.
  - Multiple gate capacitance models.
  - Diffusion capacitance and resistance models.
  - Gate leakage models.

****

1.10 : SCALING

- ❖ **Discuss the scaling principles and its limits. (MAY 2013, Nov 2017, Nov 2018)**
- ❖ **Discuss the principle of constant field and lateral scaling. Write the effects of the above scaling methods on the device characteristics. (Nov 2012, Dec 2011, Nov 2015, May 2016)**
- ❖ **Explain need of scaling, scaling principles and fundamental units of CMOS inverter. (May 2017)**
- ❖ **Highlight the need for scaling. Enumerate in detail constant electric field, constant voltage and combined electric field and voltage scaling for different parameters of MOSFET. [Nov 2019]**

- ✓ In VLSI design, the transistor size has reduced by 30% every two to three years. Scaling is reducing feature size of transistor.
- ✓ Nowadays, transistors become smaller, switch faster, dissipate less power and cheaper.
- ✓ Designers need to predict the effect of feature size scaling on chip performance to plan future products and ensure existing products for cost reduction.

Transistor scaling:

- ✓ Dennard's Scaling Law predicts that the basic operational characteristics of a MOS transistor can be preserved and the performance can be improved.
- ✓ Parameters of a device are scaled by a dimensionless factor S.
- ✓ These parameters include the following:
  - All dimensions (in the x, y, and z directions)
  - Device voltages
  - Doping concentration densities

Constant field scaling (Full Scaling):

- ✓ In **constant field scaling,** electric fields remain the same as both voltage and distance shrink.
- ✓ 1/S scaling is applied to all dimensions, device voltages and concentration densities.
  - I_{ds} per transistor are scaled by 1/S.
  - No. of transistors per unit area is scaled by S^2.
  - Current density is scaled by S and power density remains constant.
    - e.g., $(\frac{1}{S} * \frac{1}{S}) * S^2$

Lateral scaling (gate-shrink):

- ✓ Another approach is **lateral scaling**, in which only the gate length is scaled.

- ✓ This is commonly called as gate shrink, because it can be done easily to an existing mask database for a design.
  - I_{ds} per transistor are scaled by S.
  - No. of transistors per unit area is scaled by S.
  - Current density is scaled by S^2 and power density is scaled by S^2.
- ✓ The industry generally scales process generations with 30% shrink.
- ✓ It reduces the cost (area) of a transistor by a factor of two.
- ✓ A 5% gate shrink (S = 1.05) is commonly applied as a process, becomes mature to boost the speed of components in that process.

Constant voltage scaling: V_{DD} is held constant, while process is scaled.

- ✓ **Constant voltage scaling (Fixed scaling)** offers quadratic delay improvement as well as cost reduction.
- ✓ It is also maintaining continuity in I/O voltage standards. Constant voltage scaling increases the electric fields in devices.
  - I_{ds} per transistor are scaled by S.
  - No. of transistors per unit area is scaled by S^2.
  - Current density is scaled by S^3 and power density is scaled by S^3.
- ✓ A 30% shrink with Dennard scaling improves clock frequency by 40% and cuts power consumption per gate by a factor of 2.
- ✓ Maintaining a constant field has the further benefit, that many nonlinear factors and wear out mechanisms are unaffected.
- ✓ From 90nm generation technology, voltage scaling is dramatically slowed down due to leakage. This may ultimately limit CMOS scaling.

Parameter	Sensitivity	Dennard Scaling	Constant Voltage	Lateral Scaling
Scaling Parameters				
Length: L		$1/S$	$1/S$	$1/S$
Width: W		$1/S$	$1/S$	1
Gate oxide thickness: t_{ox}		$1/S$	$1/S$	1
Supply voltage: V_{DD}		$1/S$	1	1
Threshold voltage: V_{tn}, V_{tp}		$1/S$	1	1
Substrate doping: N_A		S	S	1

Table: Influence of scaling on MOS device characteristics

Interconnecting Scaling:

- ✓ Wires to be scaled equally in width and thickness to maintain an aspect ratio close to 2.
- ✓ Wires can be classified as local, semiglobal and global.
- ✓ Local wires run within functional units and use the bottom layers of metal.
- ✓ Semiglobal wires run across larger blocks or cores, typically using middle layers of metal.
- ✓ Both local and semiglobal wires are scaling with feature size.
- ✓ Global wires run across the entire chip using upper levels of metal.
- ✓ Global wires do not scale with feature size. Indeed, they may get longer (by a factor of DC, on the order of 1.1) because, die size has been gradually increasing.
- ✓ When wire thickness is scaled, the capacitance per unit length remains constant.

****

1.11 : CMOS DEVICES & TECHNOLOGIES:

The four main CMOS technologies are

- n-Well process
- p-Well process
- Twin-tub Process
- Silicon on Insulator

- **Explain the different steps involved in CMOS fabrication / manufacturing process with neat diagrams. (Nov 2007, Nov 2009, Nov 2016, NOV 2018)**
- **Describe with neat diagram the n-well and channel formation in CMOS process. (Nov/Dec 2014)(Nov/Dec 2011) (April/May 2011) (Nov/Dec 2012)**

n-WELL PROCESS:

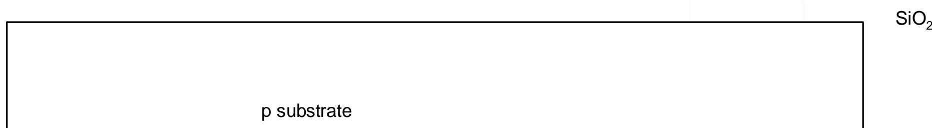
Step 1: Start with blank wafer

First step will be to form the n-well

- Cover wafer with protective layer of SiO_2 (oxide)
- Remove layer where n-well should be built.

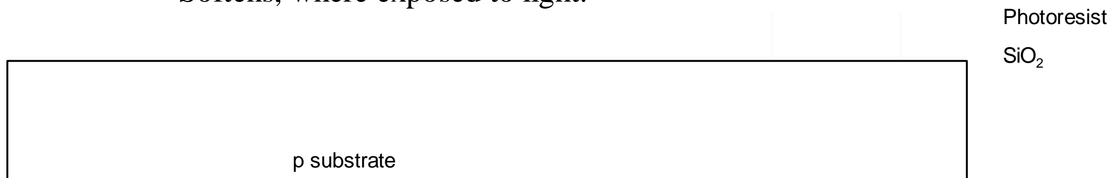
Step 2: Oxidation

Grow SiO_2 on top of Si wafer, at 900 – 1200^0 C with H_2O or O_2 in oxidation furnace.



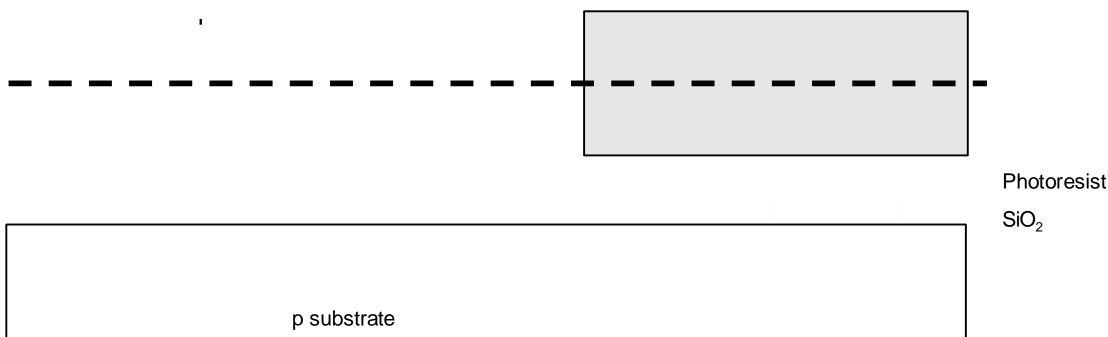
Step 3: Photoresist

- Spin on photoresist
  - Photoresist is a light-sensitive organic polymer.
  - Softens, where exposed to light.



Step 4: Lithography

- Expose photoresist through n-well mask.
- Strip off exposed photoresist.

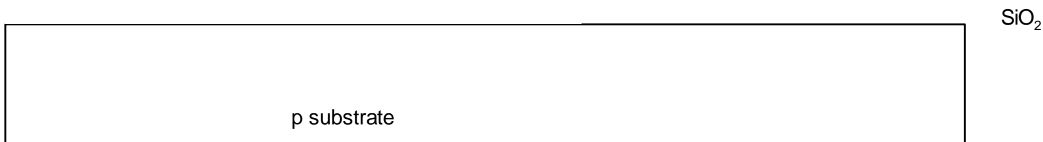


Step 5: Etch

- Etch oxide with hydrofluoric acid (HF).
- Only attracts oxide, where resist has been exposed.

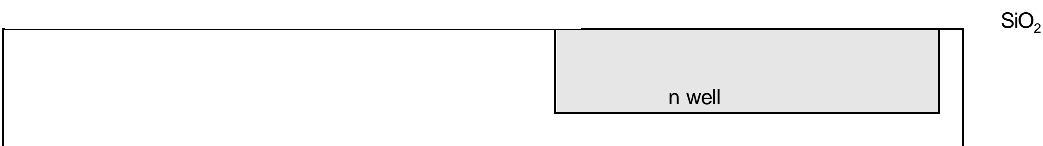
Step 6: Strip Photoresist

- Etch the remaining photoresist using a mixture of acids.



Step 7: n-well

n-well is formed with diffusion or ion implantation.

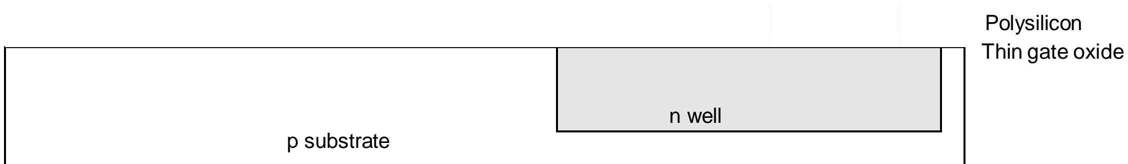


Step 8: Strip Oxide

- Strip off the remaining oxide using HF.
- Back to bare wafer with n-well.
- Subsequent steps involve similar series of steps.

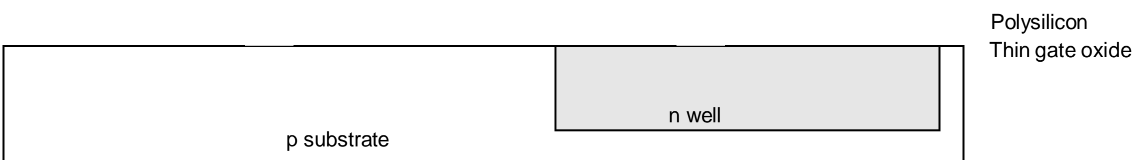
Step 9:Polysilicon

- Deposit thin layer of oxide. Use CVD to form poly and dope heavily to increase conductivity.



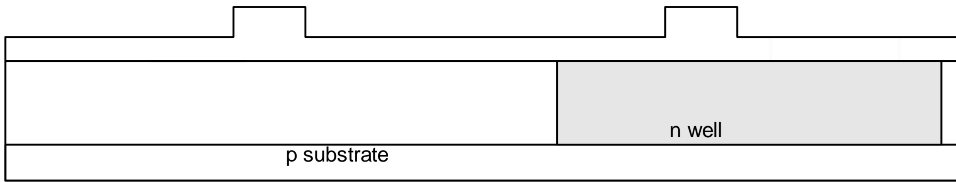
Step 10: Polysilicon Patterning

- Use same lithography process to pattern polysilicon.



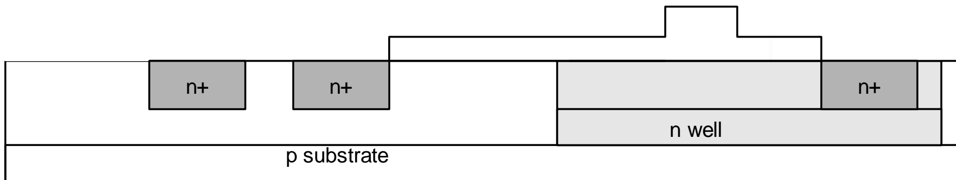
Step 11: Self-Aligned Process

Cover with oxide to define n diffusion regions.



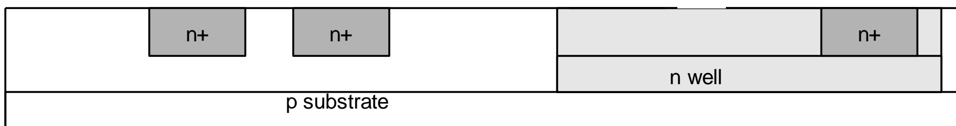
Step 12: N-diffusion

- Pattern oxide, using n+ active mask to define n diffusion regions.
- Diffusion or ion implantation is used to create n diffusion regions.



Step 13:

- Strip off oxide to complete patterning step.

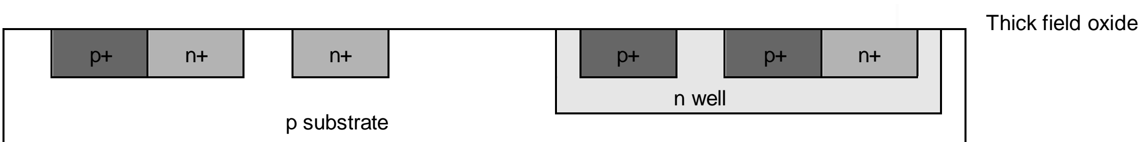


Step 14:P-Diffusion

- Similar set of steps are followed to form p+ diffusion regions for pMOS source and drain and substrate contact.

Step 15: Contacts

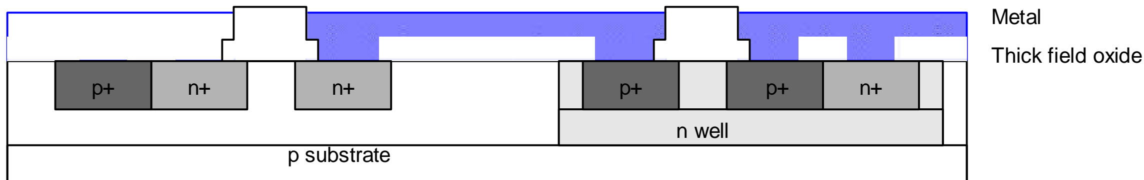
- Now, we need to wire together the devices.
- Cover chip with thick field oxide.
- Etch oxide, where contact cuts are needed.



Step 16: Metallization

- Sputter on aluminum over whole wafer.

- Pattern to remove excess metal, leaving wires.



****

P-WELL PROCESS:

- A common approach to p-well CMOS fabrication is to start with moderately doped n-type substrate (wafer), create the p-type well for the n-channel devices and build the p-channel transistor in the native n-substrate.

❖ **Explain the twin tub process with a neat diagram. (Nov 2007, April 2008)**

Twin-tub process:

Step 1:

n- Substrate is taken initially, which is shown in figure.

Step 2:

Next step is epitaxial layer deposition. Lightly doped epitaxial layer is deposited above n-substrate.

Step 3:

The next step is tub formation. Two wells are formed namely n-well and p-well.

Polysilicon layer is formed above overall substrate.

Step 4:

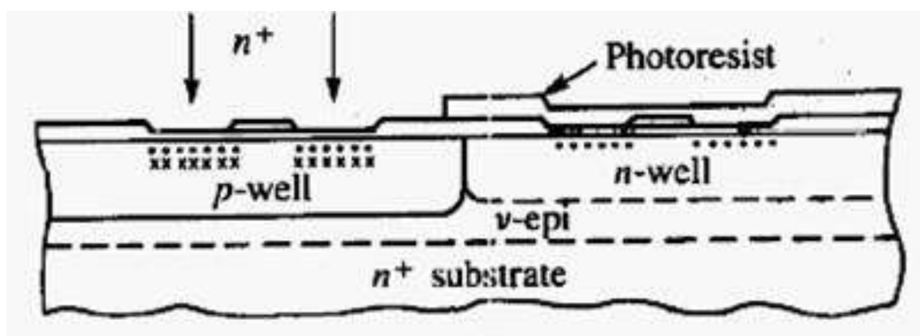
Polysilicon gates are formed for n-well and p-well by using photo-etching process.

Step 5:

n^+diffusion is formed in n-well, P^+ diffusion is formed in p-well. These are used for V_{DD} contact and V_{SS} contact. These are known as substrate formation.

Step 6:

Then, contact cuts are defined as in n-well process. Then metallization is processed.



****

Power consumption

Reduction of power consumption makes a device more reliable. The need for devices that consume a minimum amount of power was a major driving force behind the development of CMOS technologies. As a result, CMOS devices are best known for low power consumption. However, for minimizing the power requirements of a board or a system, simply knowing that CMOS devices may

use less power than equivalent devices from other technologies does not help much. It is important to know not only how to calculate power consumption, but also to understand how factors such as input voltage level, input rise time, power-dissipation capacitance, and output loading affect the power consumption of a device.

The main topics discussed are:
• Power-consumption components
• Static power consumption
• Dynamic power consumption

Power-Consumption Components

High frequencies impose a strict limit on power consumption in computer systems as a whole. Therefore, power consumption of each device on the board should be minimized. Power calculations determine power-supply sizing, current requirements, cooling/heatsink requirements, and criteria for device selection. Power calculations also can determine the maximum reliable operating frequency.

Two components determine the power consumption in a CMOS circuit:
• Static power consumption
• Dynamic power consumption

- ✓ CMOS devices have very low static power consumption, which is the result of leakage current.
- ✓ This power consumption occurs when all inputs are held at some valid logic level and the circuit is not in charging states.
- ✓ But, when switching at a high frequency, dynamic power consumption can contribute significantly to overall power consumption.
- ✓ Charging and discharging a capacitive output load further increases this dynamic power consumption. This application report addresses power consumption in CMOS logic families (5 V and 3.3 V) and describes the methods for evaluating both static and dynamic power consumption.
- ✓ Additional information is also presented to help explain the causes of power consumption, and present possible solutions to minimize power consumption in a CMOS system.
- ✓ Static Power Consumption Typically, all low-voltage devices have a CMOS inverter in the input and output stage. Therefore, for a clear understanding of static power consumption, refer to the CMOS inverter modes shown in Figure

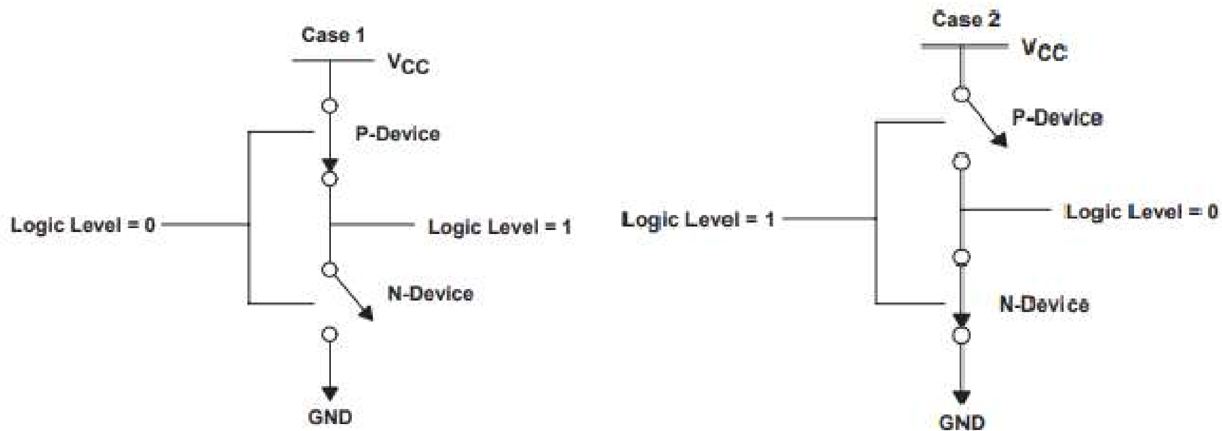


Figure: CMOS Inverter modes

- As shown in above Figure, if the input is at logic 0, the n-MOS device is OFF, and the p-MOS device is ON (Case 1).
- The output voltage is V_{CC}, or logic 1. Similarly, when the input is at logic 1, the associated n-MOS device is biased ON and the p-MOS device is OFF.
- The output voltage is GND, or logic 0. Note that one of the transistors is always OFF when the gate is in either of these logic states.
- Since no current flows into the gate terminal, and there is no dc current path from V_{CC} to GND, the resultant quiescent (steady-state) current is zero, hence, static power consumption (Pq) is zero.
- However, there is a small amount of static power consumption due to reverse-bias leakage between diffused regions and the substrate.
- This leakage inside a device can be explained with a simple model that describes the parasitic diodes of a CMOS inverter, as shown in Figure below.

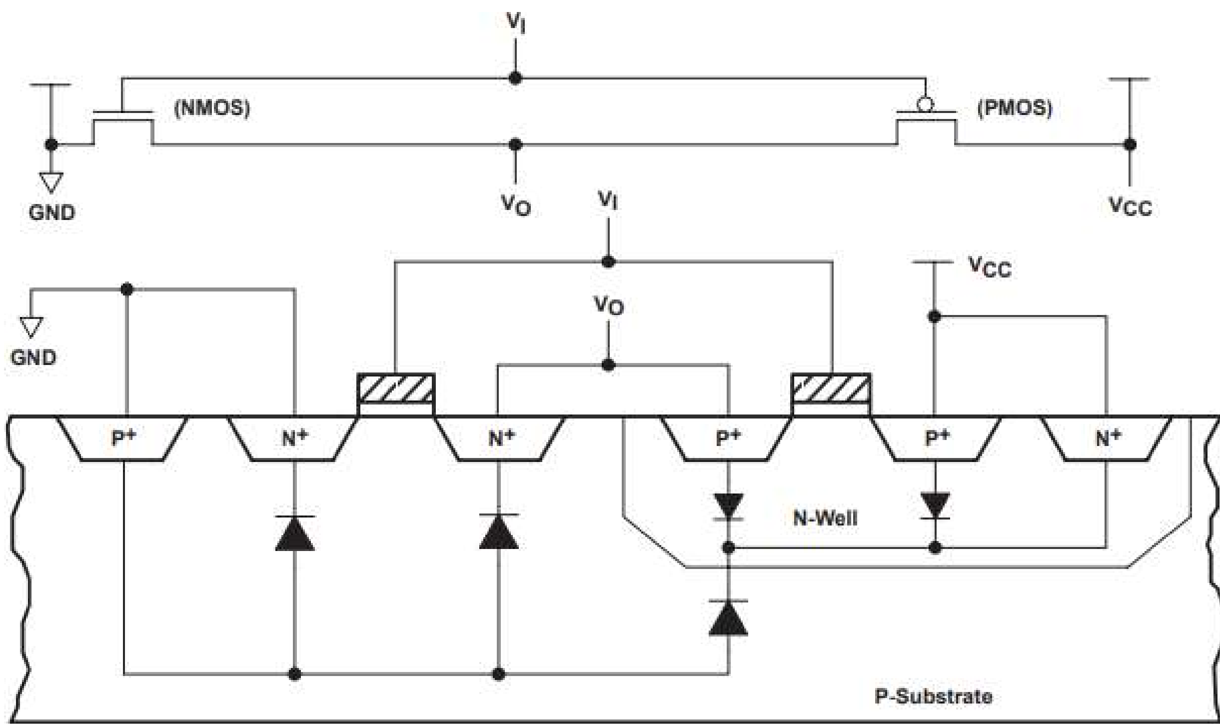


Figure: Model Describing Parasitic Diodes Present in CMOS Inverter

The source drain diffusion and N-well diffusion form parasitic diodes. In above Figure, the parasitic diodes are shown between the N-well and substrate. Because parasitic diodes are reverse biased, only their leakage currents contribute to static power consumption.

The leakage current (I_{lkg}) of the diode is described by the following equation:

$I_{lkg} = i_S\,(e^{qV/kT} - 1)$ .... (1)

Where:

i_S = reverse saturation current

V = diode voltage

k = Boltzmann's constant (1.38×10^{-23} J/K)

q = electronic charge (1.602×10^{-19} C)

T = temperature

Static power consumption is the product of the device leakage current and the supply voltage.

Total static power consumption, P_S, can be obtained as shown in equation 2.

$P_S = \Sigma$ (leakage current) x (supply voltage)...............................(2)

Most CMOS data sheets specify an ICC maximum in the 10-µA to 40-µA range, encompassing total

leakage current and other circuit features that may require some static current not considered in the simple inverter model.

The leakage current ICC (current into a device), along with the supply voltage, causes static power consumption in the CMOS devices.

This static power consumption is defined as quiescent, or P_S, and can be calculated by equation 3.

$P_S = V_{CC} \times I_{CC}$......................................(3)

Where:

V_{CC} = supply voltage

I_{CC} = current into a device (sum of leakage currents as in equation 2)

Another source of static current is ΔI_{CC}. This results when the input levels are not driven all the way to the rail, causing the input transistors to not switch off completely.

Dynamic Power Consumption

The dynamic power consumption of a CMOS IC is calculated by adding the transient power consumption (P_T), and capacitive-load power consumption (P_L).

- ✓ Transient Power Consumption Transient power consumption is due to the current that flows only when the transistors of the devices are switching from one logic state to another.
- ✓ This is a result of the current required to charge the internal nodes (switching current) plus the through current (current that flows from V_{CC} to GND when the p-channel transistor and n-channel transistor turn on briefly at the same time during the logic transition).
- ✓ The frequency at which the device is switching, plus the rise and fall times of the input signal, as well as the internal nodes of the device, have a direct effect on the duration of the current spike.
- ✓ For fast input transition rates, the through current of the gate is negligible compared to the switching current. For this reason, the dynamic supply current is governed by the internal capacitance of the IC and the charge and discharge current of the load capacitance.

Transient power consumption can be calculated using equation 4.

$P_T = C_{pd} \times V_{CC}^2 \times f_I \times NSW$..(4)

Where:

P_T = transient power consumption

V_{CC} = supply voltage

f_I = input signal frequency

NSW = number of bits switching

C_{pd} = dynamic power-dissipation capacitance

In the case of single-bit switching, NSW in equation 4 is 1.

Dynamic supply current is dominant in CMOS circuits because most of the power is consumed in moving charges in the parasitic capacitor in the CMOS gates.

As a result, the simplified model of a CMOS circuit consisting of several gates can be viewed as one large capacitor that is charged and discharged between the power-supply rails.

Therefore, the power–dissipation capacitance (C_{pd}) is often specified as a measure of this equivalent capacitance and is used to approximate the dynamic power consumption.

C_{pd} is defined as the internal equivalent capacitance of a device calculated by measuring operating current without load capacitance.

Depending on the output switching capability, C_{pd} can be measured with no output switching (output disabled) or with any of the outputs switching (output enabled).

Capacitive-Load Power Consumption

Additional power is consumed in charging external load capacitance and is dependent on switching frequency.

The following equation can be used to calculate this power if all outputs have the same load and are switching at the same output frequency.

$P_L = C_L \times V_{CC}^2 \times f_O \times NSW$ (C_L is the load per output) .... (5)

Where:

P_L = capacitive-load power consumption

V_{CC} = supply voltage

f_O = output signal frequency

C_L = external (load) capacitance

NSW = total number of outputs switching

In the case of different loads and different output frequencies at all outputs, equation 6 is used to calculate capacitive-load power consumption.

$P_L = \Sigma (C_{Ln} \times f_{On}) \times V_{CC}^2$ ....(6)

Where:

Σ = sum of n different frequencies and loads at n different outputs

f_{On} = all different output frequencies at each output, numbered 1 through n (Hz)

V_{CC} = supply voltage (V)

C_{Ln} = all different load capacitances at each output, numbered 1 through n.

Therefore, dynamic power consumption (P_D) is the sum of these two power consumptions and can be

expressed as shown in equation 7, equation 8 (single-bit switching), and equation 9 (multiple-bit switching with variable load and variable output frequencies).

$P_D = P_T + P_L$ ....(7)

$P_D = (C_{pd} \times f_I \times V_{CC}^2) + (C_L \times f_O \times V_{CC}^2)$ ....(8)

$P_D = [(C_{pd} \times f_I \times NSW) + \Sigma (C_{Ln} \times f_{On})] \times V_{CC}^2$ ....(9)

Where:

C_{pd} = power-consumption capacitance (F)

f_I = input frequency (Hz)

f_{On} = all different output frequencies at each output, numbered 1 through n (Hz)

NSW = total number of outputs switching

V_{CC} = supply voltage (V)

C_{Ln} = all different load capacitances at each output, numbered 1 through n.

Total power consumption is the sum of static and dynamic power consumption.

$P_{tot} = P_{(static)} + P_{(dynamic)}$ ....(10)

****

TWO MARKS QUESTIONS & ANSWERS

UNIT I - MOS TRANSISTOR PRINCIPLES

1. **Give the advantages of Integrated Circuit.**
   Advantages of Integrated Circuit:
   - Size is less
   - High Speed
   - Less Power Dissipation

2. **What is meant by CMOS technology?**
   Complementary metal-oxide-semiconductor (CMOS) is a technology for constructing integrated circuits. CMOS technology is used in microprocessors, microcontrollers, static RAM and other digital logic circuits.

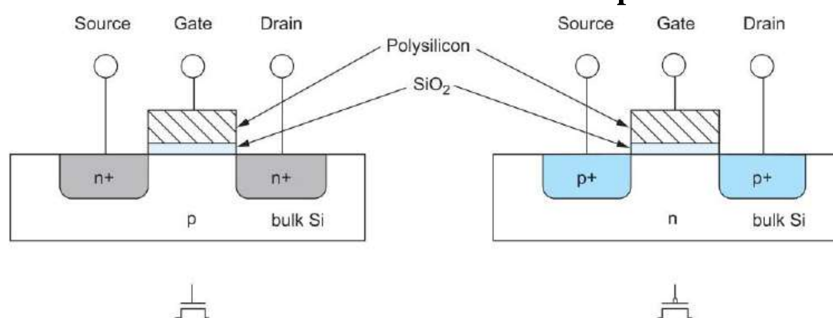
3. **How do you construct MOS transistor?**
   A Metal-Oxide-Semiconductor (MOS) structure is created by superimposing layers of conducting and insulating materials to form a structure.

4. **What is meant by MOS transistor?**
   The transistor consists of a stack of the conducting gate, an insulating layer of silicon dioxide (SiO2) and the silicon wafer (also called the substrate, body, or bulk). Gate of transistor is built from metal, so the stack is called metal oxide- semiconductor (MOS).

   Transistor operation is controlled by electric field so the device is also called Metal Oxide Semiconductor Field Effect Transistor (MOSFET).

5. **What is pull down device? (Nov 2009)**
   A device is connected to pull the output voltage to the lower supply voltage (0V) is called pull down device.

6. **Draw the schematic structure of n-MOS and p-MOS transistor with symbol.**



(a) n-MOS transistor (b) p-MOS transistor

7. **What is pull up device?**
   A device is connected to pull the output voltage to the upper supply voltage usually V_{DD} is called pull up device.

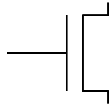
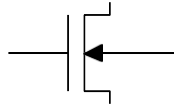
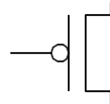
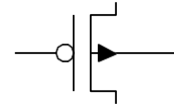
8. **Compare nMOS and pMOS devices. (or) How CMOS act as a switch? (Nov 2007) [April/May – 2023]**

✓ In nMOS, electrons are the majority carriers.

When the gate of an nMOS transistor is high, the transistor is ON. When the gate is low, the nMOS transistor is OFF.

- ✓ In pMOS, holes are the majority carriers.
When the gate of a pMOS transistor is low, the transistor is ON. When the gate is high, the pMOS transistor is OFF.

nMOS Symbol pMOS Symbol

9. Compare depletion and enhancement mode devices. [Nov/Dec-2007]

Depletion mode: The mode at which devices conduct with zero gate bias are called depletion mode.

Enhancement mode: The mode at which devices that are normally cut-off (i.e., non-conducting) with zero gate bias are called as enhancement mode.

10. Why nMOS technology is preferred more than pMOS technology?

The nMOS technology is preferred more than pMOS technology, because n- channel transistor has greater switching speed when compared to pMOS transistor.

11. Define threshold voltage of MOSFET. (April 2019) [Nov 2019]

The threshold voltage is the minimum gate to source voltage that is needed to create a conducting path between the source and drain terminals.

12. What is Moore's Law?

The Moore's law states that number of transistor on an integrated circuit will double every 18 months.

13. What are the three types of modes of MOS transistor? (or) Give the different modes of operation of MOS transistor. (or) What are the different MOS layers? (Nov 2009)

Three types of modes of MOS transistor are accumulation mode, depletion mode and inversion mode.

14. What is meant by accumulation mode in MOS transistor?

If a negative voltage is applied to the gate, so there is negative charge on the gate. The positively charged holes are attracted to the region beneath (below) the gate. This is called the accumulation mode.

15. What is meant by depletion mode in MOS transistor?

If a small positive voltage is applied to the gate, there are positive charges on the gate. The holes in the body are repelled from the region directly beneath (below) the gate, resulting in a depletion region forming below the gate.

16. What is meant by inversion mode (or) inversion layer in MOS transistor?

If applying higher positive voltage exceeding a threshold voltage (Vt), attracting more positive charges to the gate.

The holes are repelled and some free electrons in the body are attracted to the region beneath (below) the gate. This conductive layer of electrons in the p-type body is called the inversion layer.

17. List the different operating regions of MOS system. (May2012)

Three operating regions of MOS transistor (or) system are

(i) Cut off region or subthreshold region
(ii) Linear region
(iii) Saturation region

18. When the channel is said to be pinched –off? (May 2010)

If voltage between drain and source V_{ds} becomes sufficiently large that $V_{gd} < V_t$, the channel is no longer inverted near the drain and channel becomes pinched off.

Where V_g = gate-to-drain voltage, V_t= Threshold voltage.

19. When will nMOS transistor operates in cutoff region?

If $V_{gs} < V_t$, the transistor is cutoff (OFF).

Where V_{gs} = gate-to-source voltage, V_t= Threshold voltage.

20. When will nMOS transistor operates in linear region?

If $V_{gs} > V_t$, the transistor turns ON. If V_{ds} is small, the transistor acts as a linear resistor in which the current flow is proportional to V_{ds}.

Where V_{gs} = gate-to-source voltage, V_t= Threshold voltage & V_{ds}= drain-to-source voltage

21. When will nMOS transistor operates in saturation region?

If $V_{gs} > V_t$ and V_{ds} is large, the transistor acts as a current source in which the current flow becomes independent of Vds .

Where V_{gs} = gate-to-source voltage, V_t= Threshold voltage & V_{ds}= drain-to-source voltage.

22. Determine whether an nMOS transistor with a threshold voltage of 0.7V is operating in the saturation region if V_{gs}= 2V and V_{ds}=3V.(Nov 2011)

Condition for saturation region is $(V_{gs} - V_t) < V_{ds}$, So This nMOS transistor operated in the saturation region.

23. Give the expression for drain current (I_{ds}) for different modes of operation of MOS transistor.

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ \beta\left(V_{gs} - V_t - \frac{V_{ds}}{2}\right)V_{ds} & V_{ds} < V_{dsat} \quad \text{linear} \\ \frac{\beta}{2}\left(V_{gs} - V_t\right)^2 & V_{ds} > V_{dsat} \quad \text{saturation} \end{cases}$$

24. List the different capacitances of a MOS transistor.

Capacitances of a MOS transistor are

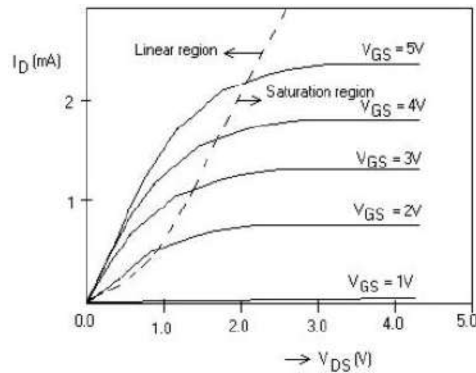
(i) Parasitic Capacitances or diffusion capacitances: C_{sb} (Source to body capacitance) & C_{db} (Drain to body capacitance)
(ii) Intrinsic capacitance: The intrinsic capacitance has three components representing the different terminals connected to the bottom plate are C_{gb} (gate-to-body), C_{gs} (gate-to-source), and C_{gd} (gate-to-drain).

(iii) Overlap capacitance: C_{gsol} (overlap capacitance to the source) & C_{gdol} (overlap capacitance to the drain)

25. Define body effect (or) substrate bias effect. [May/June-2009] [Apr/may-2010]
What is meant by body effect? (NOV. 2014)
Define body bias effect. (Nov 2016)

Threshold voltage (V_t) is not constant with respect to voltage difference between substrate and source of MOS transistor. This is known as body effect. It is otherwise known as substrate bias effect.

26. Draw the I-V characteristics of MOS transistor. (May 2012)



27. What are the secondary effects (or) Non ideal effects of MOS transistor?[May 2014]

Secondary effects are

- Mobility degradation
- Velocity saturation
- Channel length modulation
- Body effect
- Subthreshold conduction
- Junction leakage current
- Tunneling
- Short channel effect

28. What is velocity saturation effect? (April 2018)

Carriers approach a maximum velocity (v_{sat}) when high fields are applied. This phenomenon is called velocity saturation.

29. Define channel length modulation.(Nov 2011, April 2016, May 2017)

Channel length modulation defines effective length of the conductive channel. It is modulated by the external applied V_{ds}. Increasing V_{ds}, causes the depletion region at the drain junction to grow and thus reduces the length of the effective channel.

In Saturation region, I_{ds} is $I_{ds} = \frac{\beta}{2} V_{GT}^2 \left(1 + \frac{V_{ds}}{V_A}\right)$

30. What is body effect coefficient?(May 2011)

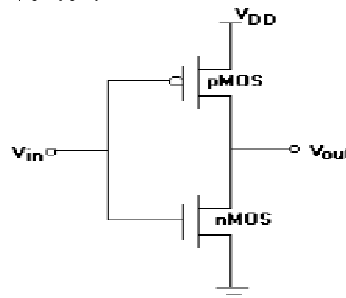
Body effect coefficient (γ) is

$$\gamma = \frac{t_{ox}}{\varepsilon_{ox}}\sqrt{2q\varepsilon_{si}N_A} = \frac{\sqrt{2q\varepsilon_{si}N_A}}{C_{ox}}$$

31. What do you mean by propagation delay time? (May 2017)

Propagation delay time (t_{pd}) (or) Maximum delay is defined as maximum time from the input crossing 50% to the output crossing 50%.

32. Draw the circuit of a CMOS inverter.



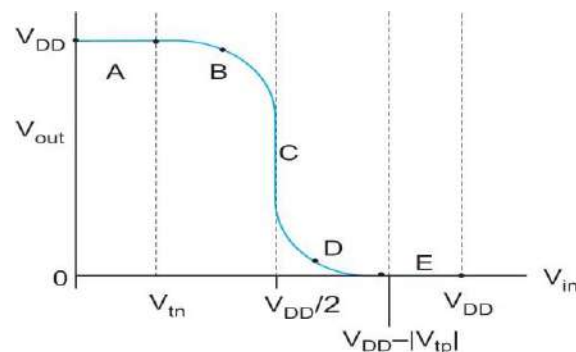
33. What are the advantages of CMOS inverter over the other inverter configurations?

a. The steady state power dissipation of the CMOS inverter circuit is negligible.

b. The voltage transfer characteristic (VTC) exhibits a full output voltage wing between 0V and V_{DD}. This results in high noise margin.

34. Draw the DC transfer characteristics of CMOS inverter. (NOV.2013, APRIL-2015)

DC transfer characteristics of CMOS inverter:



35. Define Noise margin. (May'13)

Noise margin is the amount of noise that a CMOS circuit could withstand without affect the operation of circuit.

It is basically the difference between signal value and the noise value.

NM_L (NOISE MARGIN low) = V_{IL} - V_{OL}

NM_H(NOISE MARGIN high) = V_{OH} - V_{IH}

36. What are the steps involved in the process of IC fabrication? (May 2010)

Steps involved in IC fabrication:

- Silicon wafer Preparation
- Epitaxial Growth
- Oxidation
- Photolithography

- Diffusion
- Ion Implantation
- Isolation technique
- Metallization
- Assembly processing & Packaging

37. What are the different fabrication processes available to CMOS technology?

Different CMOS processes are

- p-well process, n-well process, Twin-tub process, and Silicon On Insulator (SOI)

38. What is twin-tub process? Why it is called so?

Twin-tub process is one of the CMOS technologies in which two wells are available. The tub is also known as well. As two tubs are formed, this process is known as twin-tub process.

39. What do you mean by lateral scaling?

Lateral spacing in process of scaling the separation region between interconnect layers, keeping thickness t_{ox},, length and width as constant value.

40. What is meant by scaling? [Nov/Dec-2013]

Scaling is reducing feature size of transistor. The transistor size has reduced by 30% every two to three years. As transistors become smaller, they switch faster, dissipate less power, and are cheaper to manufacture.

41. List different types of scaling. (NOV./DEC. 2014) [APRIL-2015]

List the scaling principles. (April 2018)

Types of scaling are transistor scaling and interconnect scaling.

Types of transistor scaling are lateral scaling, constant field scaling and constant voltage scaling.

42. What are the advantages of scaling? [APRIL-2015]

Advantages of scaling are better reliability, reducing complexity and reducing IC size.

43. What is the influence of voltage scaling on power and delay? [Apr/May-2011]

Due to voltage scaling, the power dissipation will be reduced with the increase in delay (i.e) speed decreases.

44. List out the limitations of the constant voltage scaling. (Nov 2015)

Constant voltage scaling is increasing the electric fields in devices.

Voltage scaling has dramatically slowed down due to leakage. This may ultimately limit CMOS scaling.

45. Define SSI, MSI, LSI and VLSI. (May 2009)

Small scale Integration:

✓ *Small-Scale Integration* (SSI) circuits have less than 10 gates. Example: 7404 inverter.

Medium scale Integration:

✓ *Medium-Scale Integration* (MSI) circuits have up to 1000 gates. Example: 74161 counter.

Large scale Integration:

✓ *Large-Scale Integration* (LSI) circuits have up to 10,000 gates. Example: 8-bit microprocessor (8085).

Very large scale Integration:

✓ Very large scale Integration (VLSI) with gates counting upto lakhs. Example: 16-bit microprocessor (8086).

Ultra large scale Integration:

✓ Ultra Large-Scale Integration (ULSI) is the process of integrating millions of transistors on a single silicon semiconductor microchip.

46. Give objective of layout design rule.

The main objective of the layout rule is to build reliably functional circuits in as small area as possible.

47. Give propagation delay expression of CMOS inverter.

Expression of delay for rising output of CMOS inverter is $t_{PLH} = 0.69\ R_P.C_L$

Expression of delay for falling output of CMOS inverter is $t_{PHL} = 0.69\ R_N.C_L$

Propagation delay of CMOS inverter is $t_P = (t_{PLH} + t_{PHL})\ /\ 2$

48. Why NMOS device conducts strong zero and weak one? (NOV 2018)

The nMOS transistors pass 0's well but 1's poorly.

nMOS transistors attempting to pass a 1 never pull the source above $V_{DD} - V_{tn}$. This loss is called a threshold drop.

49. By What factor R_{DS} should be scaled, if constant electric filed scaling is employed? [Nov/Dec-2022]

The parameter R_{DS} [Drain Source Resistance] is scaled by 1 in constant electric field scaling.

50. List the scaling principles.

- Count of gates on chip.
- Minimum size of device.
- Power dissipation
- Maximum frequency of operation
- Die size.
- Types:
- Full scaling
- Fixed voltage scaling
- Lateral scaling

51. Why nMOS transistor is selected as pull down transistor?

- A static CMOS gate has an nMOS pull-down network to connect the output to 0 (GND).
- An NMOS device pulls the output all the way down to GND, while a PMOS lowers the output no further than $|V_{Tp}|$ — the PMOS turns off at that point, and stops contributing discharge current. NMOS transistors are hence the preferred devices in the PDN.
- NMOS device fails to raise the output above V_{DD}-V_{Tn} .

52. Name the merits of Scaling Principles.

- Minimum feature size
- Die size
- Production cost

- How Many numbers of gates are available in one chip, that also calculated.

53. What is meant channel length modulation in NMOS transistor?

The current between drain and source terminals is constant and independent of the applied voltage over the terminals. This is not entirely correct. The effective length of the conductive channel is actually modulated by the applied V_{DS}, increasing V_{DS} causes the depletion region at the drain junction to grow, reducing the length of effective channel.

The increase of depletion layer width at the drain as the drain voltage is increased. This lead to a shorter channel length and an increased drain current is called channel length modulation.

$$L_{effective} = L - \sqrt{2\frac{\varepsilon_{Si}}{qN_A}(V_{ds} - (V_{gs} - V_t))}$$

54. What are the different layers in MOS transistor?

The layers are Substrate, diffused Drain & Source, Insulator (SiO2) & Gate.

55. What are the different operating regions for an MOS transistor?

- Cutoff Region
- Non-Saturated
- (Linear) Region
- Saturated Region

56. What is Enhancement mode transistor?

The device that is normally cut-off with zero gate bias is called Enhancement mode transistor.

57. What is Depletion mode device?

The Device that conducts with zero gate bias is called Depletion mode device.

58. When the channel is said to be pinched off?

If a large V_{ds} is applied, this voltage will deplete the inversion layer. This Voltage effectively pinches off the channel near the drain.

59. What are the steps involved in manufacturing of IC?

- Silicon wafer Preparation
- Epitaxial Growth
- Oxidation
- Photolithography
- Diffusion
- Ion Implantation
- Isolation technique
- Metallization
- Assembly processing & Packaging

60. What is meant by Epitaxy?

Epitaxy means arranging atoms in single crystal fashion upon a single crystal substrate.

61. What are the processes involved in photo lithography?

Masking process
Photo etching process.

62. What is the purpose of masking in fabrication of IC?

Masking is used to identify the location in which Ion Implantation should not take place.

63. What are the materials used for masking?

Photo resist, Si02, SiN, Poly Silicon.

64. What are the types of Photo etching?

Wet etching
Dry etchings are the types of photo etching.

65. What is diffusion process? What are doping impurities?

Diffusion is a process in which impurities are diffused into the Silicon chip at 1000^0C temperature. B203 and P205 are used as impurities used.

66. What is Ion Implantation process?

It is process in which the Si material is doped with an impurity by making the accelerated impurity atoms to strike the Si layer at high temperature.

67. What are the various Silicon wafer Preparation?

- Crystal growth & doping
- Ingot trimming & grinding
- Ingot slicing
- Wafer polishing & etching
- Wafer cleaning.

68. What are the different types of oxidation?

The two types of oxidation are Dry & Wet Oxidation.

69. What is Isolation?

It is a process used to provide electrical isolation between different components and interconnections.

70. Give the different types of CMOS process.

- p well process
- n well process
- twin tub process
- SOI process

71. What is Channel stop Implantation?

In n well fabrication, n well is protected with the resist material. (Because, it should not be affected during Boron implantation). Then Boron is implanted except n well. The above said process is done using photo resist mask. This type of implantation is known as Channel stop implantation.

72. What is LOCOS?

LOCOS mean Local Oxidation of Silicon. This is one type of oxide construction.

73. What is SWAMI?

SWAMI means Side Wall Masked Isolation. It is used to reduce bird's beak effect.

74. What is LDD?

LDD means Lightly Doped Drain Structures. It is used for implantation of n region In n-well process.

75. What is Twin tub process? Why it is called so?

Twin tub process is one of the CMOS technologies. Two wells (the other name for well is Tub) are created in this process. So, because of these two tubs, this process is known as Twin tub process.

76. What are the steps involved in Twin tub process?

- Tub Formation
- Thin oxide Construction
- Source & Drain Implantation
- Contact cut definition
- Metallization.

77. Name the special features of Twin tub process.

In Twin tub process, Threshold voltage, body effects of n and p devices are independently optimized.

78. List out the advantages of Twin tub process?

Advantages of Twin tub process are (1) Separate optimized wells are available. (2) Balanced performance is obtained for n and p transistors.

79. What is SOI? What is the material used as Insulator?

SOI means Silicon on Insulator. In this process, a Silicon based transistor is built on an insulating material like Sapphire or SiO.

80. What are the advantages and disadvantages of SOI process?

Advantages of SOI process:

- There is no well formation in this process.
- There is no Field Inversion problem.
- There is no body effect problem.

Disadvantages of SOI process:

- It is very difficult to protect inputs in this process.
- Device gain is low.
- The coupling capacitance between wires always exists.

81. What are the advantages of CMOS process?

- Low Input Impedance
- Low delay Sensitivity to load.

82. Define Short Channel devices.

Transistors with Channel length less than 3- 5 microns are termed as Short channel devices. With short channel devices the ratio between the lateral & vertical dimensions are reduced.

83. What are the advantages of Silicon-on-Insulator process?

- No Latch-up
- Due to absence of bulks transistor structures are denser than bulk silicon.

84. What are the advantages of CMOS process?

- Low power Dissipation
- High Packing density
- Bi directional capability

85. What is the fundamental goal in Device modeling?

To obtain the functional relationship among the terminal electrical variables of the device that is to be modeled.

86. What is CMOS Technology?

The fabrication of an IC using CMOS transistors is known as CMOS Technology. CMOS transistor is nothing but an inverter, made up of an nMOS and pMOS transistor connected in series.

87. Give the advantages of CMOS IC?

- Size is less
- High Speed
- Less Power Dissipation

88. What are four generations of Integration Circuits?

- SSI (Small Scale Integration)
- MSI (Medium Scale Integration)
- LSI (Large Scale Integration)
- VLSI (Very Large Scale Integration)

89. Give the variety of Integrated Circuits.

- More Specialized Circuits
- Application Specific Integrated Circuits (ASICs)
- Systems On Chips.

90. Why NMOS technology is preferred more than PMOS technology?

N-channel transistors have greater switching speed when compared to PMOS transistors. Hence, NMOS is preferred than PMOS.

****

UNIT II - COMBINATIONAL LOGIC CIRCUITS

Propagation Delays, stick diagram, Layout diagrams, Examples of combinational logic design, Elmore's constant, Static Logic Gates, Dynamic Logic Gates, Pass Transistor Logic, Power Dissipation, Low Power Design principles.

2.1 : Delay estimation

- **Draw a CMOS inverter. Analyze the switching characteristics during rise time when V_{in} change from high to low. (April 2019-7M)**
- **Derive an expression for the rise time, fall time and propagation delay of a CMOS inverter. (DEC 2013, APRIL-2015) [Nov 2019] [Nov/Dec 2022]**

- ✓ Important definitions for delay estimation:

Propagation delay time (t_{pd}):

- ✓ Propagation delay time is defined as maximum time from the input crossing 50% to the output crossing 50%.

Contamination delay time (t_{cd}):

- ✓ Contamination delay time is defined as minimum time from the input crossing 50% to the output crossing 50%.

Rise time (t_r):

- ✓ Rise time is defined as time for a waveform to rise from 20% to 80% of its steady-state value

Fall time (t_f):

- ✓ Fall time is defined as time for a waveform to fall from 80% to 20% of its steady-state value
- ✓ Edge rate is average of rise and fall time, (t_{rf}) = (t_r + t_f)/2

Delay estimation response curve:

- ✓ When an input changes, the output will retain its old value for at least the contamination delay and take on its new value in, at most the propagation delay.
- ✓ Delays for the output rising is t_{pdr} /t_{cdr} and the output falling is t_{pdf} /t_{cdf} .
- ✓ Rise/fall times are also called as slopes or edge rates.
- ✓ Propagation and contamination delay times are also called as max-time and min-time respectively.

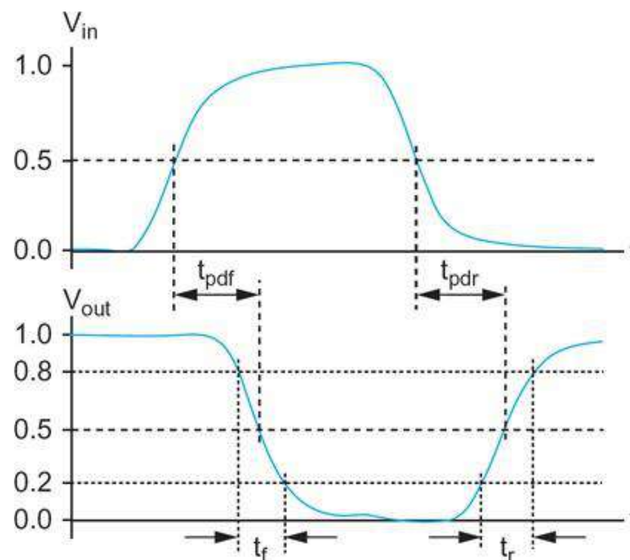


Figure: Delay estimation of CMOS inverter

- ✓ The gate that charges or discharges a node is called the driver. The gates and wires being driven, are called the load. Propagation delay is usually called as delay.
- ✓ Arrival times and propagation delays are defined separately for rising and falling transitions.
- ✓ The delay of a gate may be different from different inputs. Earliest arrival times can also be computed based on contamination delays.
- ✓ Expression of delay for rising output is $t_{PLH} = 0.69\ R_P.C_L$
  Where, R_P – effective resistance of pMOS transistor
  C_L - load capacitance of CMOS inverter.
- ✓ Expression of delay for falling output is $t_{PHL} = 0.69\ R_N.C_L$
  Where, R_N – effective resistance of nMOS transistor
- ✓ Propagation delay of CMOS inverter is $t_P = (t_{PLH} + t_{PHL}) / 2$

2.1.1 : RC Delay Model:

Discuss in detail about the resistive and capacitive delay estimation of a CMOS inverter circuit. (MAY 2013) (or)
Briefly explain about the RC delay model.

- ✓ RC delay model approximates the nonlinear transistor I-V and C-V characteristics with an average resistance and capacitance over the switching range of the gate.

Effective Resistance:

- ✓ The RC delay model treats a transistor as a switch in series with a resistor.
- ✓ The effective resistance is the ratio of V_{ds} to I_{ds}.
- ✓ A unit nMOS transistor is defined to have effective resistance R.
- ✓ An nMOS transistor of k times unit width has resistance R/k, because it delivers k times as much current.
- ✓ A unit pMOS transistor has greater resistance, generally in the range of 2R–3R, because of its lower mobility.

- ✓ According to the long-channel model, current decreases linearly with channel length (L) and hence resistance is proportional to L.

Gate and Diffusion Capacitance:

- ✓ Each transistor has gate and diffusion capacitance.
- ✓ C is the gate capacitance of a unit transistor. A transistor of k times unit width has capacitance kC.
- ✓ Diffusion capacitance depends on the size of the source/drain region.
- ✓ Wider transistors have proportionally greater diffusion capacitance. Increasing channel length, increases gate capacitance proportionally but does not affect diffusion capacitance.

Equivalent RC Circuits:

- ✓ Figure shows equivalent RC circuit models for nMOS and pMOS transistors of width k with contacted diffusion on both source and drain.
- ✓ The pMOS transistor has approximately twice the resistance of the nMOS transistor, because holes have lower mobility than electrons.

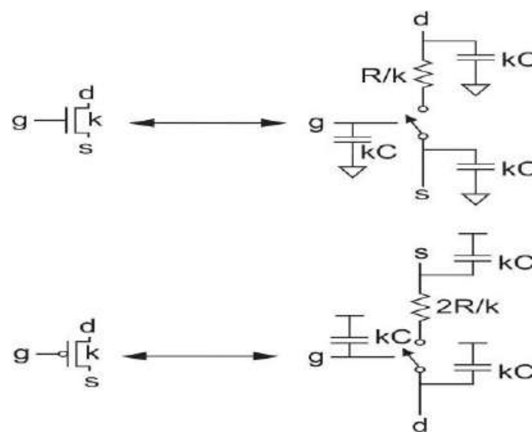


Figure: RC model of nMOS &pMOS transistors

****

2.2 : Stick diagram

- ❖ **Explain about stick diagram in VLSI design. (April 2008)**
- ❖ **Draw the static diagram of CMOS inverter. (April 2019-7M)**

- ✓ A **stick diagram** is a cartoon of a chip layout. A "stick diagram" is a paper and pencil tool that use to plan the layout of a cell.
- ✓ The stick diagram resembles the actual layout, but uses "sticks" or lines to represent the devices and conductors. Figure 17, shows a stick diagram for an inverter.

- ✓ The stick diagram represents the rectangles with lines, which represent wires and component symbols.
- ✓ The stick diagram does not represent all the details of a layout, but it makes some relationship much clearer and it is simple to draw.
- ✓ Layouts are constructed from rectangles, but stick diagrams are built from cartoon symbols for components and wires.

Stick diagram Rules:

- ✓ **Rule 1:** When two or more 'sticks' of the same type cross or touch each other, that represents electrical contact.
- ✓ **Rule 2:** When two or more 'sticks' of the different type cross or touch each other, there is no electrical contact. If electrical contact is needed, we have to show the connection explicitly.
- ✓ **Rule 3:** When a poly crosses diffusion, it represents a transistor. If a contact is shown, then it is not a transistor. A transistor exists where a polysilicon (red) stick crosses either an n-diffusion (green) stick or a p-diffusion (yellow) stick.
- ✓ **Rule 4:** In CMOS, a demarcation line is drawn to avoid touching of p-diff with n-diff. All pMOS must lie on one side of the line and all nMOS will have to be on the other side.

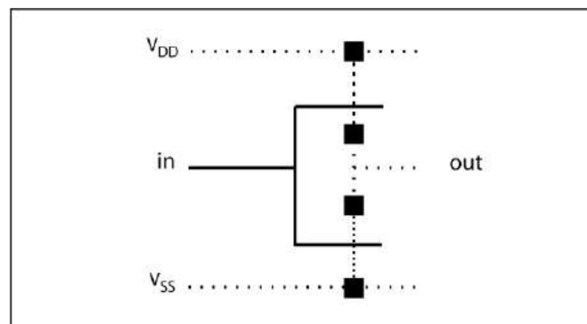


Figure 17: Stick diagram for an inverter

The symbols for wires used on various layers are shown in Figure 18.

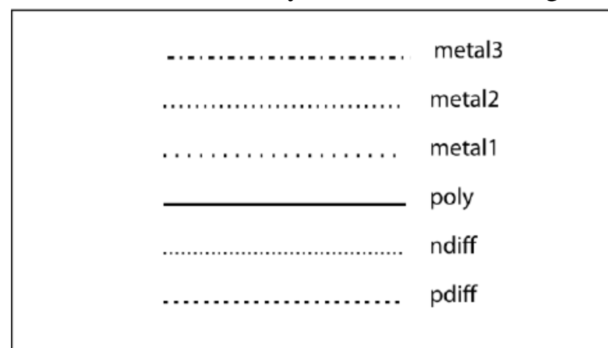


Figure 18: Symbols for wires used on various layers

- ✓ **Drawing stick diagrams in color**: Red for poly, green for n-diffusion, yellow for p-diffusion, and shades of blue for metal are typical colors.
- ✓ A few simple rules for constructing wires from straight-line segments ensure that, the stick diagram corresponds to a feasible layout.
- ✓ Wires cannot be drawn at arbitrary angles. Only horizontal and vertical wire segments are allowed.
- ✓ Two wire segments on the same layer, which cross are electrically connected.

- ✓ Vias to connect wires, which do not normally interact, are drawn as black dots.
- ✓ Figure 19, shows the stick figures for transistors.
- ✓ Each type of transistor is represented as poly and diffusion crossings, much as in the layout.

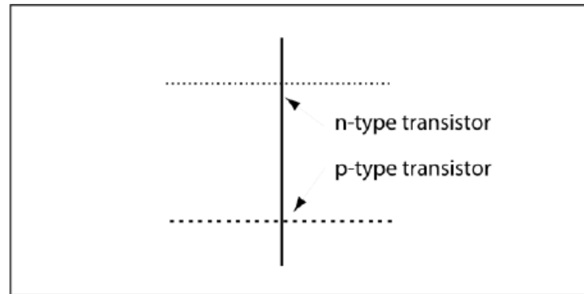
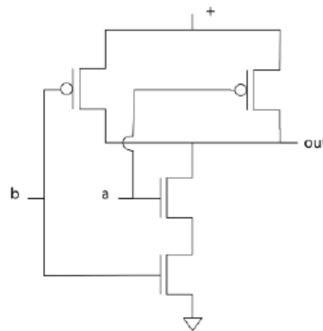


Figure 19: Stick figures for transistors

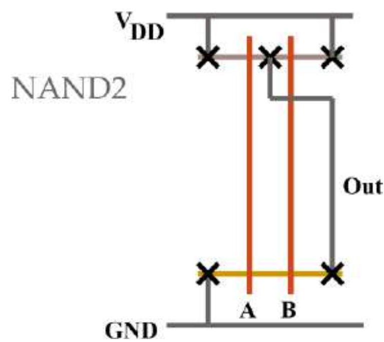
- ✓ Area and aspect ratio are also difficult to estimate from stick diagrams.
- ✓ Stick diagrams are especially important tools for layouts built from large cells and for testing the connections between cells.

Example:1

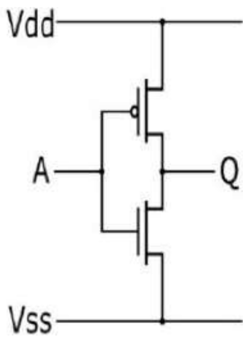
Here is the transistor schematic for a two-input NAND gate:

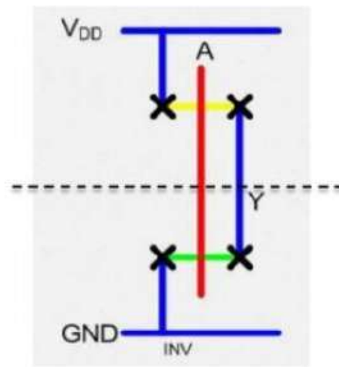


And here is a stick diagram for the two-input NAND:

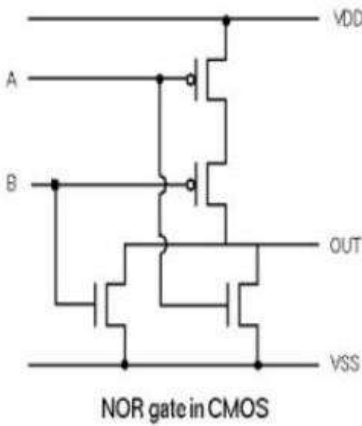


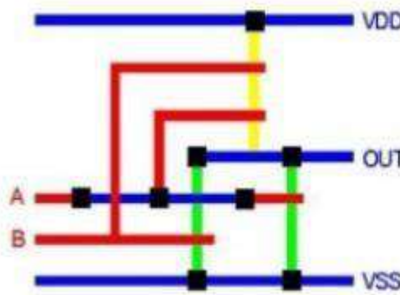
Example: 2
Draw the Stick diagram of CMOS Inverter



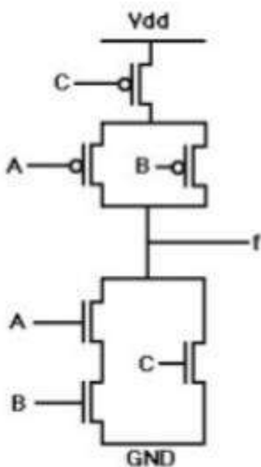


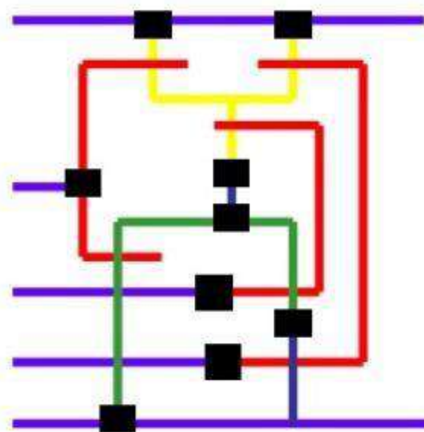
Example: 3
Draw the Stick diagram of CMOS NOR gate





Example: 4
Draw the stick diagram of [(A . B) +C]'





****

2.3 : Layout Design Rules and Gate Layouts

- ❖ **Draw and explain briefly the n-well CMOS design rules. (NOV 2007, April 2008, MAY 2014)**
- ❖ **Discuss in detail with a neat layout, the design rules for a CMOS inverter.**
- ❖ **Write the layout design rules and draw diagram for four input NAND and NOR. (Nov 2016) (April 2018)**
- ❖ **State the minimum width and minimum spacing lambda based design rules to draw the layout. (April 2019-6M)**

- ✓ Layout rules also referred to as **design rules**.
- ✓ It can be considered as prescription for preparing the photomasks, which are used in the fabrication of integrated circuits.
- ✓ The rules are defined in terms of feature sizes (widths), separations and overlaps.
- ✓ The main **objective of the layout rules is to build reliable functional circuits in as small area as possible.**
- ✓ Layout design rules describe how small features can be and how closely they can be reliably packed in a particular manufacturing process.
- ✓ Design rules are a set of geometrical specifications that dictate the design of the layout masks.
- ✓ A design rule set provides numerical values for minimum dimensions and line spacing.
- ✓ Scalable design rules are based on a single parameter (λ), which characterizes the resolution of the process. λ is generally half of the minimum drawn transistor channel length.
- ✓ This length is the distance between the source and drain of a transistor and is set by the minimum width of a polysilicon wire.

Lambda based rule (Scalable design rule):

- ✓ Lambda-based rules are round up dimensions of scaling to an integer multiple of λ.
- ✓ Lambda rules make scaling layout small. The same layout can be moved to a new process, simply by specifying a new value of λ.
- ✓ The minimum feature size of a technology is characterized as 2λ.

Micron Design Rules (Absolute dimensions):

- ✓ The MOSIS rules are expressed in terms of lambda.
- ✓ These rules allow some degree of scaling between processes.
- ✓ Only need to reduce the value of lambda and the designs will be valid in the next process down in size.
- ✓ These processes rarely shrink uniformly.
- ✓ Thus, industry usually uses the actual micron design rules for layouts.
- ✓ There are set of micron design rules for a hypothetical 65 nm process.
- ✓ We can observe that, these rules differ slightly but not immensely from lambda based rules with lambda = 0.035 micro meter.
- ✓ Upper level metal rules are highly variable depending on the metal thickness. Thicker wires require greater widths, spacing and bigger vias.

- ✓ Two metal layers in an n-well process has the following:

- Metal and diffusion have minimum width and spacing of 4 λ.
- Contacts are 2 λ × 2 λ and must be surrounded by 1 λ on the layers above and below.
- Polysilicon uses a width of 2 λ.
- Polysilicon overlaps diffusion by 2 λ where a transistor is desired and has a spacing of 1 λ away where no transistor is desired.
- Polysilicon and contacts have spacing of 3 λ from other polysilicon or contacts.
- N-well surrounds pMOS transistors by 6 λ and avoids nMOS transistors by 6 λ.

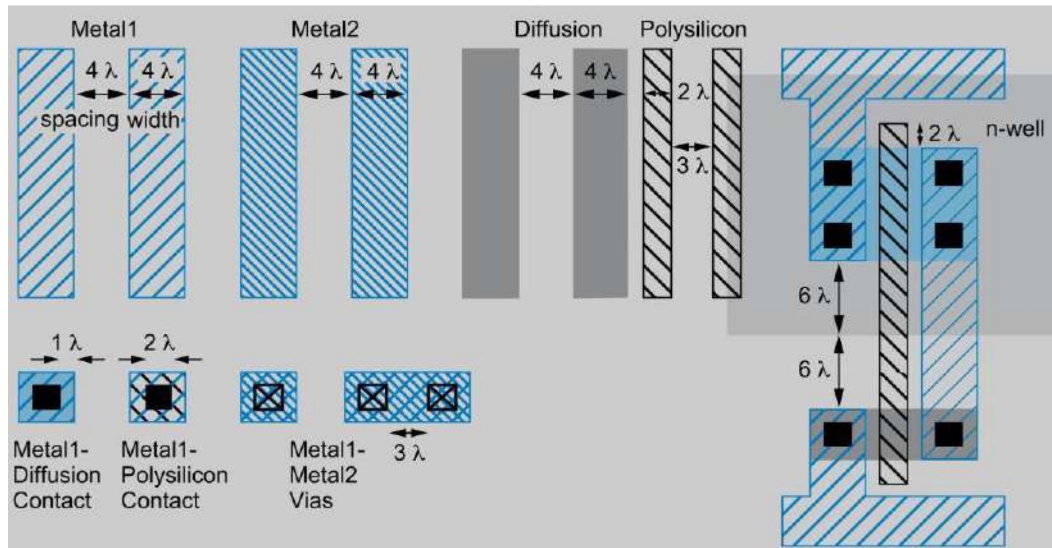


Figure: Simplified λ -based design rules with CMOS inverter layout diagram

Design Rule:

Well Rules:

- ✓ The n-well is usually a deeper implant than the transistor source/drain implants.
- ✓ Therefore, it is necessary to provide sufficient clearance between the n-well edges and the adjacent *n*+ diffusions.

Transistor Rules:

- ✓ CMOS transistors are generally defined by at least four physical masks.
- ✓ There are active (also called diffusion, diff, thinox, OD, or RX), n-select (also called n-implant, n-imp, or nplus), p-select (also called p-implant, pimp, or pplus) and polysilicon (also called poly, polyg, PO, or PC).
- ✓ The active mask defines all areas, where n- or p-type diffusion is to be placed *or* where the gates of transistor are to be placed.

Contact Rules:

- ✓ There are several generally available contacts:
  - Metal to p-active (p-diffusion)
  - Metal to n-active (n-diffusion)
  - Metal to polysilicon
  - Metal to well or substrate

Metal Rules:

- ✓ Metal spacing may vary with the width of the metal line.

- ✓ Metal wire width of minimum spacing may be increased. This is due to etch characteristics versus large metal wires.

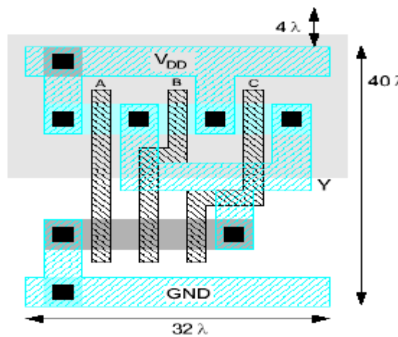
Via Rules:

- ✓ Processes may allow vias to be placed over polysilicon and diffusion regions.
- ✓ Some processes allow vias to be placed within these areas, but do not allow the vias to the boundary of polysilicon or diffusion.

Example: NAND3

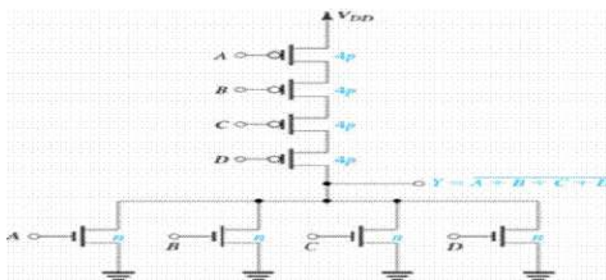
Draw the gate layout diagram of NAND. (May 2017)

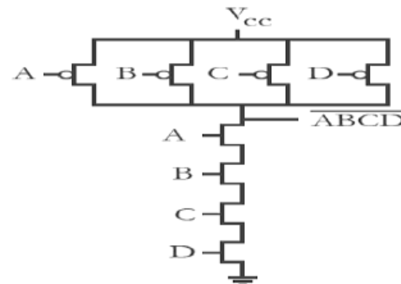
- Horizontal N-diffusion and p-diffusion strips
- Vertical polysilicon gates
- Metal1 V_{DD} rail at top
- Metal1 GND rail at bottom



❖ **Draw diagram for four input NAND and NOR gate. (Nov 2017)**

4 input NOR gate **4 input NAND gate**



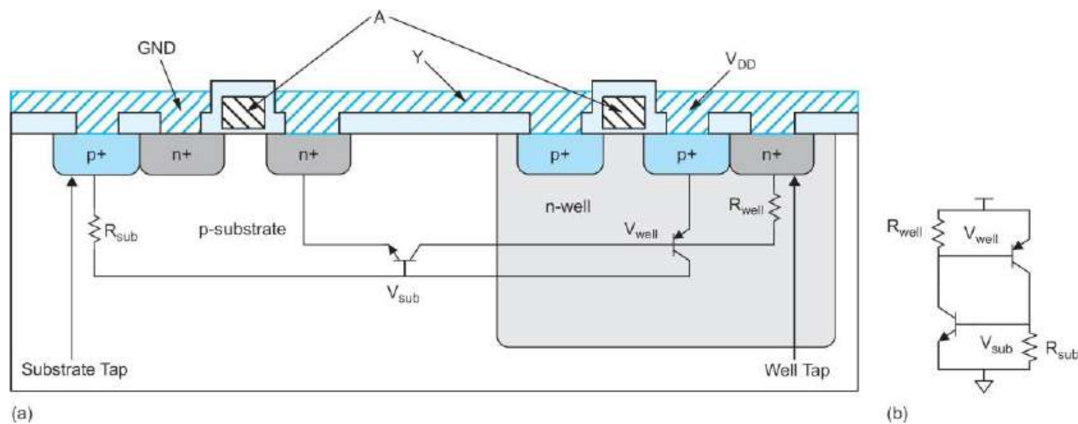


****

2.4 : Latchup problem:

❖ ***Discuss the orgin of latch up problems in CMOS circuits with necessary diagrams. Explain the remedial measures. (Nov 2007, April 2008)***

- ✓ A CMOS process is slowed down by developing low-resistance paths between V_{DD} and GND, causing catastrophic meltdown. The phenomenon is called latchup.
- ✓ Latchup problem arises when parasitic bipolar transistors are formed by the substrate, well and diffusion.
- ✓ The cause of the latchup effect can be understood by examining the process cross-section of a CMOS inverter, as shown in Figure (a).
- ✓ The schematic shows, a circuit composed of an npn-transistor, a pnp-transistor, and two resistors connected between the power and ground rails (Figure (b)).



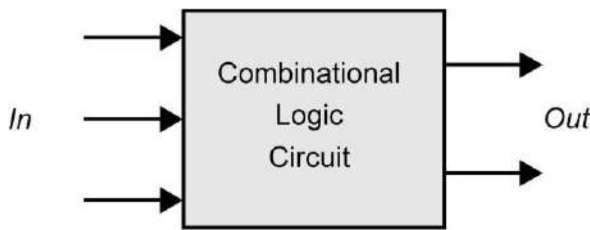
- ✓ The npn transistor is formed between the grounded n-diffusion source of the nMOS transistor, the p-type substrate and the n-well.
- ✓ The resistors are due to the resistance through the substrate or well to the nearest substrate and well taps.
- ✓ The cross-coupled transistors form a bistable silicon-controlled rectifier (SCR). Both parasitic bipolar transistors are OFF.
- ✓ Latchup can be triggered, when transient currents flow through the substrate during normal chip power-up.

- ✓ Latchup prevention is easily accomplished by
  - Minimizing Rsub and Rwell.
  - Use of guard rings
- ✓ SOI process avoids latchup entirely, because they have no parasitic bipolar structures.

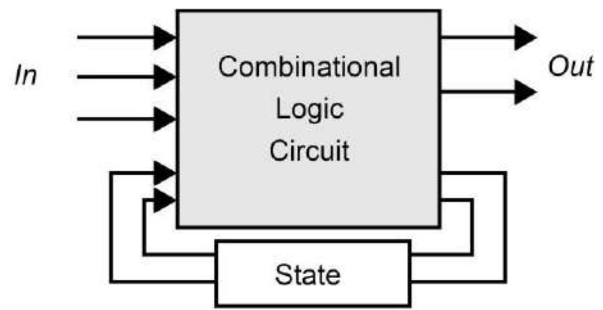
2.5 Introduction (Combinational Logic Circuit):

CMOS logic

- Digital logics are divided into combinational and sequential circuits.
- Combinational circuits are circuits where outputs depend only on the present inputs.
- For sequential or regenerative circuit, the output is not only a function of the current input data, but also of previous values of the input signals.
- A sequential circuit includes a combinational logic portion and a memory module that holds the state. Example are registers, counters and memory.
- The building blocks for combinational circuits are logic gates, while the building blocks for sequential circuits are registers and latches.
- The delay of a logic gate depends on its output current I, load capacitance C and output voltage swing ΔV.



(a) Combinational



(b) Sequential

- Alternative (ratioed circuits, dynamic circuits and pass transistor circuits) CMOS logic configurations are called circuit families.
- nMOS transistors provide more current than pMOS for the same size and capacitance, so nMOS networks are preferred.

Examples of combinational circuits

(i) CMOS inverter:

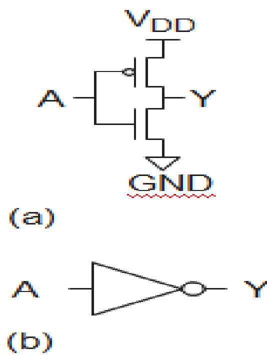


Figure: Inverter (a) schematic (b) symbol $Y = \bar{A}$

(ii) **Two input NAND gate:**

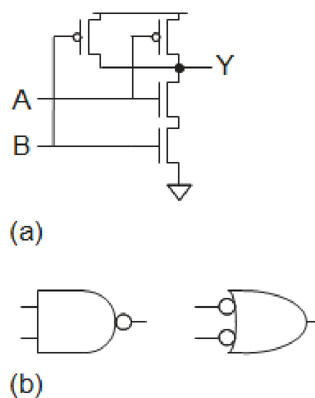


Figure: 2 input NAND gate (a) schematic (b) symbol

(iii) **Three input NAND gate:**

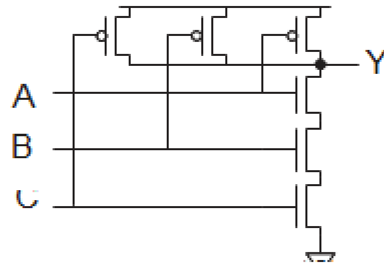


Figure: 3 –input NAND gate Schematic Y=A.B.C

(iv) **Two input NOR gate:**

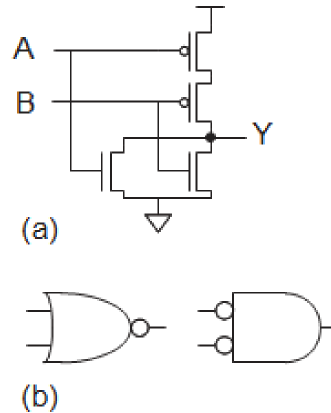
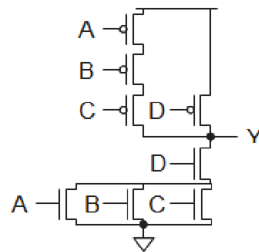


Figure: 2-input NOR gate (a) schematic (b) Symbol $Y = \overline{A + B}$

Example:

Sketch a static CMOS gate computing $Y = \overline{(A + B + C) \cdot D}$.



****

2.6 : *Elmore's Delay*

What is meant by Elmore's delay and give expression for Elmore's delay?

- The Elmore delay model estimates the delay from a source, switching to one of the leaf nodes. Delay is the sum over each node i of the capacitance C_i on the node multiplied by the effective resistance R.

Propagation delay time :

$$t_{pd} \approx \sum_{\text{nodes } i} R_{i-to-source} C_i$$
$$= R_1C_1 + (R_1 + R_2)C_2 + ... + (R_1 + R_2 + ... + R_N)C_N$$

- Delay of an ideal fanout-of-1 inverter with no parasitic capacitance is $\tau = 3RC$.
- The normalized delay d relative to this inverter delay:

$$d = \frac{t_{pd}}{\tau}$$

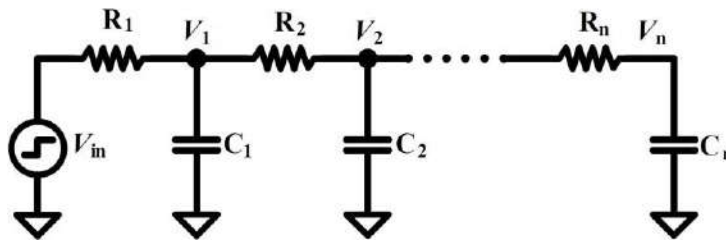


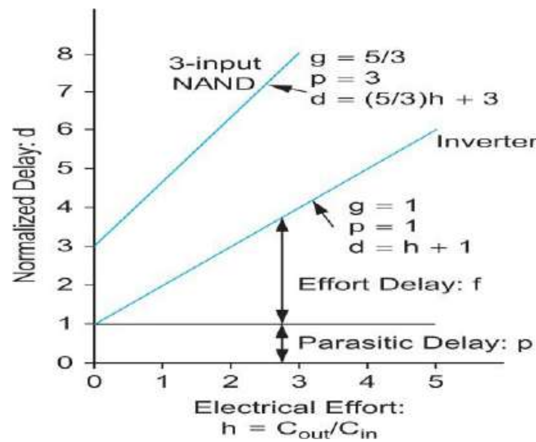
Figure: RC delay equivalent for series of transistors

Linear delay model

- The RC delay model is one, where delay is a linear function of the fanout of a gate.
- The normalized delay of a gate can be expressed in units of Y as d = f + p.
  Where p is the parasitic delay inherent to the gate when no load is attached.
  f is the effort delay or stage effort that depends on the complexity.
- Effort delay of the gate is f = gh.
  Where g is the logical effort (An inverter has a logical effort of 1).

- Logical effort is defined as the ratio of the input capacitance of a gate to the input capacitance of an inverter delivering the same output current.
- h is the fanout or electrical effort. Electrical effort is defined as ratio of the output capacitance to input capacitance.
- More complex gates have greater logical efforts, indicating that they take longer time to drive a given fanout.
- For example, the logical effort of the 3-input NAND gate is 5/3.
- The electrical effort can be computed as $h = \frac{C_{out}}{C_{in}}$

  Where C_{out} is the capacitance of the external load being driven and C_{in} is the capacitance of the gate.
- Normalized delay vs electrical effort for an idealized inverter and 3-input NAND gate shown in diagram.
- The y-intercepts indicate the parasitic delay. The slope of the lines is the logical effort.
- The inverter has a slope of 1. The NAND gate has a slope of 5/3.



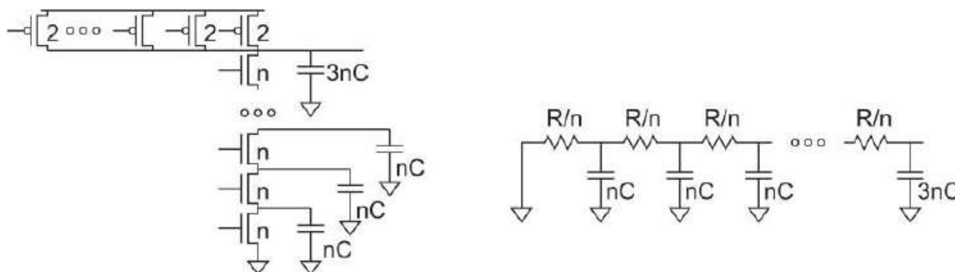
Design a four input NAND gate and obtain its delay during the transition from high to low. (April 2018)

Figure shows a model of an n-input NAND gate in which the upper inputs were all 1 and the bottom input rises. The gate must discharge the diffusion capacitances of all of the internal nodes as well as the output.

Elmore delay is

$$t_{pd} = R(3nC) + \sum_{i=1}^{n-1}\left(\frac{iR}{n}\right)(nC) = \left(\frac{n^2}{2} + \frac{5}{2}n\right)RC$$

Figure: n-input NAND gate parasitic delay



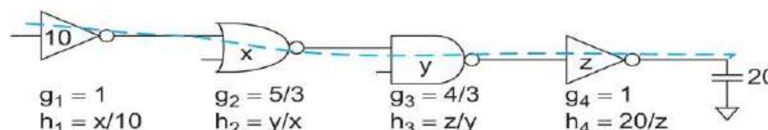
Delay for 4 input NAND gate: $\left(\frac{n^2}{2} + \frac{5}{2}n\right)RC = \left(\frac{4^2}{2} + \frac{5}{2}4\right)RC = \left(\frac{16}{2} + \frac{20}{2}\right)RC = 18RC$

Logical effort

Obtain the logical effort and path efforts of the given circuit. (April 2018)

Delay in Multistage Logic Networks:

The figure shows the logical and electrical efforts of each stage in a multistage path as a function of the sizes of each stage.



The path of interest (the only path in this case) is marked with the dashed blue line. Observe that logical effort is independent of size, while electrical effort depends on sizes.

The path logical effort G can be expressed as the products of the logical efforts of each stage along the path.

$$G = \prod g_i$$

The path electrical effort H can be given as the ratio of the output capacitance the path must drive divided by the input capacitance presented by the path

$$H = \frac{C_{out(path)}}{C_{in(path)}}$$

The path effort F is the product of the stage efforts of each stage.

$$F = \prod f_i = \prod g_i h_i$$

Introduce an effort to account for branching between stages of a path. This branching effort b is the ratio of the total capacitance seen by a stage to the capacitance on the path.

$$b = \frac{C_{onpath} + C_{offpath}}{C_{onpath}}$$

The path branching effort B is the product of the branching efforts between stages.

$$B = \prod b_i$$

The path effort (F) is defined as the product of the logical, electrical, and branching efforts of the path. The product of the electrical efforts of the stages is actually BH, not just H.

F = GBH

Compute the delay of a multistage network. The path delay D is the sum of the delays of each stage. It can also be written as the sum of the path effort delay DF

$$D = \sum d_i = D_F + P$$
$$D_F = \sum f_i$$
$$P = \sum p_i$$

The product of the stage efforts is F, independent of gate sizes. The path effort delay is the sum of the stage efforts. The sum of a set of numbers whose product is constant is minimized by choosing all the numbers to be equal.

The path delay is minimized when each stage bears the same effort. If a path has N stages and each bears the same effort, that effort must be

f̂ = gi hi = F 1 / N

Thus, the minimum possible delay of an N-stage path with path effort F and path parasitic delay P is

D = NF 1/ N + P

It shows that the minimum delay of the path can be estimated knowing only the number of stages, path effort, and parasitic delays without the need to assign transistor sizes.

The capacitance transformation formula is used to find the best input capacitance for a gate given the output capacitance it drives. $C_{in_i} = \frac{C_{out_i} * g_i}{\hat{f}}$

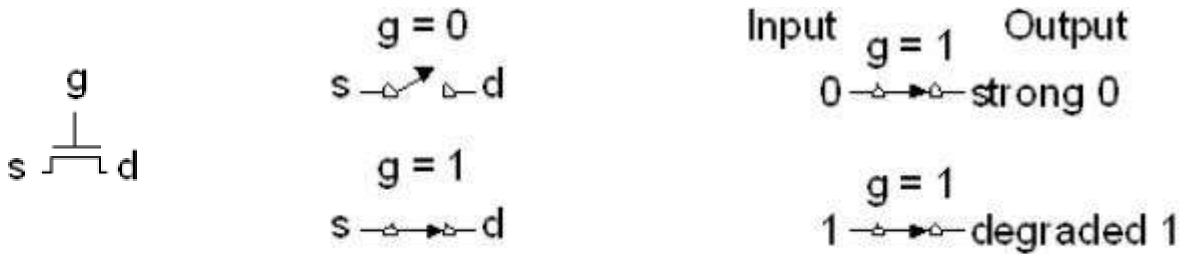
At the end of the path, apply the capacitance transformation to determine the size of each stage. Check the arithmetic by verifying that the size of the initial stage matches the specification.

****

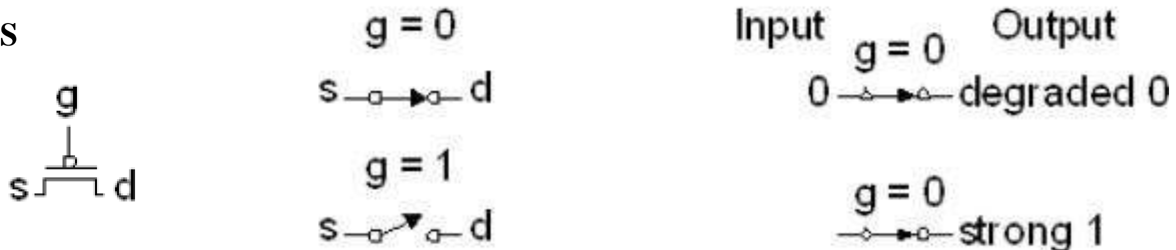
2.7 Pass Transistor

Write short notes on pass transistor.

nMos



pMOS



- Transistors can be used as switches

Signal Strength

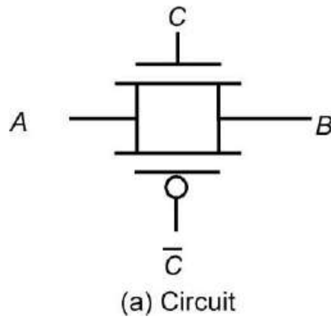
- How close it approximates ideal voltage source
- VDD and GND rails are strongest 1 and 0
- nMOS pass strong 0, But degraded or weak 1
- pMOS pass strong 1,But degraded or weak 0
- Thus NMOS are best for pull-down network
- Thus PMOS are best for pull-up network

****

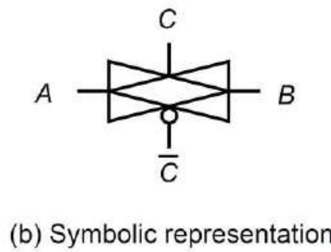
2.8 Transmission Gates

Write short notes on transmission gates (TG).

- By connecting an nMOS and a pMOS transistor in parallel, we obtain a switch that turns on when a 1 is applied to the gate terminal in which 0's and 1's are both passed in an acceptable fashion.
- We term this a ***transmission gate or pass gate***.



(a) Circuit



(b) Symbolic representation

Input Output

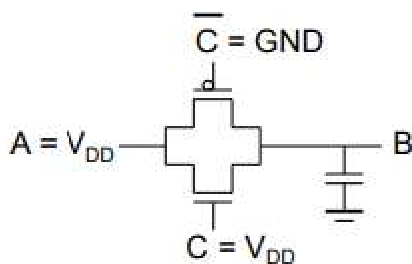
g = 0, gb = 1
a b

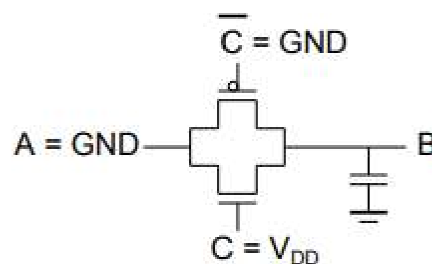
g = 1, gb = 0
a b

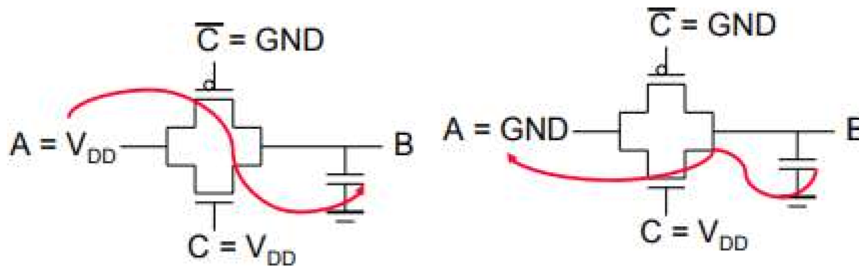
g = 1, gb = 0
0 strong 0

g = 1, gb = 0
1 strong 1

- In a circuit where only a 0 or a 1 has to be passed, the appropriate transistor (n or p) can be deleted, reverting to a single nMOS or pMOS device.
- Note that, both the control input and its complement are required by the transmission gate. This is called ***double rail logic***.
- When the control input is low(control =0), the switch is open, and when the control is high (control=1) the switch is closed.







- Full swing bidirectional switch controlled by the gate signal C, A = B if C = 1

****

2.9 Cascaded CMOS Inverter

Derive the generalized expression for propagation delay of N-cascaded CMOS inverters if 'N' is even and if 'N' is odd. [Nov 2019]

- Assume a signal is available at the output of a minimum size inverter and that it is to drive a load C_L.

The average propagation delay associated with driving this load directly is,

$$t_{dir} = \frac{t_{apd} C_L}{C_G}$$

Where, t_{apd} is the average logic stage delay and C_G is the input capacitance of the reference inverter.

For any integer n ≥ 1, define **α by the expression**,

$$\alpha = \left(\frac{C_L}{C_G}\right)^{1/n}$$

Alternatively, **n can be represented in terms of α as**

$$n = \frac{\ln(C_L/C_G)}{\ln \alpha}$$

This structure is composed of a cascade of n inverters each sized by the 4 : 1 sizing rule and each with a drive capability that is α times as large as the previous stage.

The width and length of the kth stage can be characterized by the equations,

$$W_{dk} = \alpha^{k-1} W_{d1}$$

$$W_{uk} = W_{dk}$$

$$L_{dk} = L_{d1}$$

$$L_{uk} = 4L_{dk}$$

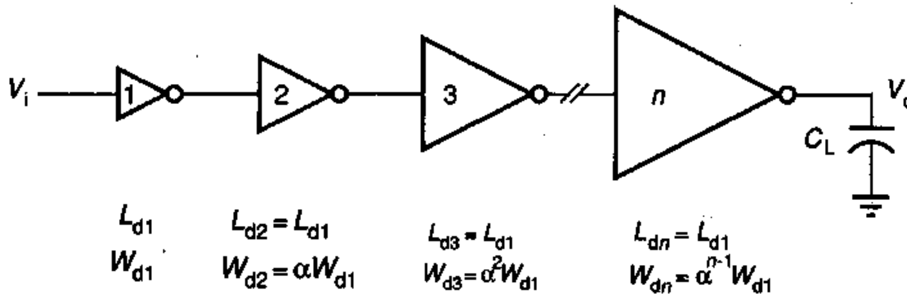
where,

W_{dk} and L_{dk} are device dimensions correspond to the pull down transistor

W_{uk} and L_{uk} are device dimensions correspond to the pull up transistor

The load on the kth stage C_{LK} is,

$$C_{Lk} = \alpha^k C_G$$



$$t_{path} \approx t_{apd} \sum_{i=1}^{N} \frac{m_i + f_i}{\theta_i}$$

The average propagation delay of the first inverter is αt_{apd}

Hence, it follows from above equation with $m_i = 0$ and $(f_i / \theta_i) = \alpha$

The total delay for the cascade is,

$$t_{cas} = n \alpha t_{apd}$$

Let r be the ratio between the propagation delays of the direct drive circuit and of the geometric cascade approach.

$$r = \frac{t_{cas}}{t_{dir}} = \frac{n \alpha t_{apd}}{t_{apd} C_L / C_G} = \frac{n \alpha C_G}{C_L}$$

It is our goal to determine n and α to minimize r and thus minimize the propagation delay in driving the load.

Therefore, n can be eliminated from the expression for r to obtain the expression.

$$r = \frac{\ln(C_L/C_G)}{C_L/C_G}\frac{\alpha}{\ln \alpha}$$

****

2.10 Circuit Families

Briefly discuss about the classification of circuit families and comparison of the circuit families. (May 2014, APRIL-2015)
Draw the CMOS logic circuit for the Boolean expression Z= $\overline{A(B+C)+DE}$ **and explain. (April 2018)**

Draw and explain the function of static CMOS.

2.10.1: Static CMOS

- Static CMOS circuits with complementary nMOS pulldown and pMOS pullup networks are used for the majority of logic gates in integrated circuits.

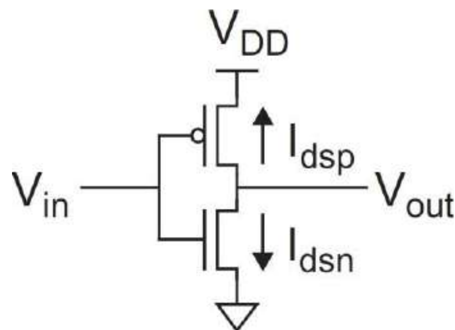


Figure: Static CMOS inverter

Advantages of static CMOS:

- Static CMOS circuits have good noise margins
- Static CMOS circuits are fast, low power, easy to design.
- Static CMOS circuits are widely supported by CAD tools.
- Static CMOS circuits are available in standard cell libraries.

Drawback of static CMOS

- It requires both nMOS and pMOS transistors for each input.
- It has a relatively large logical effort.
- Gate delay is increased.

a. Bubble pushing

- CMOS stages are inherently inverting, so AND and OR functions must be built from NAND and NOR gates.
- DeMorgan's law helps with this conversion:

$$\overline{A.B} = \overline{A} + \overline{B}$$
$$\overline{A+B} = \overline{A}.\overline{B}$$

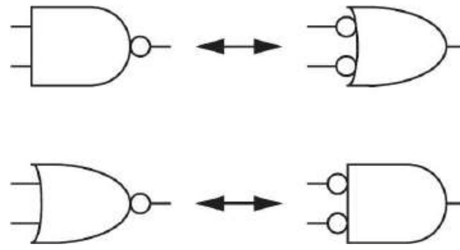


Figure: Bubble pushing with DeMorgan's law

- A NAND gate is equivalent to an OR of inverted inputs.
- A NOR gate is equivalent to an AND of inverted inputs.
- The same relationship applies to gates with more inputs.
- Switching between these representations is easy and is often called bubble pushing.

b. Compound Gates

- Static CMOS also efficiently handles compound gates computing various inverting combinations of AND/OR functions in a single stage.
- The function F = AB +CD can be computed with an AND-OR INVERT- 22 (AOI22) gate and an inverter, as shown in Figure.

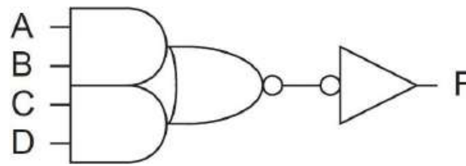


Figure: Logic using AOI22 gate

- Logical effort of compound gates can be different for different inputs.
- Figure shows, how logical efforts can be estimated for the AOI21, AOI22 and a more complex compound AOI gate.

Q: Design a circuit described by the Boolean function Y=[A.(B+C)(D+E)]'using CMOS logic. (NOV 2021)

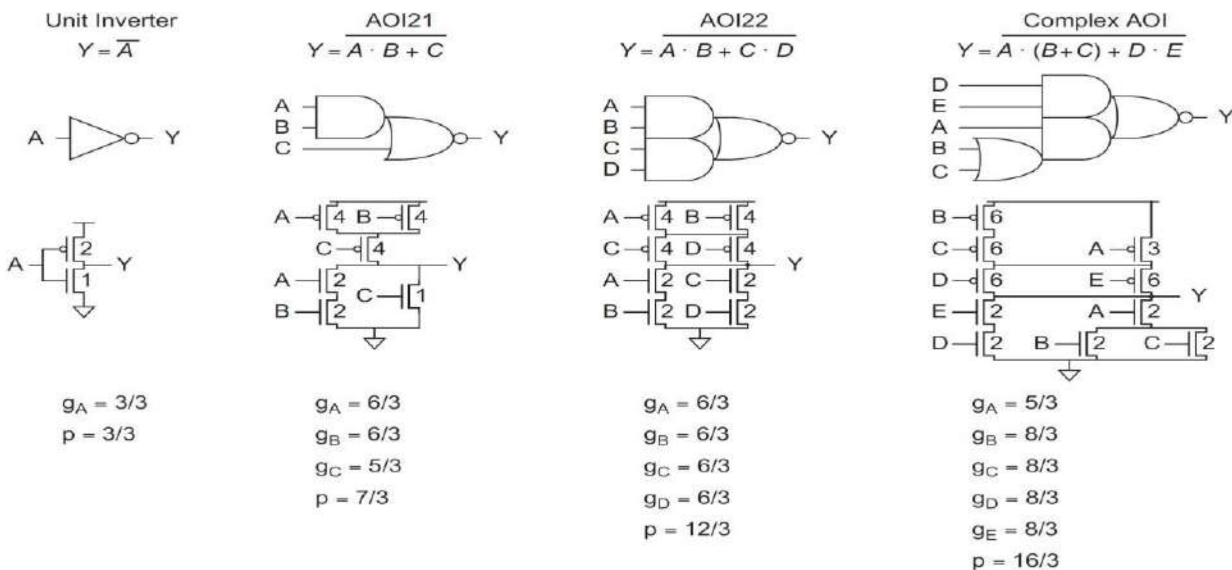


Figure: Logical efforts and parasitic delays of AOI gates

c. Input ordering delay effect

- The logical effort and parasitic delay of different gate inputs are different.
- Consider the falling output transition occurring, when one input hold a stable 1 value and the other rises from 0 to 1.
- If input B rises last, node x will initially be at $V_{DD} – V_t = V_{DD}$, because it was pulled up through the nMOS transistor on input A.
- The Elmore delay is (R/2)(2C) + R(6C) =7RC=2.33 τ
- If input A raises last, node x will initially be at 0 V, because it was discharged through the nMOS transistor on input B.
- No charge must be delivered to node x, so the Elmore delay is simply R(6C) =6RC =2τ.

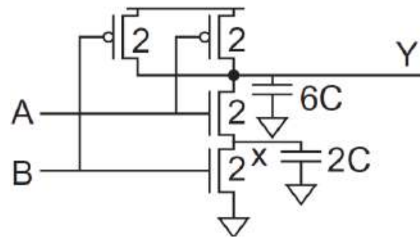


Figure: 2 –input NAND gate Schematic Y=A.B

- We define the outer input to be the input closer to the supply rail (e.g., B) and the inner input to be the input closer to the output (e.g., A).
- Therefore, if one signal is known to arrive later than the others, the gate is faster when that signal is connected to the inner input.

d. Asymmetric gates

- When one input is far less critical than another, even symmetric gates can be made asymmetric to favor the late input at the expense of the early one.
- In a series network, this involves connecting the early input to the outer transistor and making the transistor wider, so that, it offers less series resistance when the critical input arrives.
- In a parallel network, the early input is connected to a narrower transistor to reduce the parasitic capacitance.
- Consider the path in Figure (a). Under ordinary conditions, the path acts as a buffer between A and Y.
- When reset is asserted, the path forces the output low.
- If reset only occurs under exceptional circumstances and take place slowly, the circuit should be optimized for input-to-output delay at the expense of reset.
- This can be done with the asymmetric NAND gate in Figure (b).

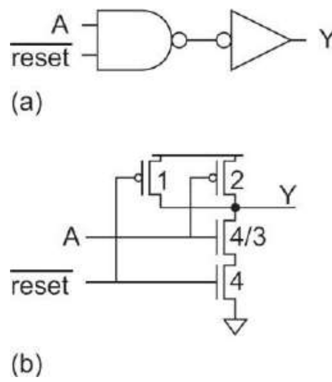


Figure: Resettable buffer optimized for data input

e. **Skewed gates**

What is meant by skewed gate and give functions of skewed gate with schematic diagrams?

- One input transition is more important than the other. HI-skew gates to favor the rising output transition. LO-skew gates to favor the falling output transition.
- This favoring can be done by decreasing the size of the noncritical transistor.
- The logical efforts for the rising (up) and falling (down) transitions are called g_u and g_d, respectively.
- Figure (a) shows, how a HI-skew inverter is constructed by downsizing the nMOS transistor.
- This maintains the same effective resistance for the critical transition, while reducing the input capacitance relative to the unskewed inverter of Figure (b).
- Thus reducing the logical effort on that critical transition to g_u = 2.5/3 =5/6.
- The logical effort for the falling transition is estimated by comparing the inverter to a smaller unskewed inverter with equal pulldown current, shown in Figure (c), giving a logical effort of g_d =2.5/1.5 =5/3.

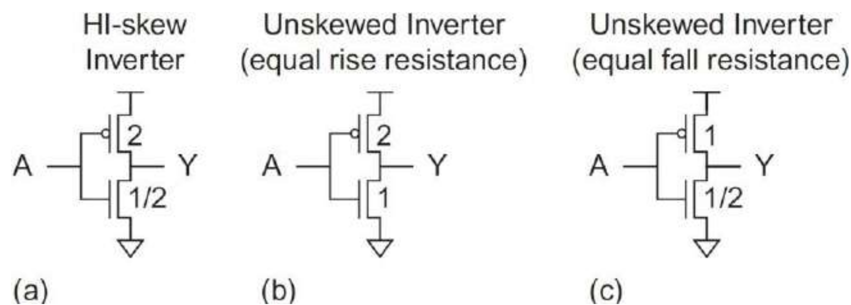


Figure: Logical effort calculation for HI-skew inverter

- Figure shows, HI skew and LO-skew gates with a skew factor of two. Skewed gates are sometimes denoted with an H or an L on their symbol in a schematic.

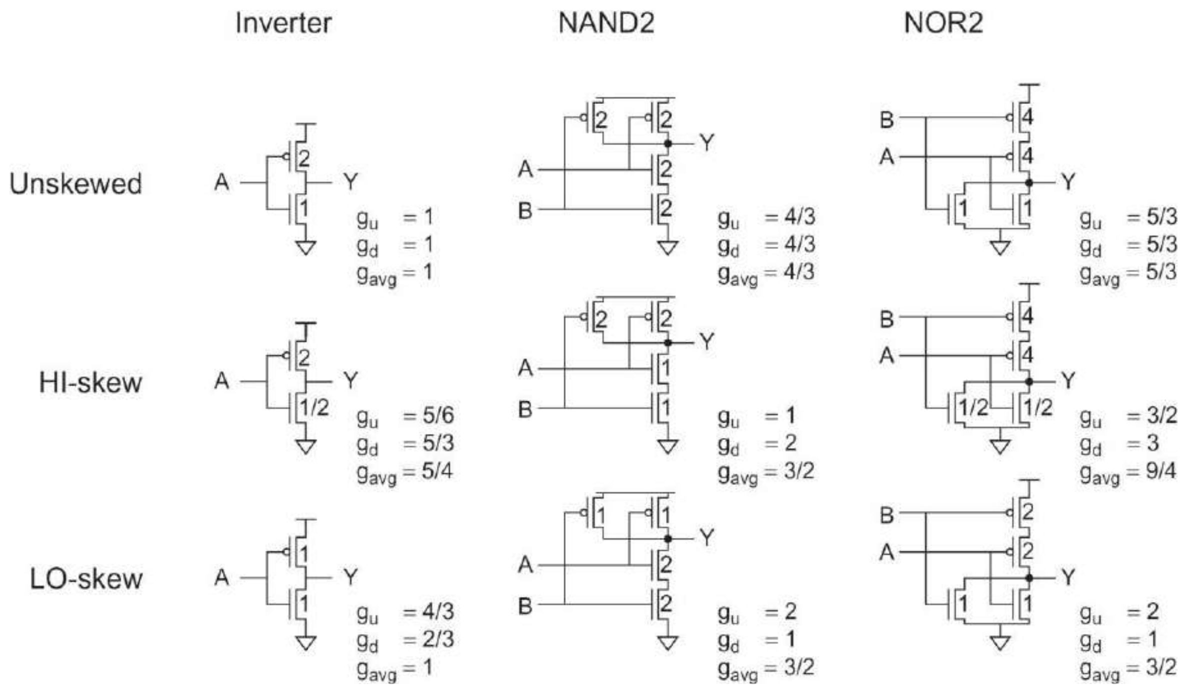


Figure: List of skewed gates

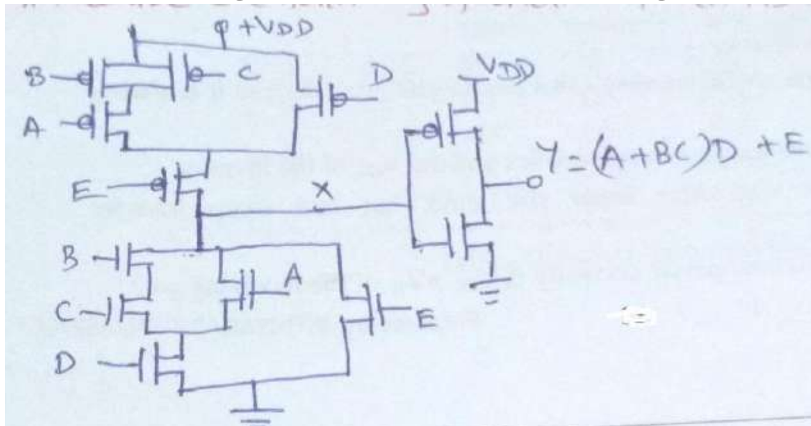
f. P/N ratios

- By accepting a slower rise delay, the pMOS transistors can be downsized to reduce input capacitance and average delay significantly.
- P/N ratio is defined as the ratio of pMOS to nMOS transistor width. For processes, a mobility ratio of μn/μp = 2.

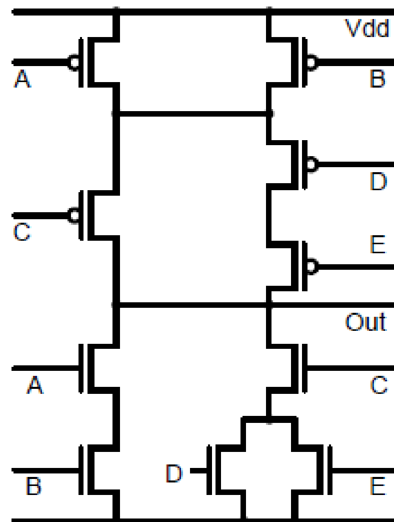
g. Multiple threshold voltages

- Some CMOS processes offer two or more threshold voltages.
- Transistors with lower threshold voltages produce more ON current, but also leak exponentially more OFF current.
- Libraries can provide both high- and low-threshold versions of gates.
- The low-threshold gates can be used carefully to reduce the delay of critical paths.
- Skewed gates can use low-threshold devices on, only the critical network of transistors.

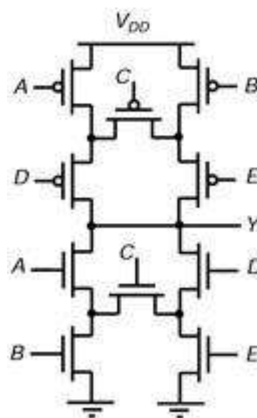
❖ **Realize the following function Y=(A+BC)D+E using static CMOS logic. (April 2019-6M)**



Example: Realize the following function Y= [AB+C (D+E)]' using static CMOS logic. [May 2021 (model)]



Example: Implement the following expression in static CMOS logic fashion using no more than 10 transistors. Y = (AB + ACE + DE + DCB)' [Nov 2019]



****

2.10.2 : Ratioed Circuits:

Write short notes on ratioed circuits. (Nov 2016)

- The ratioed gate consists of an nMOS pulldown network and pullup device called the static load.
- When the pulldown network is OFF, the static load pulls the output to 1.
- When the pulldown network turns ON, it fights the static load.
- The static load must be weak enough that, the output pulls down to an acceptable 0. Hence, there is a ratio constraint between the static load and pulldown network.

Advantage: Stronger static loads produce faster rising outputs.

Disadvantages:

- Degrade the noise margin and burn more static power when the output is 0.

  - A resistor is a simple static load, but large resistors consume a large layout area in typical MOS processes.
- Another technique is to use an nMOS transistor with the gate tied to V_{GG} (Shown in fig.(b)). If $V_{GG} = V_{DD}$, the nMOS transistor will only pull up to $V_{DD} - V_t$.
- Figure (c) shows depletion load ratioed circuit.

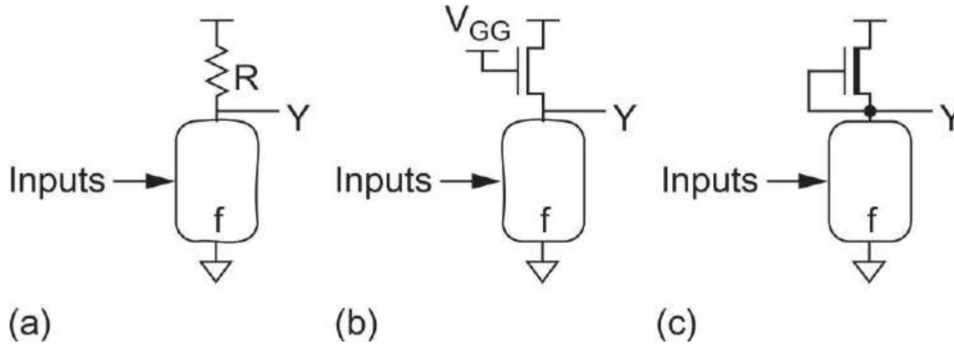


Figure: nMOS ratioed gates

2.10.3 : pseudo nMOS

Explain the detail about pseudo-nMOS gates with neat circuit diagram. (April/May 2011) (Nov/Dec 2013)

- Figure (a) shows a pseudo-nMOS inverter.
- The static load is built from a single pMOS transistor that has its gate grounded, so it is always ON.
- The beta ratio affects the shape of the transfer characteristics and the V_{OL} of the inverter.
- Larger relative pMOS transistor size offer faster rise time, but less sharp transfer characteristics.
- **Drawback:** Pseudo-nMOS gates will not operate correctly if $V_{OL} > V_{IL}$ of the receiving gate.

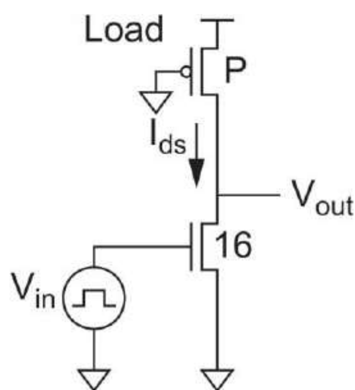


Figure (a): pseudo nMOS inverter

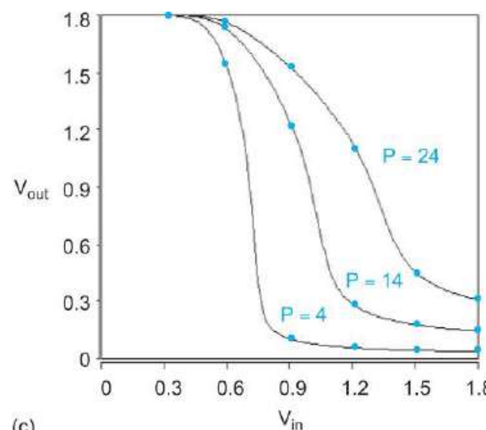


Figure: D.C Charateristics

- Figure shows several pseudo-nMOS logic gates.

Implement NAND gate using pseudo- nMOS logic. (Nov 2013, May 2021[model])

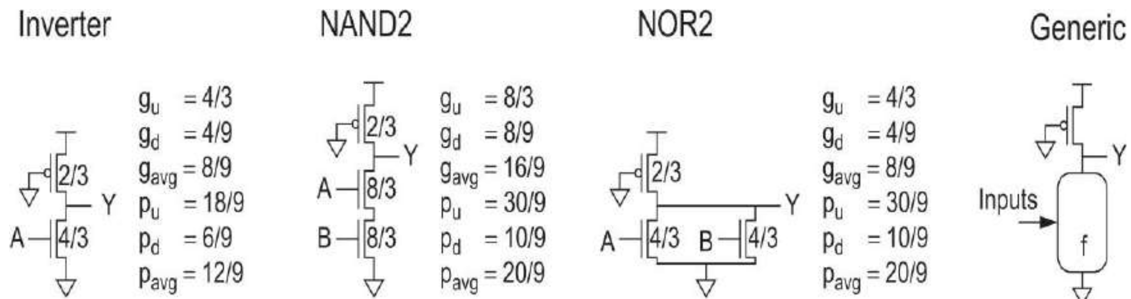


Figure: Pseudo-nMOS logic gates

2.10.4 : Ganged capacitor:

- Figure shows pairs of CMOS inverters ganged together.
- The truth table is given in Table, showing that the pair compute the NOR function. Such a circuit is sometimes called a symmetric ^{2} NOR, or ganged CMOS.

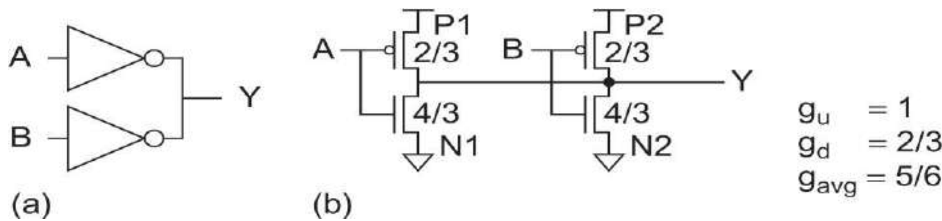


Figure: symmetric ^{2} NOR gate.

A	B	N1	P1	N2	P2	Y
0	0	OFF	ON	OFF	ON	1
0	1	OFF	ON	ON	OFF	~ 0
1	0	ON	OFF	OFF	ON	~ 0
1	1	ON	OFF	ON	OFF	0

Table: Operation of symmetric NOR

- When one input is 0 and the other 1, the gate can be viewed as a pseudo-nMOS circuit with appropriate ratio constraints.
- When both inputs are 0, both pMOS transistors turn on in parallel, pulling the output high faster than they would, in an ordinary pseudo nMOS gate.
- When both inputs are 1, both pMOS transistors turn OFF, saving static power dissipation.

2.10.5 : Differential Cascode voltage switch with pass gate logic (DCVSPG)

Explain about DCVSL logic with suitable example. (May 2017)

- Cascode Voltage Switch Logic (CVSL) seeks the benefits of ratioed circuits without the static power consumption.
- It uses both true and complementary input signals and computes both true and complementary outputs using a pair of nMOS pulldown networks, as shown in Figure (a).

- The pulldown network f implements the logic function as in a static CMOS gate, while $\overline{f}$ uses inverted inputs feeding transistors arranged in the conduction complement.
- For any given input pattern, one of the pulldown networks will be ON and the other OFF.
- The pulldown network that is ON will pull that output low.
- This low output turns ON the pMOS transistor to pull the opposite output high.
- When the opposite output rises, the other pMOS transistor turns OFF, so no static power dissipation occurs.
- Figure (b) shows a CVSL AND/NAND gate.

Advantage:

- CVSL has a potential speed advantage because all of the logic is performed with nMOS transistors, thus reducing the input capacitance.

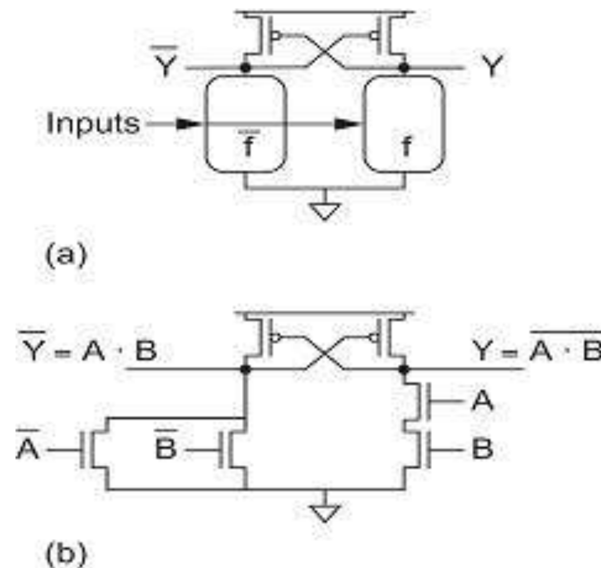


Figure: CVSL gates

****

2.11:Dynamic CMOS design:

Describe the basic principle of operation of dynamic CMOS, domino and NP domino logic with neat diagrams. (NOV 2011) [April / May 2023]

Dynamic Circuits:

- Ratioed circuits reduce the input capacitance by replacing the pMOS transistors connected to the inputs with a single resistive pullup.
- The drawbacks of ratioed circuits include
  - Slow rising transitions,
  - Contention on the falling transitions,
  - Static power dissipation and a nonzero V_{OL}.
- Dynamic circuits avoid these drawbacks by using a clocked pullup transistor rather than a pMOS that is always ON.
- Figure compares (a) static CMOS, (b) pseudo-nMOS, and (c) dynamic inverters.

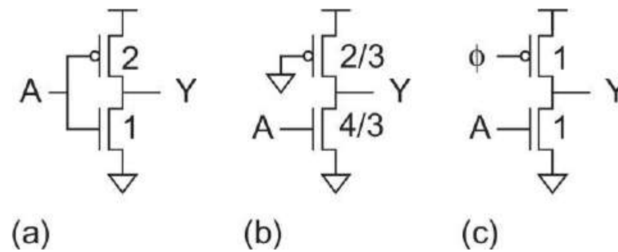


Figure: Comparison of (a) static CMOS, (b) pseudo-nMOS, and (c) dynamic inverters

- Dynamic circuit operation is divided into two modes, as shown in Figure.
  (i) During precharge, the clock ϕ is 0, so the clocked pMOS is ON and initializes the output Y high.
  (ii) During evaluation, the clock is 1 and the clocked pMOS turns OFF. The output may remain high or may be discharged low through the pulldown network.

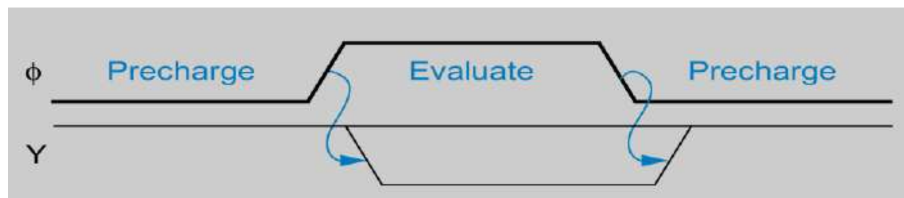


Figure: Precharge and evaluation of dynamic gates

<u>Advantages:</u>

- Dynamic circuits are the fastest used circuit family because they have lower input capacitance and no contention during switching.
- Zero static power dissipation.

<u>Disadvantags:</u>

- They require careful clocking, consume significant dynamic power and are sensitive to noise during evaluation mode.

Foot transistor:

- In Figure (c), if the input A is 1 during precharge, contention will take place because both the pMOS and nMOS transistors will be ON.
- When the input cannot be guaranteed to be 0 during precharge, an extra clocked evaluation transistor can be added to the bottom of the nMOS stack.
- To avoid contention as shown in the below figure, extra transistor is sometimes called as foot is added.

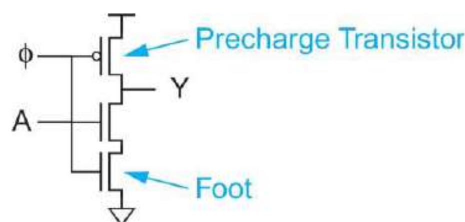


Figure: Footed dynamic inverter

- The given below figure shows generic footed and unfooted gates.

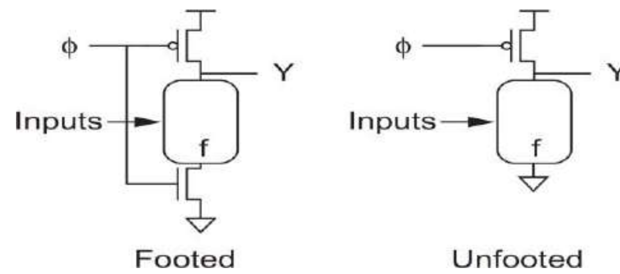


Figure: Generalized footed and unfooted dynamic gates

- The given below figure estimates the falling logical effort of both footed and unfooted dynamic gates.

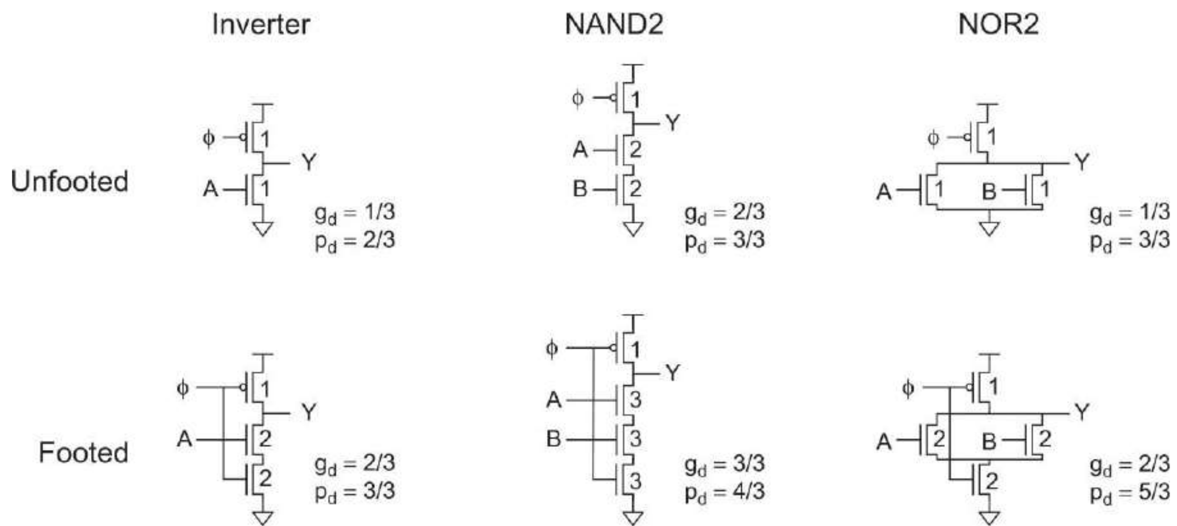


Figure: List of dynamic gates

- The pull down transistor's width is chosen to give unit resistance. Precharge occurs while the gate is idle and takes place more slowly.
- Therefore, the precharge transistor width is chosen for twice unit resistance.
- This reduces the capacitive load on the clock and the parasitic capacitance at the expense of greater rising delays.
- Footed gates have higher logical effort than their unfooted concept but are still an improvement over static logic.
- The parasitic delay does increase with the number of inputs, because there is more diffusion capacitance on the output node.
- A fundamental difficulty with dynamic circuits is the monotonicity requirement. While a dynamic gate is in evaluation, the inputs must be monotonically rising.
- That is, the input can start LOW and remain LOW, start LOW and rise HIGH, start HIGH and remain HIGH, but not start HIGH and fall LOW.
- Figure shows waveforms for a footed dynamic inverter in which the input violates monotonicity.

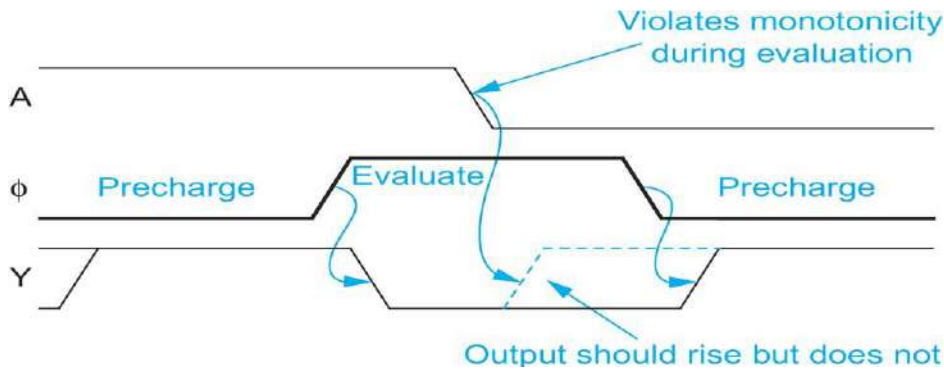


Figure: Monotonicity problem

- During precharge, the output is pulled HIGH.
- When the clock rises, the input is HIGH, so the output is discharged LOW through the pulldown network.
- The input later falls LOW, turning off the pulldown network. However, the precharge transistor is also OFF, so the output floats, staying LOW rather than rising.
- The output will remain low until the next precharge step.
- The inputs must be monotonically rising for the dynamic gate to compute the correct function.
- Unfortunately, the output of a dynamic gate begins HIGH and monotonically falls LOW during evaluation.
- This monotonically falling output X is not a suitable input to a second dynamic gate expecting monotonically rising signals, as shown in the below figure.
- Dynamic gates sharing the same clock cannot be directly connected.
- This problem is often overcome with domino logic.

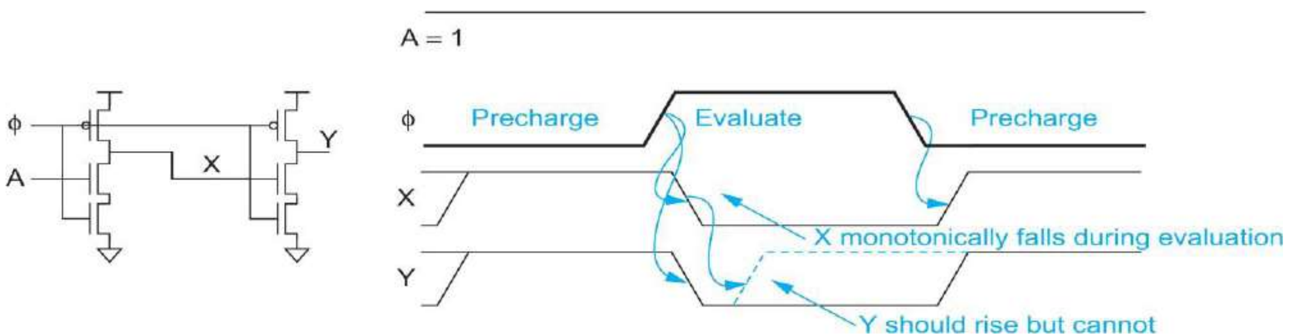


Figure: Incorrect connection of dynamic gates

The **charge sharing problem** occurs when the charge which is stored at the output node in the pre-charge phase is shared among the junction capacitance of transistors in the evaluation phase. Charge sharing may degrade the output voltage level or even cause an erroneous output value.

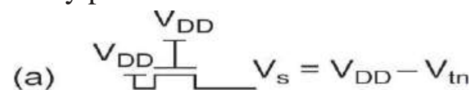
To **overcome the dynamic charge sharing** and soft- node leakage problems in NORA CMOS structures, a circuit technique called Zipper CMOS can be used. The basic circuit architecture of Zipper CMOS is essentially identical to NORA CMOS, with the exception of the clock signals.

****

2.12: Pass Transistor Logic:

Explain Pass transistor logic with neat sketches. (April 2008)
Explain the pass transistor logic and show how complementary pass transistor logic and double pass transistor logic are applied for 2: 1 multiplexer. [May 2021][Apr/May 2022]

- In pass-transistor circuits, inputs are applied to the source/drain diffusion terminals.
- These circuits build switches using either nMOS pass transistors or parallel pairs of nMOS and pMOS transistors called as transmission gates.
- The nMOS transistors pass '0's well but 1's poorly. Figure (a) shows an nMOS transistor with the gate and drain tied to V_{DD}.
- Initially at $V_S = 0$. $V_{gs} > V_{tn}$, so the transistor is ON and current flows.
- Therefore, nMOS transistors attempting to pass a 1 never pull the source above $V_{DD} – V_{tn}$. This loss is called a threshold drop.
- The pMOS transistors pass 1's well but 0's poorly.
- If the pMOS source drops below $|V_{tp}|$, the transistor cuts off.
- Hence, pMOS transistors only pull down to a threshold above GND, as shown in Figure (b).



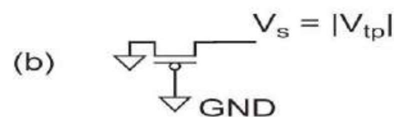
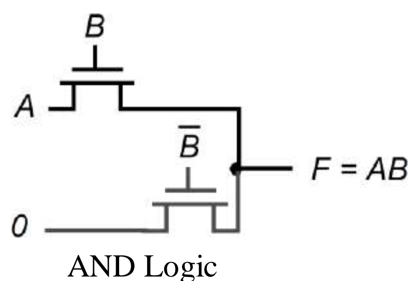
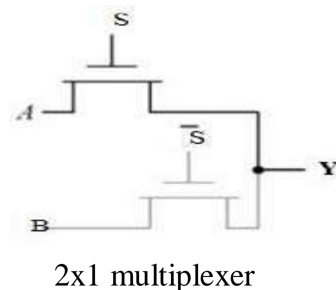


Figure : Pass Transistor threshold drops

- Figures show an implementation of the AND function and 2x1 multiplexer using only NMOS transistors.



AND Logic



2x1 multiplexer

- In AND gate, if the *B* input is high, the top transistor is turned ON and copies the input *A* to the output *F*.
- When *B* is low, the bottom pass transistor is turned ON and passes a 0.
- In 2x1 multiplexer, if the S selection input is high, the top transistor is turned ON and allows input *A* to the output Y.
- When *S* is low, the bottom pass transistor is turned ON and passes the B input.
- An NMOS device is effective at passing a 0 but is poor at pulling a node to V_{DD}. When the pass transistor pulls a node high, the output only charges up to $V_{DD} - V_{tn}$.

Application:

- Pass transistors are essential to the design of efficient 6-transistor static RAM cells used in modern systems.

Formal Method for P-T Logic Derivation

Complementary function can be implemented from the same circuit structure by applying complementary principle:

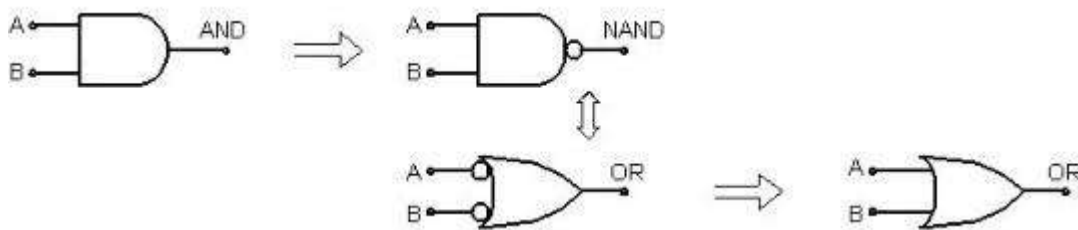
Complementary Principle: Using the same circuit topology, with pass signals inverted, complementary logic function is constructed in CPL.

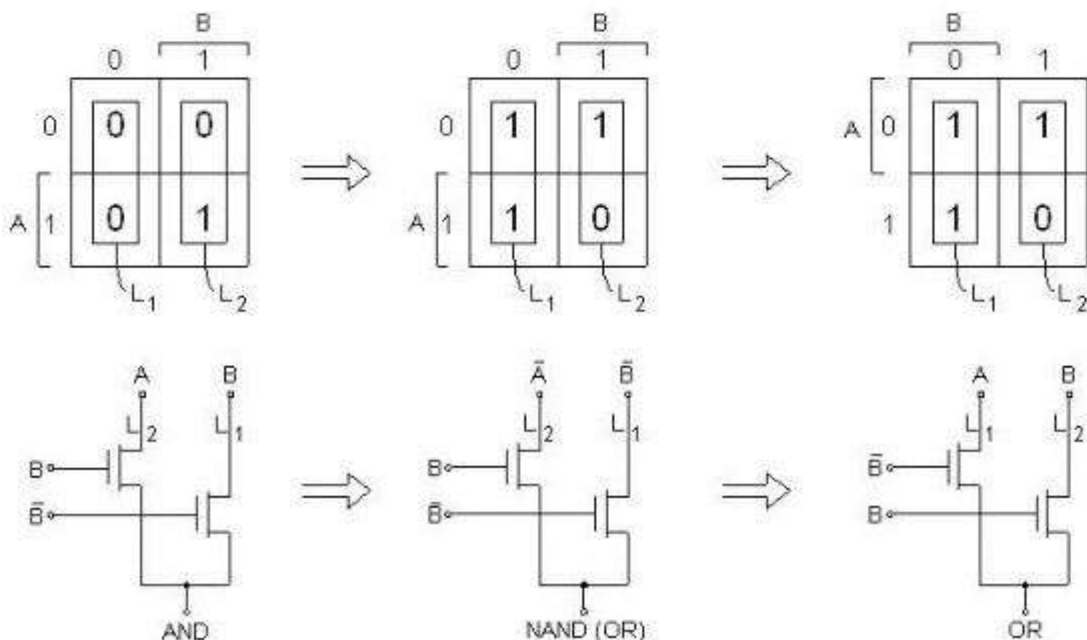
By applying duality principle, a dual function is synthesized:

Duality Principle: Using the same circuit topology, with gate signals inverted, dual logic function is constructed.

Following pairs of basic functions are dual:

- AND-OR (and vice-versa)
- NAND-NOR (and vice-versa)
- XOR and XNOR are self-dual (dual to itself)

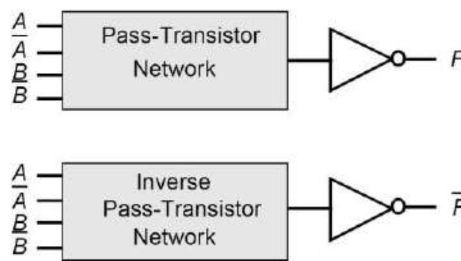




Complementary: AND ⟹ NAND
Duality: AND ⟹ OR

2.12.1 : Differential Pass Transistor Logic / Complementary Pass Transistor Logic (CPL)

- For high performance design, a differential pass-transistor logic family, called CPL, is commonly used.
- The basic idea is to accept true and complementary inputs and produce true and complementary outputs.
- A number of CPL gates (AND/NAND, OR/NOR, and XOR/NXOR) are shown in Figure.
- Since the circuits are *differential*, complementary data inputs and outputs are always available.
- Both polarities of every signal eliminate the need for extra inverters, as is often the case in static CMOS or pseudo-NMOS.
- CPL belongs to the class of *static* gates, because the output-defining nodes are always connected to either V_{DD} or *GND* through a low resistance path.
- This is advantage for the noise flexibility.



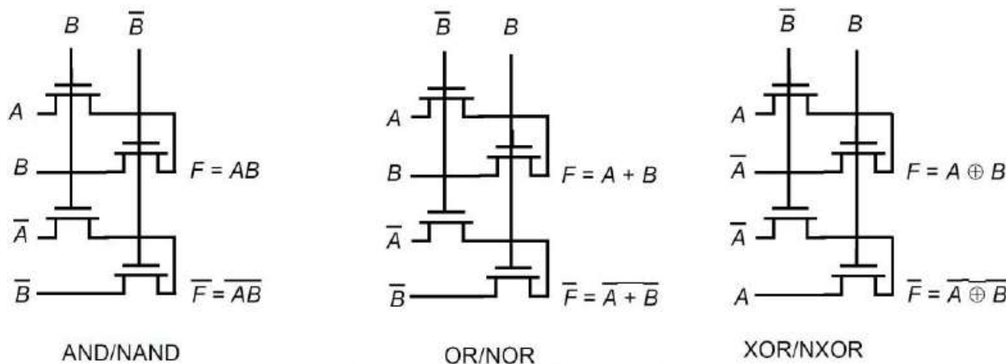


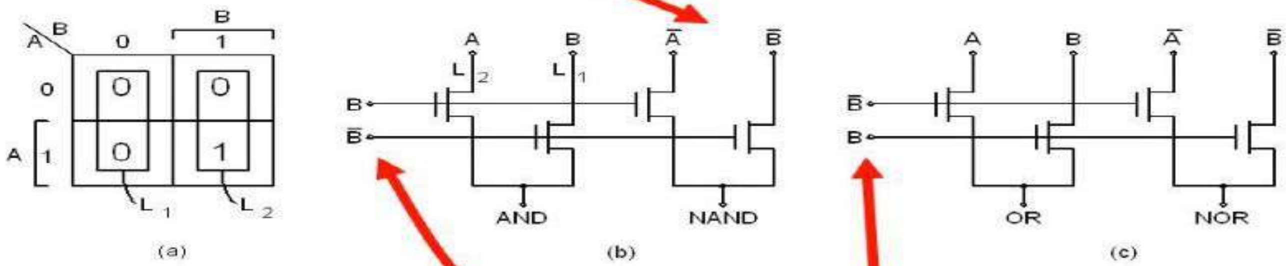
Figure: Complementary pass-transistor logic (CPL).

Formal Method for CPL Logic Derivation (AND, NAND, OR, NOR)

(a) Cover the Karnaugh-map with largest possible cubes (overlapping allowed)

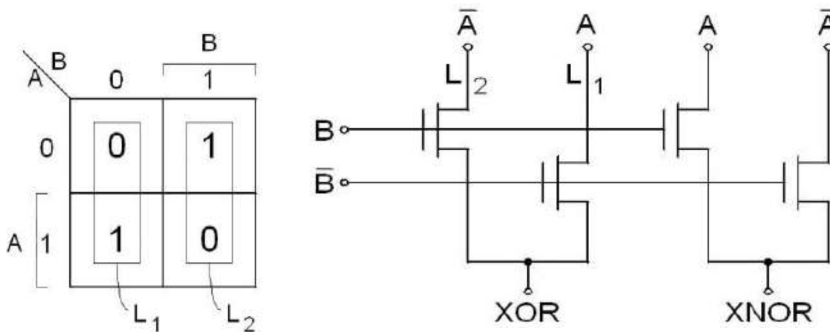
(b) Express the value of the function in each cube in terms of input signals

(c) Assign one branch of transistor(s) to each of the cubes and connect all the branches to one

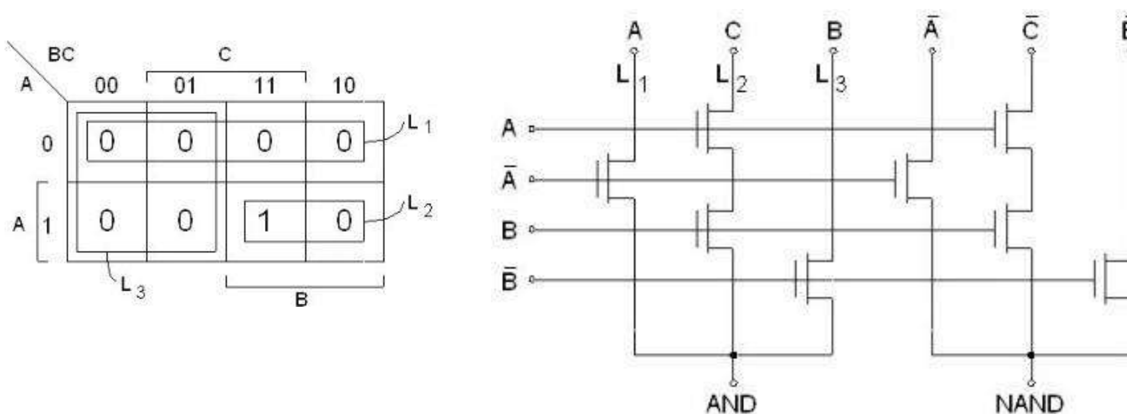


common node, which is the output of NMOS pass-transistor network

Example: Realize XOR and XNOR gate using CPL. [Nov 2019]



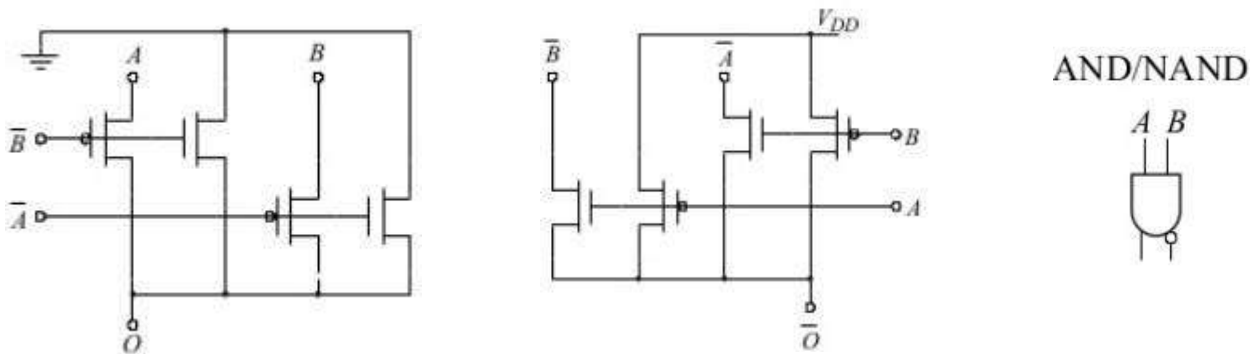
Example: Realize 3 – input AND and NAND gate using CPL.



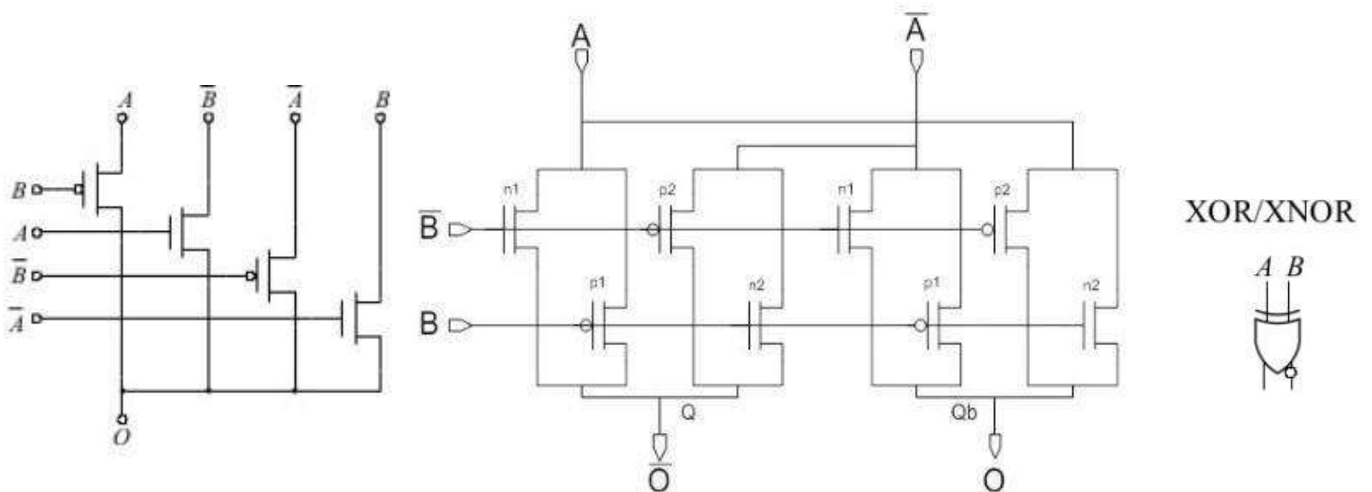
2.12.2 Double Pass Transistor Logic (DPL)

- Double Pass Transistor Logic is a double rail form of CMOS transmission gate optimized to use single pass transistors where only a known 0 or 1 needs to be passed.
- It passes good high and low logic levels without the need for level restoring devices.
- DPL uses both nMOS and pMOS switches to realize the desired functions.

- This provides a full swing on the output.
- No extra transistors are required for swing restoration.
- A DPL gate consists of both true and complementary inputs / outputs and hence is a *dual rail logic circuit.*



Example: Realize XOR gate using Double Pass Transistor Logic (DPL).



Synthesis Rules

- Two NMOS branches cannot be overlapped covering logic 1s. Similarly, two PMOS branches cannot be overlapped covering logic 0s.
- Pass signals are expressed in terms of input signals or supply. Every input vector has to be covered with exactly two branches.
- At any time, excluding transitions, exactly two transistor branches are active (any of the pairs NMOS/PMOS, NMOS/NMOS and PMOS/PMOS are possible), i.e. they both provide output current.

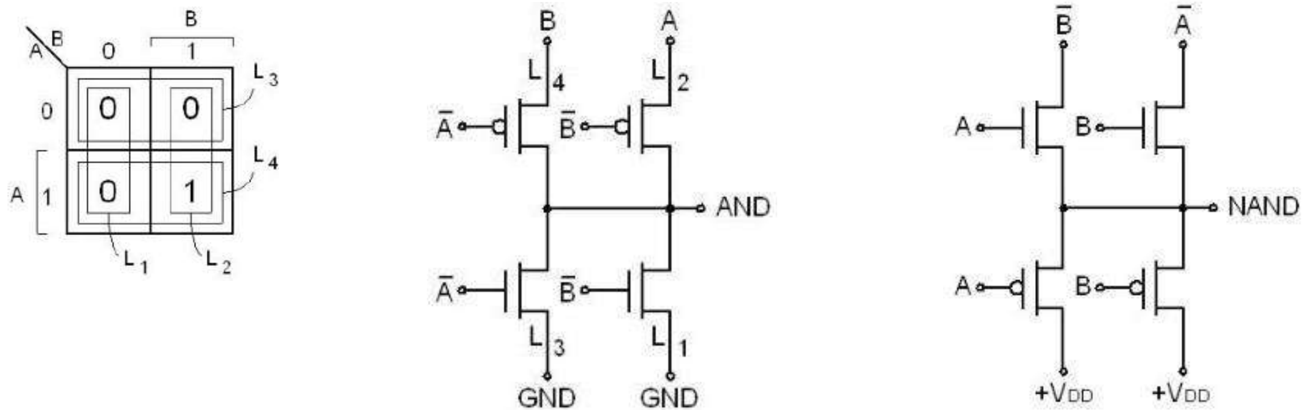
Complementary Principle: Complementary logic function in DPL is generated after the following modifications:

- Exchange PMOS and NMOS devices. Invert all pass and gate signals

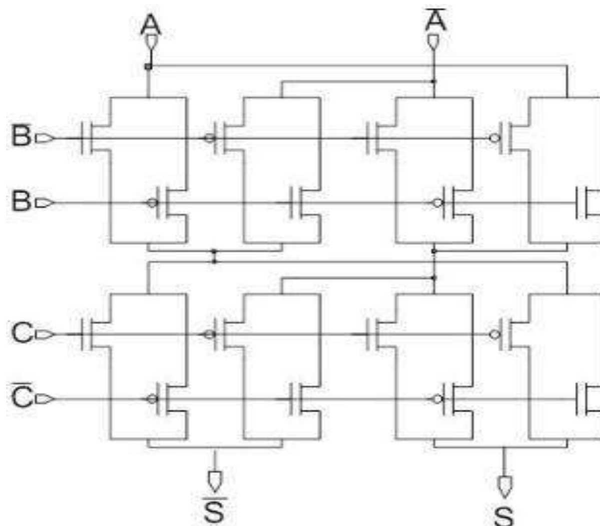
Duality Principle: Dual logic function in DPL is generated when:

- PMOS and NMOS devices are exchanged, and VDD and GND signals are exchanged.

Example: Realize AND and NAND gate using DPL.



Example: Realize full adder (sum circuit) using Double Pass Transistor Logic (DPL).



<u>*2.13 : CMOS with transmission gates*</u>

- ❖ **Discuss in detail the characteristics of CMOS Transmission gates.(May 2016, May 2017, Nov 2017)**
- ❖ **Explain Transmission gates with neat sketches. (April 2008, April 2018)**
- ❖ **List out limitations of pass transistor logic. Explain any two techniques used to overcome limitations. (NOV 2018)**
  - A transmission gate in conjunction with simple static CMOS logic is called CMOS with transmission gate.
  - A transmission gate is parallel pairs of nMOS and pMOS transistor.
  - A single nMOS or pMOS pass transistor suffers from a threshold drop.
  - Transmission gates solve the threshold drop but require two transistors in parallel.
  - The resistance of a unit-sized transmission gate can be estimated as R for the purpose of delay estimation.

- Current flow the parallel combination of the nMOS and pMOS transistors. One of the transistors is passing the value well and the other is passing it poorly.
- A logic-1 is passed well through the pMOS but poorly through the nMOS.
- Estimate the effective resistance of a unit transistor passing a value in its poor direction as twice the usual value: 2R for nMOS and 4R for pMOS.

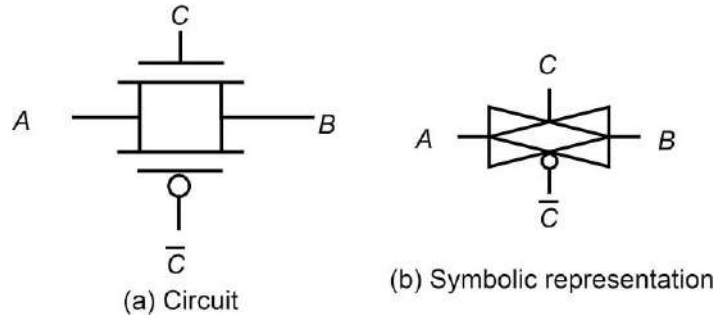


Figure: CMOS Transmission gate

- The given below figure shows the parallel combination of resistances. When passing a 0, the resistance is R || 4R = (4/5)R.
- The effective resistance passing a 1 is 2R || 2R = R.
- Hence, a transmission gate made from unit transistors is approximately R in either direction.
- Transmission gates are built using equal-sized nMOS and pMOS transistors.
- Boosting the size of the pMOS transistor only slightly improves the effective resistance while significantly increasing the capacitance.

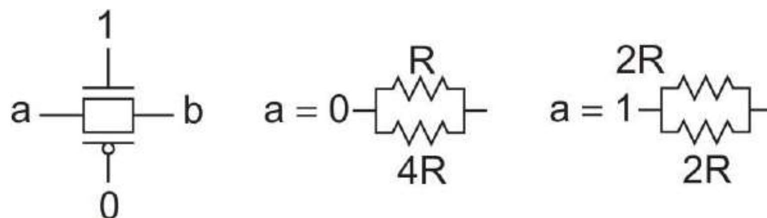


Figure: Effective resistance of a unit transmission gate

- Figure (a) redraws the multiplexer to include the Inverters that drive the diffusion inputs but to exclude the output inverter. Figure (b) shows this multiplexer drawn at the transistor level.

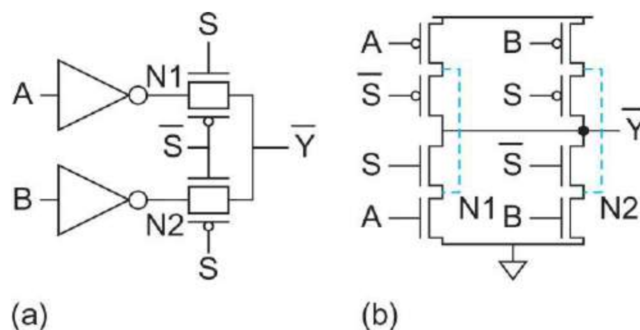


Figure: CMOSTG in a 2-input inverting multiplexer

****

2.14 : Domino logic

Explain the domino logic families with neat diagrams. (NOV 2012, APRIL-2015, Nov 2017)

- The dynamic-static pair together is called a domino gate.
- The monotonicity problem can be solved by placing a static CMOS inverter between dynamic gates, as shown in figure (a).
- This converts the monotonically falling output into a monotonically rising signal suitable for the next gate, as shown in figure (b).
- A single clock can be used to precharge and evaluate all the logic gates within the chain.
- The dynamic output is monotonically falling during evaluation, so the static inverter output is monotonically rising.
- Therefore, the static inverter is usually a HI-skew gate to favor this rising output. Observe that precharge occurs in parallel, but evaluation occurs sequentially.

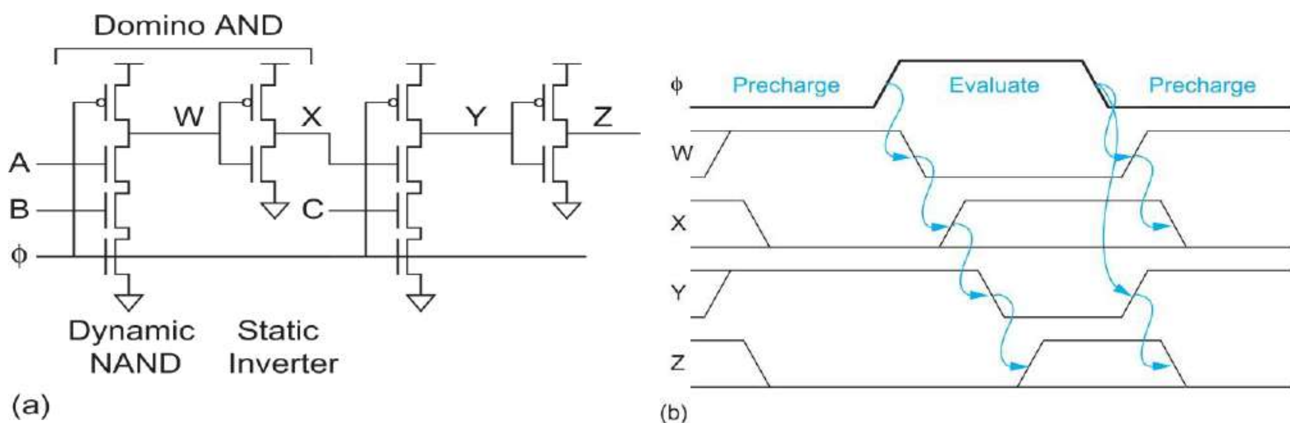


Figure: Domino gates

2.15 : Dual Rail Domino Logic:

- Dual-rail domino gates encode each signal with a pair of wires. The input and output signal pairs are denoted with _h and _l, respectively.
- Table summarizes the encoding. The _h wire is asserted to indicate that the output of the gate is "high" or 1. The _l wire is asserted to indicate that the output of the gate is "low" or 0.
- When the gate is precharged, neither _h nor _l is asserted. The pair of lines should never be both asserted simultaneously during correct operation.

sig_h	sig_l	Meaning
0	0	Precharged
0	1	'0'
1	0	'1'
1	1	Invalid

Table: Dual-rail domino signal encoding

- Dual-rail domino gates accept both true and complementary inputs and compute both true and complementary outputs, as shown in Figure (a).
- This is identical to static CVSL circuits except that the cross-coupled pMOS transistors are instead connected to the precharge clock.
- Therefore, dual-rail domino can be viewed as a dynamic form of CVSL, sometimes called DCVS.
- Figure (b) shows a dual-rail AND/NAND gate and Figure (c) shows a dual-rail XOR/XNOR gate. The gates are shown with clocked evaluation transistors, but can also be unfooted.

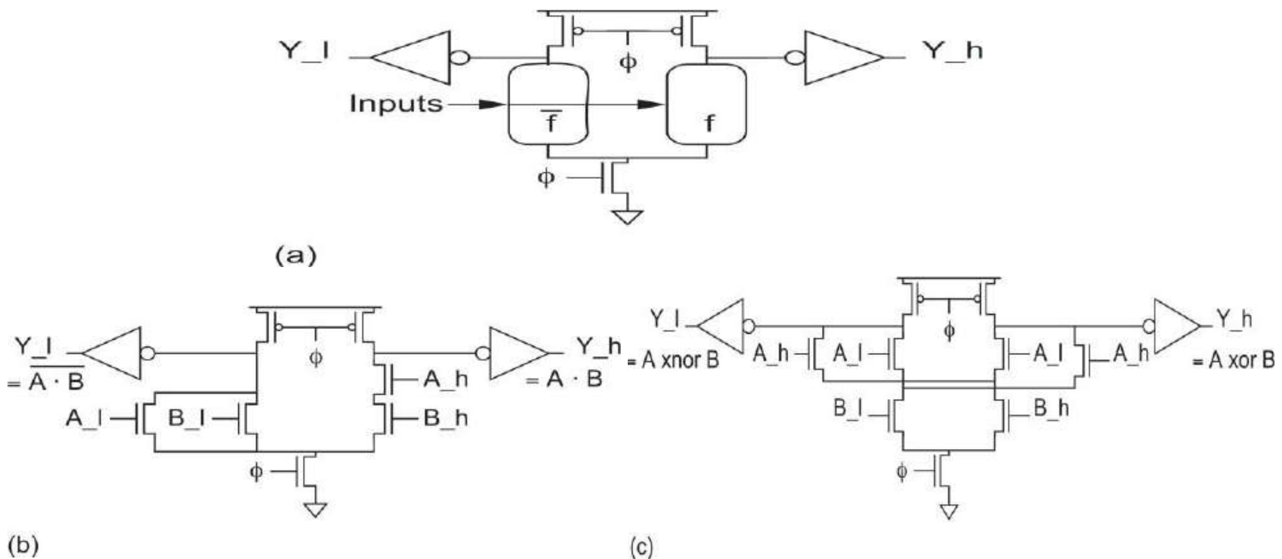


Figure: Dual-rail domino gates

Disadvantages:

- It requires more area, wiring and power.
- Dual-rail structures lose the efficiency of wide dynamic NOR gates.

Application:

- It is useful for asynchronous circuits.

2.16 : Keepers

Explain the keeper logic family with neat diagrams.
Briefly discuss the signal integrity issues in dynamic design. (April 2018, NOV 2018)

- Dynamic circuits have poor input noise margins.
- If the input rises above V_t,, while the gate is in evaluation, the input transistors will turn ON weakly and can incorrectly discharge the output.
- Both leakage and noise margin problems can be addressed by adding a keeper circuit.
- Figure shows a conventional keeper on a domino buffer. The keeper is a weak transistor that holds, or staticizes, the output at the correct level when it would otherwise float.
- When the dynamic node X is high, the output Y is low and the keeper is ON to prevent X from floating.
- When X falls, the keeper initially opposes the transition, so it must be much weaker than the pulldown network.

- Eventually Y rises, turning the keeper OFF and avoiding static power dissipation.

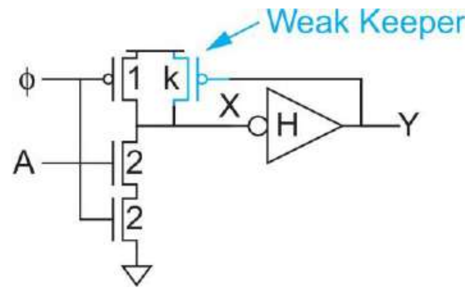


Figure: Conventional keeper

- The keeper must be strong enough to compensate for any leakage current drawn when the output is floating and the pulldown stack is OFF.
- Strong keepers also improve the noise margin, because when the inputs are slightly above V_t, the keeper can supply enough current to hold the output high.

2.16.1 : Differential keeper:

- Figure shows a differential keeper for a dual-rail domino buffer.
- When the gate is precharged, both keeper transistors are OFF and the dynamic outputs float. As one of the rails evaluates low, the opposite keeper turns ON.
- The differential keeper is fast, because it does not oppose the falling rail.
- As long as one of the rails is guaranteed to fall promptly, the keeper on the other rail will turn on before excessive leakage or noise causes failure.

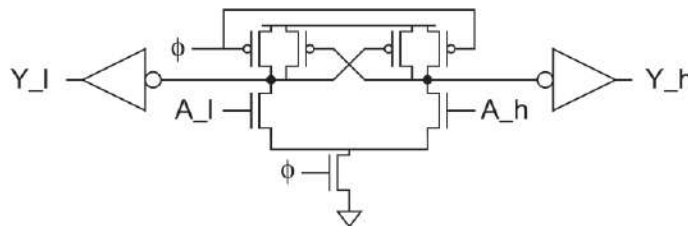


Figure: Differential keeper

2.16.2 : Secondary precharge devices

- Dynamic gates are subject to problems with charge sharing.
- For example, consider the 2-input dynamic NAND gate in Figure (a). Suppose the output Y is precharged to V_{DD} and inputs A and B are low.

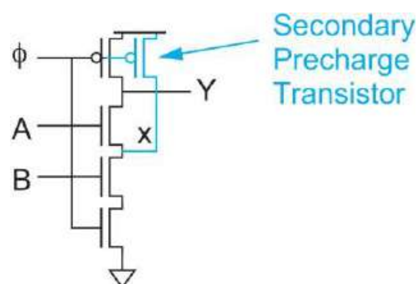


Figure: Secondary precharge transistor

- Also suppose that the intermediate node x had a low value from a previous cycle.
- During evaluation, input A rises, but input B remains low, so the output Y should remain high.
- However, charge is shared between C_X and C_Y, shown in Figure (b). This behaves as a capacitive voltage divider and the voltages equalize at

$$V_x = V_Y = \frac{C_Y}{C_x + C_Y} V_{DD}$$

2.16.3 : Charge sharing:

- Charge sharing is serious when the output is lightly loaded (small C_Y) and the internal capacitance is large.
- If the charge-sharing noise is small, the keeper will eventually restore the dynamic output to V_{DD}.
- If the charge-sharing noise is large, the output may flip and turn off the keeper, leading to incorrect results.
- Charge sharing can be overcome by precharging some or all of the internal nodes with secondary precharge transistors.
- These transistors should be small, because they only charge the small internal capacitances and their diffusion capacitance slows the evaluation.
- It is sufficient to precharge every other node in a tall stack.

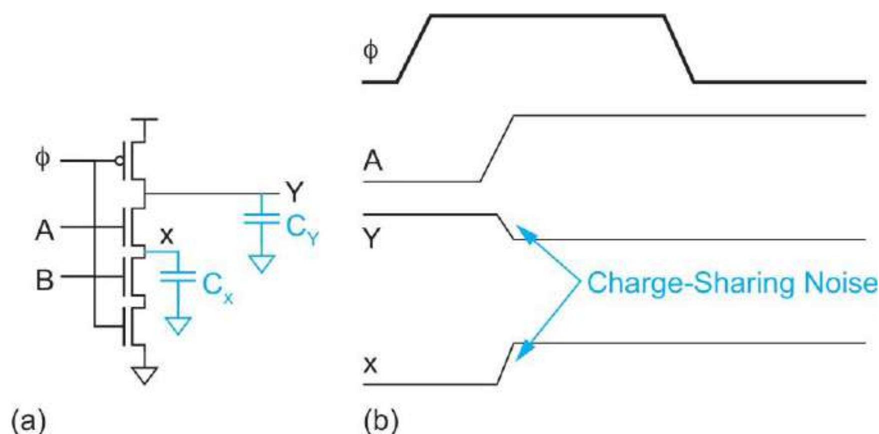


Figure: Charge-sharing noise

2.16.4 : NP and Zipper Domino

Describe the basic principle of operation of NP domino logic. (NOV 2011)

- The HI-skew inverting static gates are replaced with predischarged dynamic gates using pMOS logic.
- A footed dynamic p-logic NAND gate is shown in Figure (b). When ϕ is 0, the first and third stages precharge high while the second stage predischarges low.
- When ϕ rises, all the stages evaluate. Domino connections are possible, as shown in Figure (c).
- The design style is called NP Domino or NORA Domino (NO RAce).

- NORA has two major drawbacks.
  (i) The logical effort of footed p-logic gates is worse than that of HI-skew gates.

(ii) NORA is extremely susceptible to noise.

- In an ordinary dynamic gate, the input has a low noise margin (about V_t), but is strongly driven by a static CMOS gate.
- The floating dynamic output is more prone to noise from coupling and charge sharing, but drives another static CMOS gate with a larger noise margin.
- In NORA, however, the sensitive dynamic inputs are driven by noise prone dynamic outputs.
- Besides drawback and the extra clock phase requirement, there is little reason to use NORA.
- Zipper domino is a closely related technique, that leaves the precharge transistors slightly ON during evaluation by using precharge clocks. This swing between 0 and $V_{DD} - |V_{tp}|$ for the pMOS precharge and V_{tn} and V_{DD} for the nMOS precharge.

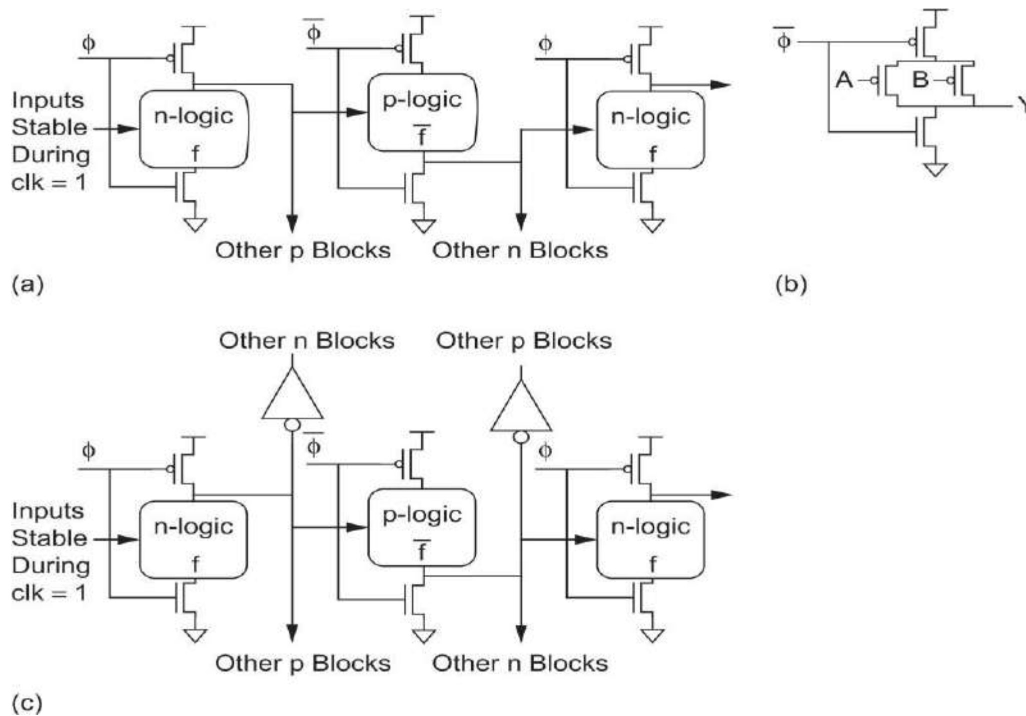


Figure : NP Domino

****

2.17 : Power dissipation:

- **Explain the static and dynamic power dissipation in CMOS circuits with necessary diagrams and expressions. (DEC 2011, Nov 2015, NOV 2016, May 2017, May 2010)**
- **What are the sources of power dissipation in CMOS and discuss various design techniques to reduce power dissipation in CMOS? (Nov 2012, May 2013, Nov 2014, May 2016)**
- **Derive an expression for dynamic power dissipation. (April 2019, Nov 2019, May 2021)[April / May 2023]**

- The instantaneous power P (t) consumed by a circuit element is the product of the current and the voltage of the element

$$P(t) = I(t)V(t)$$

- The energy consumed over time interval T is the integral of the instantaneous power

$$E = \int_0^T P(t)\,\mathrm{dt}$$

- The average power is $P_{avg} = \frac{E}{T} = \frac{1}{T}\int_0^T P(t)\,\mathrm{dt}$

  Power is expressed in units of Watts (W). Energy is usually expressed in Joules (J)

  - By Ohm's Law, V = IR, so the instantaneous power dissipated in the resistor is

$$P_R(t) = \frac{V_R^2(t)}{R} = I_R^2(t)\,\mathrm{R}$$

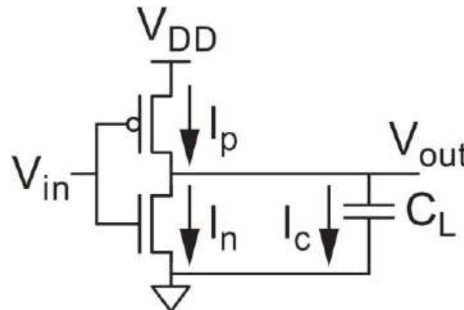
- This power is converted from electricity to heat. V_{DD} supplies power proportional to its current

  $P_{VDD}(t) = I_{DD}(t)\, V_{DD}$

- When the capacitor is charged from 0 to V_C, it stores energy E_C

$$E_C = \int_0^\infty I(t)V(t)dt = \int_0^\infty C\frac{dV}{dt}V(t)dt = C\int_0^{V_C} V(t)dV = \tfrac{1}{2}CV_C^2$$

- Figure shows a CMOS inverter driving a load capacitance.



- When the input switches from 1 to 0, the pMOS transistor turns ON and charges the load to V_{DD}.
- According to E_C equation the energy stored in the capacitor is

$$E_C = \tfrac{1}{2}C_L V_{DD}^2$$

- The energy delivered from the power supply is

$$E_C = \int_0^\infty I(t)V_{DD}dt = \int_0^\infty C\frac{dV}{dt}V_{DD}dt = CV_{DD}\int_0^{V_{DD}} dV = CV_{DD}^2$$

- Gate switches at some average frequency f_{sw}.
- Over some interval T, the load will be charged and discharged Tf_{sw} times.
- Then, the average power dissipation is

$$P_{\text{switching}} = \frac{E}{T} = \frac{Tf_{\text{sw}}CV_{DD}^2}{T} = CV_{DD}^2 f_{\text{sw}}$$

- This is called the dynamic power because it arises from the switching of the load.
- Because most gates do not switch every clock cycle, it is often more convenient to express switching frequency f_{sw} as an activity factor α times the clock frequency f.
- The dynamic power dissipation may be rewritten as

$$P_{\text{switching}} = \alpha C V_{DD}^2 f$$

- The activity factor is the probability that the circuit node transitions from 0 to 1, because that is the only time the circuit consumes power.
- A clock has an activity factor of $\alpha = 1$ because it rises and falls every cycle.
- The total power of a circuit is calculated as,

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

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

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

2.17.1 : Dynamic power:

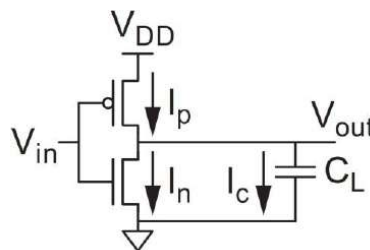
- Dynamic power consists mostly of the switching power.

$$P_{\text{switching}} = \alpha C V_{DD}^2 f$$

- The supply voltage V_{DD} and frequency f are known by the designer.
- To estimate dynamic power, one can consider each node of the circuit.
- The capacitance of the node is the sum of the gate, diffusion, and wire capacitances on the node.
- The activity factor can be estimated using switching probability or measured from logic simulations.
- The effective capacitance of the node is, its true capacitance multiplied by the activity factor.
- The switching power depends on the sum of the effective capacitances of all the nodes.

2.17.1.1 :Sources of dynamic power dissipation:

- Dynamic dissipation due to
  - Charging and discharging load capacitances as gates switchs.
  - “Short-circuit” current while both pMOS and nMOS stacks are partially ON



2.17.1.2 : Low Power Design Principles / Reducing dynamic power dissipation:

❖ **Explain various ways to minimize the static and dynamic power dissipation. (Nov 2013, May 2015)**
❖ **Discuss the low power design principles in detail. (Nov 2017)**

- Low power design involves considering and reducing each of the terms in switching power.
  i. As V_{DD} is a quadratic term, it is good to select the minimum V_{DD}.
  ii. Choose the lowest frequency.
  iii. The activity factor is reduced by putting unused blocks to sleep.
  iv. Finally, the circuit may be optimized to reduce the overall load capacitance.

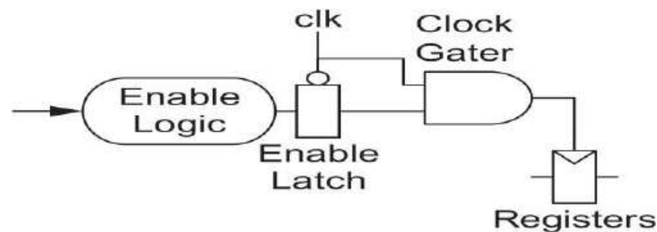
- Switching power is consumed by delivering energy to charge a load capacitance, then dumping this energy to GND.

Activity factor:

- If a circuit can be turned OFF entirely, the activity factor and dynamic power go to zero.
- Blocks are typically turned OFF, by stopping the clock called as clock gating.
- The activity factor of a logic gate can be estimated by calculating the switching probability.

(a) Clock gating:

- Clock gating, AND's a clock signal with an enable to turn OFF the clock to idle blocks.
- The clock enable must be stable, while the clock is active.
- Figure shows how an enable latch can be used to ensure the enable does not change before the clock falls.



Capacitance:

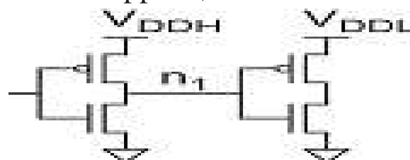
- Switching capacitance comes from the wires and transistors in a circuit.
- Wire capacitance is minimized through good floor planning and placement.
- Device-switching capacitance is reduced by choosing smaller transistors.

Voltage:

- Voltage has a quadratic effect on dynamic power.
- Therefore, choosing a lower power supply significantly reduces power consumption.
- The chip may be divided into multiple voltage domains, where each domain is optimized for the needs of certain circuits.

a. Voltage domains:

- Selecting, which circuits belong in which domain and routing power supplies to multiple domains.
- Figure **(Voltage domain crossing)** shows direct connection of inverters in two domains using high and low supplies, V_{DDH} and V_{DDL}, respectively.

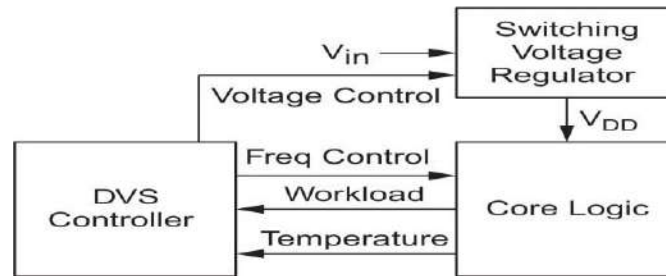


b. Dynamic voltage scaling (DVS):

Q: Describe how dynamic voltage scaling can reduce dynamic power dissipation. (NOV 2021)

- Systems can save large amounts of energy by reducing the clock frequency, then reducing the supply voltage.
- This is called dynamic voltage scaling (DVS) or dynamic voltage/frequency scaling (DVFS).

- It determines the supply voltage and clock frequency sufficient to complete the workload on schedule or to maximize performance without overheating.



- Figure shows a block diagram for a basic DVS system.

Frequency:

- Dynamic power is directly proportional to frequency, so a chip should not run faster than necessary.
- Reducing the frequency allows downsizing transistors or using a lower supply voltage.

Low Power Architecture

- Device Level
  - ✓ Low Capacitance in device and Multi Threshold Devices
- DVFS – Dynamic Voltage Frequency Scaling
- Multi VDD
- Gate Sizing
- Voltage Islands
- Power Gating
- Clock Gating
- Parallelism and Pipelined micro-architecture

Parallel Computations

• Multiple cores
• Multiple Issue pipelines
• Linear power increase

Pipelining

• Faster clock
• Exponential power increase
• Longer branch miss-predictions

2.17.2 : Static power:

- Static power is consumed even when a chip is not switching.
- Static CMOS gates have no contention current.

2.17.2.1 :Sources of static power dissipation:

- Static dissipation due to
  - Subthreshold leakage through OFF transistors.
  - Gate leakage through gate dielectric.
  - Junction leakage from source/drain diffusions.
  - Contention current in ratioed circuits.

$$P_{static} = (I_{sub} + I_{gate} + I_{junc} + I_{contention})V_{DD}$$

1. **Subthreshold leakage current:**
   - Subthreshold leakage current flows when a transistor is OFF.
   - Subthreshold leakage current equation is

$$I_{sub} = I_{off} 10^{\frac{V_{gs} + \eta(V_{ds} - V_{DD}) - k_{\gamma} V_{sb}}{S}}$$

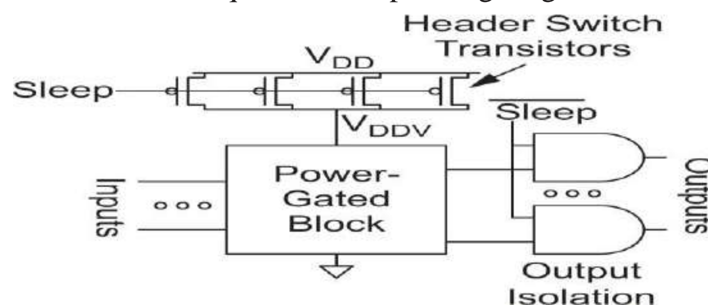
where I_{off} is the subthreshold current at $V_{gs} = 0$ and $V_{ds} = V_{DD}$, and S is the subthreshold slope.

2. **Gate leakage:**
   - Gate leakage occurs when carriers tunnel through a thin gate dielectric, when a voltage is applied across the gate (e.g., when the gate is ON).
   - Gate leakage is a strong function of the dielectric thickness.
3. **Junction leakage:**
   - Junction leakage occurs when a source or drain diffusion region is at a different potential from the substrate.
   - Leakage of reverse-biased diodes is usually negligible.
4. **Contention current:**
   - Static CMOS circuits have no contention current. However, certain alternative circuits inherently draw current even while quiescent.

2.17.2.2 :Methods of reducing static power:

Power gating:

- To reduce static current during sleep mode is, to turn OFF the power supply to the sleeping blocks. This technique is called power gating.



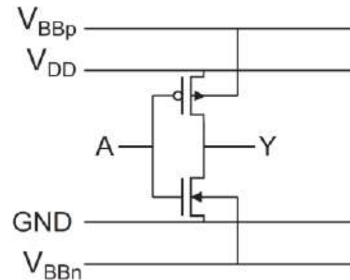
- The logic block receives its power from a virtual V_{DD} rail, V_{DDV}.
- When the block is active, the header switch transistors are ON, connecting V_{DDV} to V_{DD}.
- When the block goes to sleep, the header switch turns OFF, allowing V_{DDV} to float and gradually sink toward 0.

Multiple threshold voltage and oxide thickness:

- Selective application of multiple threshold voltages can maintain performance on critical paths with low-V_t transistors, while reducing leakage on other paths with high-V_t transistors.

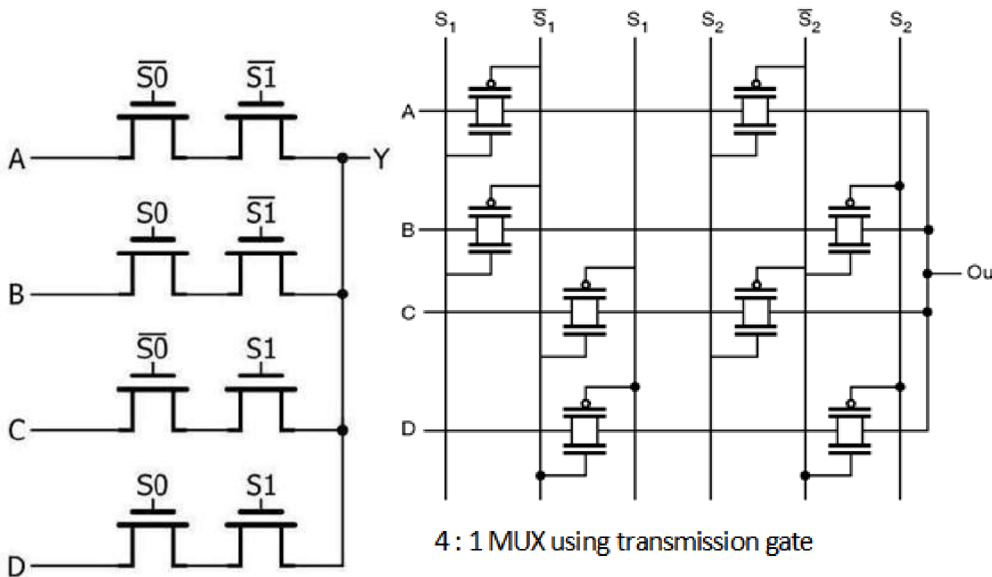
Variable threshold voltage:

- Method to achieve high I_{on} in active mode and low I_{off} in sleep mode is, by adjusting the threshold voltage of the transistor by applying a body bias.
- This technique is sometimes called variable threshold CMOS (VTCMOS).
- Figure shows a schematic of an inverter using body bias.



****

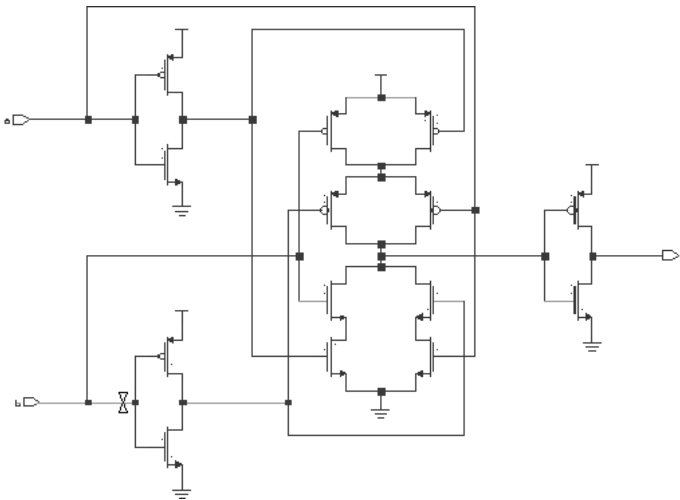
❖ **Let A, B, C and D be the inputs of a data selector and S_0 & S_1 be the select lines. Realize a 4:1 data selector using nMOS pass transistor and transmission gate approach. Compare the hardware complexity. (April 2019-13M)**



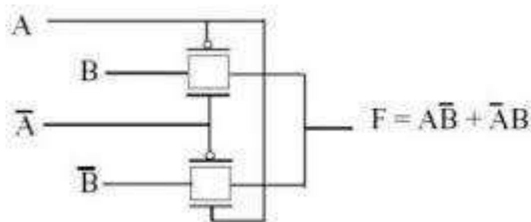
4:1 MUX using pass transistor

Need double of transistors to design 4:1 MUX using transmission gate compare with pass transistor.

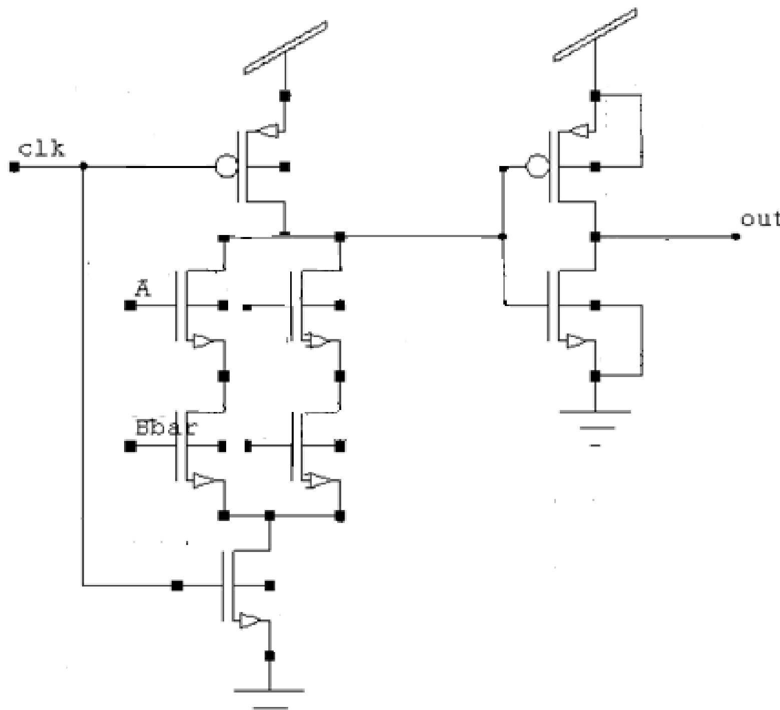
❖ **Realize a 2-input XOR using static CMOS, transmission gate and dynamic CMOS logic. Analyze the hardware complexity. (April 2019-15M)**
Draw a static CMOS XOR gate. [Nov 2019]



2-input XOR using static CMOS



2-input XOR using transmission gate



2-input XOR using dynamic CMOS logic

****

2.18 Circuit Pitfalls

- Circuit designers use simple circuits due to their robustness.
- Larger circuits with more transistors add more area & capacitance and has more possibility of error.

- Static CMOS is the most robust family.

Certain circuit pitfalls that can cause chips to fail include

- ✓ Threshold drop
- ✓ Leakage
- ✓ Charge sharing
- ✓ Power supply noise
- ✓ Hot spots
- ✓ Minority carrier injection
- ✓ Back gate coupling
- ✓ Process sensitivity
- ✓ Soft errors

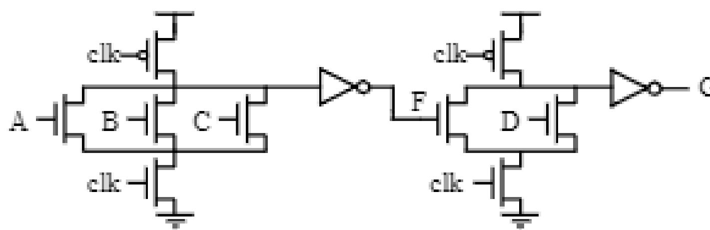
****

Example:

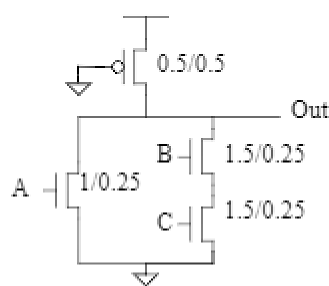
Suppose we wish to implement the two logic functions given by F=A+B+C andG=A+B+C+D. Assume both true and complementary signals are available. Implement these functions in dynamic CMOS as cascaded stages so as to minimize the total transistor count. [Nov 2019]

Solution:

Dynamic gates with NMOS pull-down networks cannot be directly cascaded. This solution uses a domino logic approach.



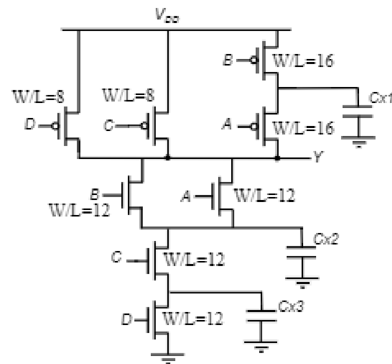
Q:What logic function does the circuit implement? To which logic family does the circuit belong? Does the circuit have any advantages over fully complementary CMOS? [Nov 2019]



Solution:

- ✓ The circuit implements Out = (A+BC)’. It is in the pseudo NMOS family.
- ✓ The circuit uses less area than a fully complementary CMOS implementation.

Q:Consider the circuit of Figure **[Nov 2019]**



What is the logic function implemented by the CMOS transistor network? Size the NMOS and PMOS devices so that the output resistance is the same as that of an inverter with an NMOS W/L = 4 and PMOS W/L = 8.

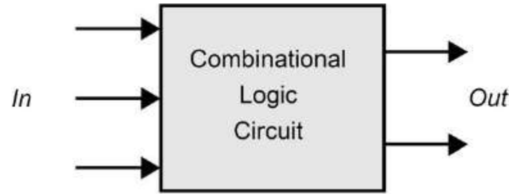
Solution:

The logic function is: Y = [(A+B) CD]’. The transistor sizes are given in the figure above.

TWO MARK QUESTIONS & ANSWERS

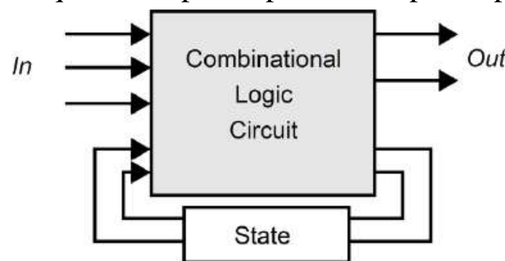
UNIT II - COMBINATIONAL LOGIC CIRCUITS

1. **Define combinational circuit and give an example.**
   A combinational circuit can be defined as a circuit, whose output is dependent only on the inputs. Example: full adder.



(a) Combinational

2. **Define sequential circuit and give an example.**
   A sequential circuit can be defined as a circuit, whose output depends not only on the present value of its inputs but on the sequence of past inputs. Example: flip-flop.



(b) Sequential

3. **What is the static CMOS inverter?**
   Static CMOS inverter circuit is the combination of nMOS pulldown and pMOS pullup network.

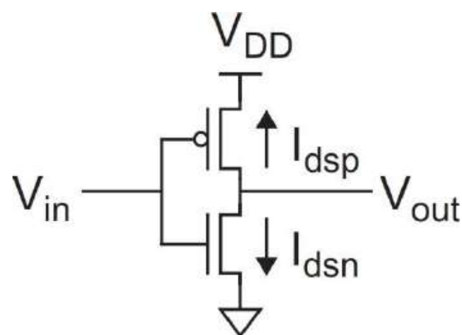


Figure: Static CMOS inverter

4. **What are the advantages of static CMOS circuits?**
   Advantages of static CMOS circuits:
   - Static CMOS circuits have good noise margins
   - Static CMOS circuits are fast, low power, easy to design.
   - Static CMOS circuits are widely supported by CAD tools,

- Static CMOS circuits are available in standard cell libraries.

5. **What are the disadvantages of static CMOS circuits?**
   Disadvantages of static CMOS circuits:
   - It requires both nMOS and pMOS transistor on each input.
   - It has large logical effort.
   - Gate delay is increased.

6. **What is bubble pushing? (May 2010)**
   DeMorgan's law :

$$\overline{A.B} = \overline{A} + \overline{B}$$
$$\overline{A+B} = \overline{A}.\overline{B}$$

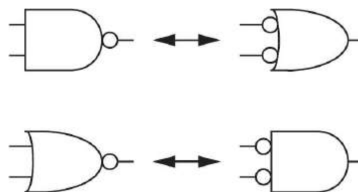


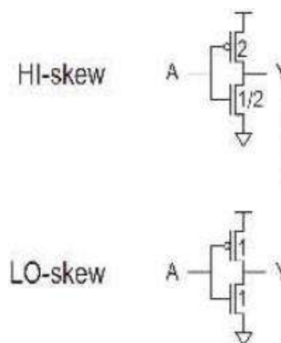
Figure: Bubble pushing with DeMorgan's law

- A NAND gate is equivalent to an OR of inverted inputs.
- A NOR gate is equivalent to an AND of inverted inputs.
- The same relationship applies to gates with more inputs.
- Switching between these representations is easy and is often called bubble pushing.

7. **What is meant by compound gate?**
   Static CMOS efficiently handles compound gates computing various inverting combinations of AND/OR functions in a single stage.

8. **What is the function of skewed gate?**
   One input transition is more important than the other. HI-skew gates to favor the rising output transition and LO-skew gates to favor the falling output transition.

9. **What are the types of skewed gate?**
   Two types of skewed gate are HI-skew gate and LO-skew gate.



10. Define P/N ratio.

P/N ratio is defined as the ratio of pMOS to nMOS transistor width. For processes, a mobility ratio of $\mu n/\mu p = 2$.

11. What is meant by ratioed logic?

In ratioed logic, a gate consists of an nMOS pull-down network that realizes the logic function and a simple load device, which replace the entire pull-up network.

12. What is meant by pseudo nMOS logic?

A pseudo nMOS logic (ratioed logic) which uses a grounded pMOS load is referred to as a pseudo nMOS gate.

13. Draw a pseudo nMOS inverter.(Nov 2011)

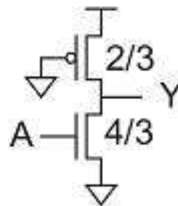


Figure: Pseudo nMOS inverter

14. What are the disadvantages of using a pseudo nMOS gate instead of a full CMOS gate?(May 2012)

What is the drawback of pseudo nMOS logic?

- Pseudo-nMOS gates will not operate correctly if (Maximum low level output) VOL >VIL (Maximum low level input) of the receiving gate.
- Ratioed circuits dissipate power continually in certain states and have poor noise margin.
- Ratioed circuits used in situations where smaller area is needed.

15. What are advantages and disadvantages of ratioed logic?

Advantage: Stronger static loads produce faster rising outputs.

Disadvantages:

a. Degrade the noise margin and burn more static power when the output is 0.
b. A resistor is a simple static load, but large resistors consume a large layout area in typical MOS processes.

16. Compare CMOS combinational logic gates with reference to the equivalent nMOS depletion load logic with reference to the area requirement.(May 2012)

For CMOS, the area required is $533\mu m^2$, for pseudo nMOS the area required is $288\ \mu m^2$

17. What is AOI logic function?

AND OR Invert logic function (AOI) implements operation in the order of AND, OR, NOT operations. So this logic function is known as AOI logic function.

18. What is AOI 221 Gate?

AOI 221, here 221 refers to number of inputs in each section.

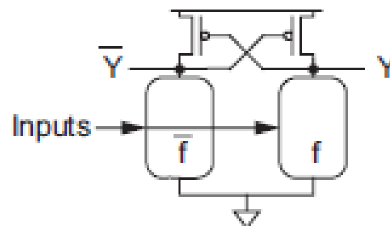
19. What is meant by Asymmetric Gates?

When one input is far less critical than another, even nominally symmetric gates can be made asymmetric to favor the late input at the expense of the early one.

20. What is meant by Cascode Voltage Switch Logic?

Cascode Voltage Switch Logic (CVSL) seeks the benefits of ratioed circuits without the static power consumption.

It uses both true and complementary input signals and computes both true and complementary outputs using a pair of nMOS pulldown networks.



21. What are the advantages of Cascode Voltage Switch Logic?
State the reasons for the speed advantages of CVSL family. (Nov 2012)

Advantage: CVSL has a potential speed advantage, because all of the logic is performed with nMOS transistors, thus reducing the input capacitance.

22. Define rise & fall time. [April 2008, Nov/Dec-2008] [Nov/Dec-2009]

Rise time (tr):

- It is defined as time for a waveform to rise from 20% to 80% of its steady state value.

Fall time (tf):

- It is defined as time for a waveform to fall from 80% to 20% of its steady-state value.

23. What is edge rate?

Edge rate is defined as an average value of rise time and fall time.

Edge rate (t_{rf}) = (t_r + t_f)/2 .

24. What do you mean by propagation delay time?

Propagation delay time (tpd) (or) Maximum delay is defined as maximum time from the input crossing 50% to the output crossing 50%.

25. What do you mean by contamination delay time?

Contamination delay time (tcd) (or) Minimum delay is defined as minimum time from the input crossing 50% to the output crossing 50%.

26. What is meant by average contamination delay time?

Average contamination delay time (**t_{cd}**) is defined as an average value of rising contamination delay time (**t_{cdr}**) and falling contamination delay time (**t_{cdf}**).

Contamination delay time (t_{cd}) = (t_{cdr} + t_{cdf})/2

27. What is meant by RC delay model?

RC delay model approximates the nonlinear transistor I-V and C-V characteristics with an average resistance and capacitance over the switching range of the gate.

28. Draw equivalent RC delay model for a MOS transistor?

Equivalent RC delay model for an nMOS and pMOS transistor:

nMOS transistor

pMOS transistor

29. Define electrical or fanout.

Electrical effort is defined as ratio of the output capacitance to input capacitance of a gate.

Electrical effort (h) = C_{out} / C_{in}

30. What is parasitic delay?

The parasitic delay (P) of a gate is the delay of the gate when it drives zero load. It can be estimated with RC delay models.

31. Write the general expression of parasitic delay for n inputs NAND and NOR gate?

Expression of parasitic delay for n inputs NAND and NOR is n. Where, n – no. of inputs.

32. Write the expression for the logical effort and parasitic delay of n input NOR gate. [Nov/Dec-2011]

Logical effort for n inputs NOR gate is (2n+1)/3

Parasitic delay for n inputs NOR gate is n

33. What is meant by dynamic logic?

- Dynamic logic using a clocked pullup transistor rather than a pMOS that is always ON.

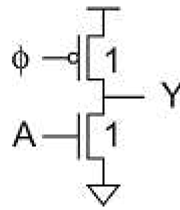


Figure: Dynamic logic

34. What are the two modes of operation in dynamic logic and give its functions? (NOV 2021)

Dynamic circuit operation has two modes, as shown in Figure.

(i) During precharge, the clock ϕ is 0, so the clocked pMOS is ON and output Y is high.

(ii) During evaluation, the clock is 1 and the clocked pMOS turns OFF. The output may remain high or may be discharged low through the pulldown network.

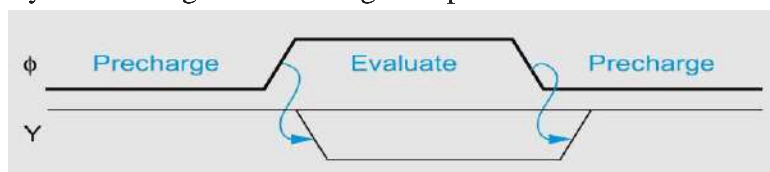


Figure: Precharge and evaluation of dynamic gates

35. What are the disadvantages of dynamic logic?

Disadvantages of dynamic logic:

- Dynamic circuits require careful clocking.
- Dynamic circuits consume significant dynamic power.
- Dynamic circuits are sensitive to noise during evaluation mode.
- Monotonicity problem
- Dynamic circuits suffer from charge leakage.

36. What are the advantages of dynamic logic?

Advantages of dynamic logic:

- Dynamic circuit has lower input capacitance and no contention during switching.
- Zero static power dissipation.

37. What is the use of footed transistor in dynamic logic circuit?

An extra clocked evaluation transistor can be added to the bottom of the nMOS stack to avoid contention as shown in the below figure. The extra transistor is called a foot.

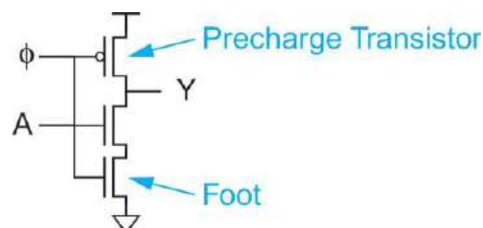


Figure: Footed dynamic inverter

38. What is meant by Monotonicity problem?

- During precharge, the output is pulled HIGH. When the clock rises, the input is HIGH, so the output is discharged LOW through the pulldown network. The input later falls LOW, turning OFF the pulldown network.
- However, the precharge transistor is also OFF, so the output floats, staying LOW rather than rising. This is called monotonicity problem in dynamic circuit.

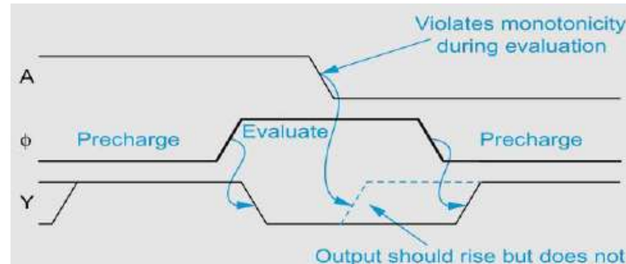


Figure: Monotonicity problem

39. What is meant by domino logic?

The monotonicity problem can be solved by placing a static CMOS inverter between dynamic gates. This converts the monotonically falling output into a monotonically rising signal. The dynamic-static pair together is called domino logic.

40. Write the features of CMOS Domino Logic?

Features of CMOS Domino Logic:

- These structures occupy small area.
- Parasitic capacitance is to be small to increase the speed.
- Each gate can make one 'logic 1' to 'logic 0' transition.

41. What is the use of keeper circuit?

The keeper is a weak transistor that holds or *staticizes* the output at the correct level when it floats.

42. What is meant by pass transistors?

In pass-transistor circuits, inputs are applied to the source/drain diffusion terminals.
A single nMOS or pMOS pass transistor suffers from a threshold drop.

43. Which MOS can pass logic 1 and logic 0 strongly?

p-MOS can pass strong logic 1.
n-MOS can pass strong logic 0.

44. What is meant by CMOS Transmission gate? (Nov 2007, May 2011)(or) Define Transmission gate. (May 2009)

A parallel pair of nMOS and pMOS transistors is called *transmission gate*.
Transmission gates solve the threshold drop problem but require two transistors in parallel.

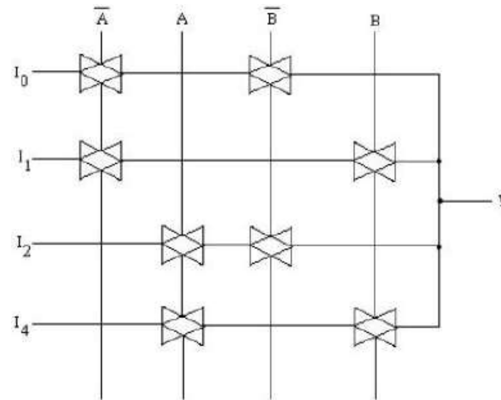
45. State the advantages of Transmission gate. (April 2017, May 2021)

Transmission gates solve the threshold drop problem.

It provides good conducting path between input and output.

46. Draw the CMOS implementation of 4-to-1 MUX using transmission gates.[Nov/Dec 2022]

CMOS implementation of 4-to-1 MUX using transmission gates:



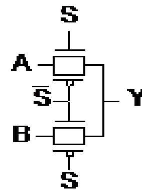
47. What are the various forms of inverter based CMOS logic?

Various forms of inverter based CMOS logic:

i. Pseudo nMOS logic
ii. Dynamic CMOS logic
iii. Clocked CMOS logic
iv. CMOS domino logic

48. Draw 2:1 MUX using transmission gate. (Nov 2008, APRIL-2015, 2016)[April/May 2023]

2:1 MUX using transmission gate:

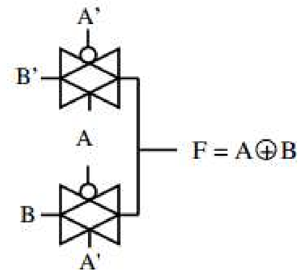


49. Draw XOR and XNOR using transmission gates. [Apr/may-2010]

- **2 I/P XOR using TGs:**

  F = A.B' + A'.B , we need this: if A=1 → F = B' (pass B' to F)
  if A=0 → F = B (pass B to F)

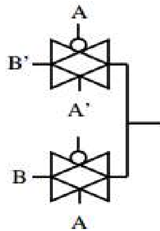
  using TGs:



8 Ts (2 inverters for A and B and two TGs)
Versus 12 Ts for regular CMOS

2 I/P XNOR

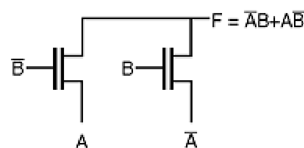
F = A.B + A'B'   if A=1 → pass B to F
if A=0 → pass B' to F



50. Draw a two input XOR using nMOS pass transistor logic. April 2019

A	B	F
0	0	0
0	1	1
1	0	1
1	1	0

Truth table of 'XOR' gate



'XOR' gate using pass transistor logic

51. Define power dissipation. [Nov/Dec-2013]

Power dissipation is defined as power consumed by the transistor unnecessarily, therefore increasing the power requirement to the logic.

52. List the types of power dissipation. [APRIL 2015, April 2018, Nov 2017]
List the various power losses in CMOS circuits. (May 2013)

Types of power dissipation are static and dynamic power dissipation.

53. What do you understand by static & dynamic power dissipation?
State the various types of power dissipation. (April 2019)

- Dynamic power dissipation is power consumed by transistor when it operates.

- At some stage both transistor pMOS and nMOS are in ON stage which, leads to short circuit formation between V_{DD} and GND, thus unwanted power dissipation occurs.
- Static power dissipation is power consumed by transistor when it is not in operating stage.

54. What do you mean by low power design?

When both static and dynamic powers are reduced then, the circuit is said to be low power designed circuit.

55. What are the factors that cause dynamic power dissipation in CMOS circuits? (Nov 2016, NOV 2021)

Dynamic dissipation due to

- Charging and discharging load capacitances as gates switch.
- “Short-circuit” current while both pMOS and nMOS stacks are partially ON.

56. How can dynamic power dissipation reduced? (or) State any two criteria for low power logic design. (Nov 2015, MAY 2014)

Dynamic power dissipation ($P_{dynamic}$) expressed as below,

$$P_{dynamic} = \alpha C V_{DD}^2 f$$

To reduce dynamic power, use the following

- α: clock gating, sleep mode
- C: small transistors (esp. on clock), short wires
- V_{DD}: lowest suitable voltage
- f: lowest suitable frequency

57. Write the expression for power dissipation in CMOS inverter. [Nov/Dec-2008]

Total power dissipation P_{total} is the sum of dynamic power dissipation ($P_{dynamic}$) and static power dissipation (P_{static}).

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

Where, $P_{dynamic} = \alpha C V_{DD}^2 f$

- α:activity factor
- C: capacitor
- V_{DD}: Supply voltage
- f: Supply frequency

$$P_{static} = (I_{sub} + I_{gate} + I_{junc} + I_{contention}) V_{DD}$$

58. What are the factors that cause static power dissipation in CMOS circuits? [Nov-2012] List the sources of static power consumption. (Nov 2016, NOV 2021)

Static dissipation due to

- Subthreshold leakage through OFF transistors
- Gate leakage through gate dielectric
- Junction leakage from source/drain diffusion

- Contention current in ratioed circuits

59. How can static power dissipation reduced?

To reduce static power

- Selectively use ratioed circuits
- Selectively use low V_t devices
- Leakage reduction: Use stacked devices, body bias and low temperature

60. Why single phase dynamic logic structure cannot be cascaded? Justify.(May 2016)

No, single phase dynamic logic structure cannot be cascaded. Because monotonicity problem will be raised, so static logic should be used in between dynamic logics structure.

61. What is Complementary Pass Transistor logic? (NOV/DEC-2014)

Complementary Pass Transistor logic has complementary data inputs and outputs.

It reduces the count of transistors used to make different logic gates, by eliminating redundant transistors.

62. Give the effects of supply voltage and temperature variations CMOS circuits. [Nov-2012]

Supply Voltage:

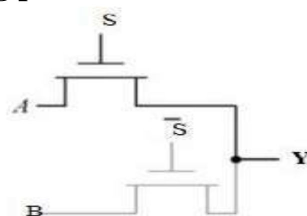
- Supply voltage may vary due to tolerance of voltage regulators, IR drop along the supply rail and di/dt noise.
- Typically the supply is specified as +/- 10% around nominal (uniform distribution).
- Speed is proportional to VDD, also noise budgets are affected.

Temperature

Parts must operate over a range of temperatures.

Standard	Minimum	Maximum
Commercial	0°C	70°C
Industrial	-40°C	85°C
Military	-55°C	125°C

63. Implement a 2:1 multiplexer using pass transistor. (NOV/DEC-2013, April 2015)

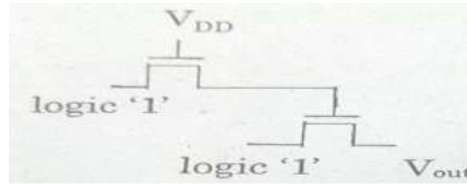


64. Compare static and dynamic power dissipation. [Nov 2019]

- Static power is power consumed while there is no circuit activity. For example, the power consumed by a D flip-flop when neither the clock nor the D input have active inputs (i.e., all inputs are "static" because they are at fixed dc levels).

- Dynamic power is power consumed while the inputs are active. When inputs have ac activity, capacitors are charging and discharging and the power increases as a result. The dynamic power includes both the ac component as well as the static component.

65. What is the value of V_{out} for the figure shown below, where V_{tn} is threshold voltage of transistor? (Nov 2016)



Output voltage, $V_{out} = V_{DD} - 2V_{tn}$, Where V_{tn} : Threshold voltage

66. How does a transmission gate produce fully restored logic output? (NOV 2021)

A transmission gate is parallel pairs of nMOS and pMOS transistor. A single nMOS or pMOS pass transistor suffers from a threshold drop. Transmission gates solve the threshold drop but require two transistors in parallel.

One of the transistors is passing the value well and the other is passing it poorly. A logic-1 is passed well through the pMOS but poorly through the nMOS. A logic-0 is passed well through the nMOS but poorly through the pMOS.

67. What is charge sharing in dynamic CMOS logic? [Nov/Dec-2022]

Charge sharing problem occurs when the charge which is stored at the output node in the pre-charge phase is shared among the junction capacitance of transistor in the evaluation phase. Charge sharing may degrade the output voltage level or even cause an erroneous output value.

68. What is use of transmission gates? [April/May-2022]., [Nov/Dec-2020 & April/May-2021]

<u>Used as a</u>

- Logic structure
- Switch
- Latch
- Used solution to deal with the voltage-drop problem.
- Complex gates can be implemented using minimum number of transistors, which also reduces parasitics.

69. List the sources of power dissipation in CMOS circuits. [April/May-2022]

<u>Static CMOS design:</u>

- Bubble Pushing
- Compound gates
- Skewed gates

<u>Dynamic CMOS design:</u>

- Dual rail domino logic

- Multiple output domino logic

****

Design a half adder using static CMOS logic. [Nov/Dec 2022]

$S = A \oplus B$ ...... Sum

$S = \bar{A}B + A\bar{B}$

Let $\bar{S} = \overline{\bar{A}B + A\bar{B}}$

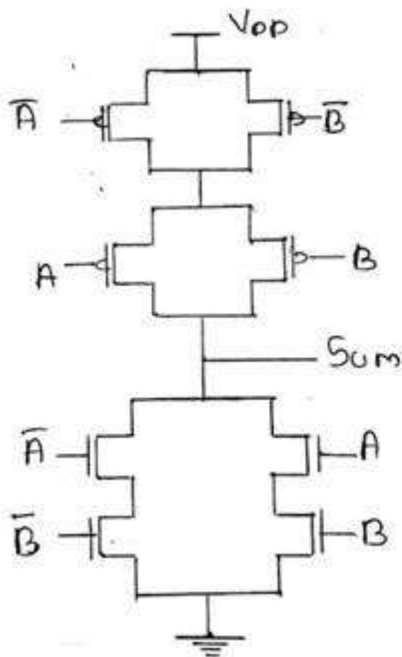
$\therefore \bar{\bar{S}} = \bar{A}\bar{B} + AB$ ......implement as pull down network (nMOS)

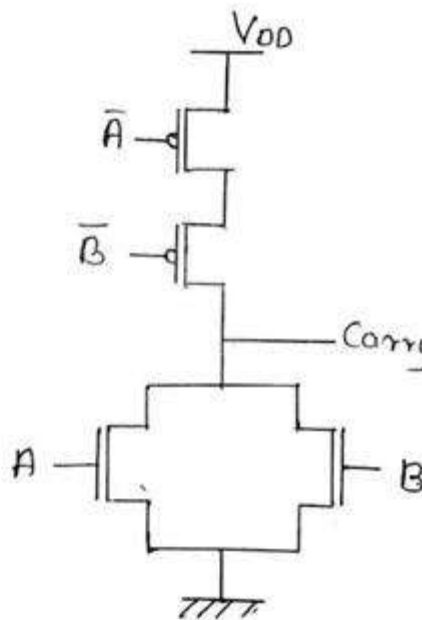
$\bar{S}_{dual} = (\bar{A} + \bar{B}).(A + B)$ ......Implement as pull up network (pMOS)

$C = A.B$ ..... Carry

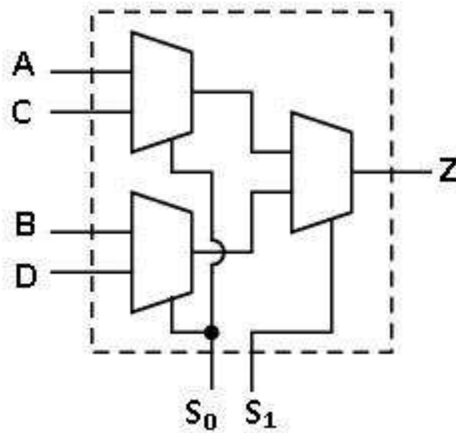
Let $\bar{C} = \overline{A.B}$ ......implement as pull down network (nMOS)

$\bar{C}_{dual} = \bar{A} + \bar{B}$ ......Implement as pull up network (pMOS)





Design a 4:1 MUX using 2:1 MUX. Realize it using transmission gate. [Nov/Dec 2022]



$$Z = \left(A.\bar{S}_0.\bar{S}_1\right) + \left(B.\bar{S}_0.S_1\right) + \left(C.S_0.\bar{S}_1\right) + \left(D.S_0.S_1\right)$$

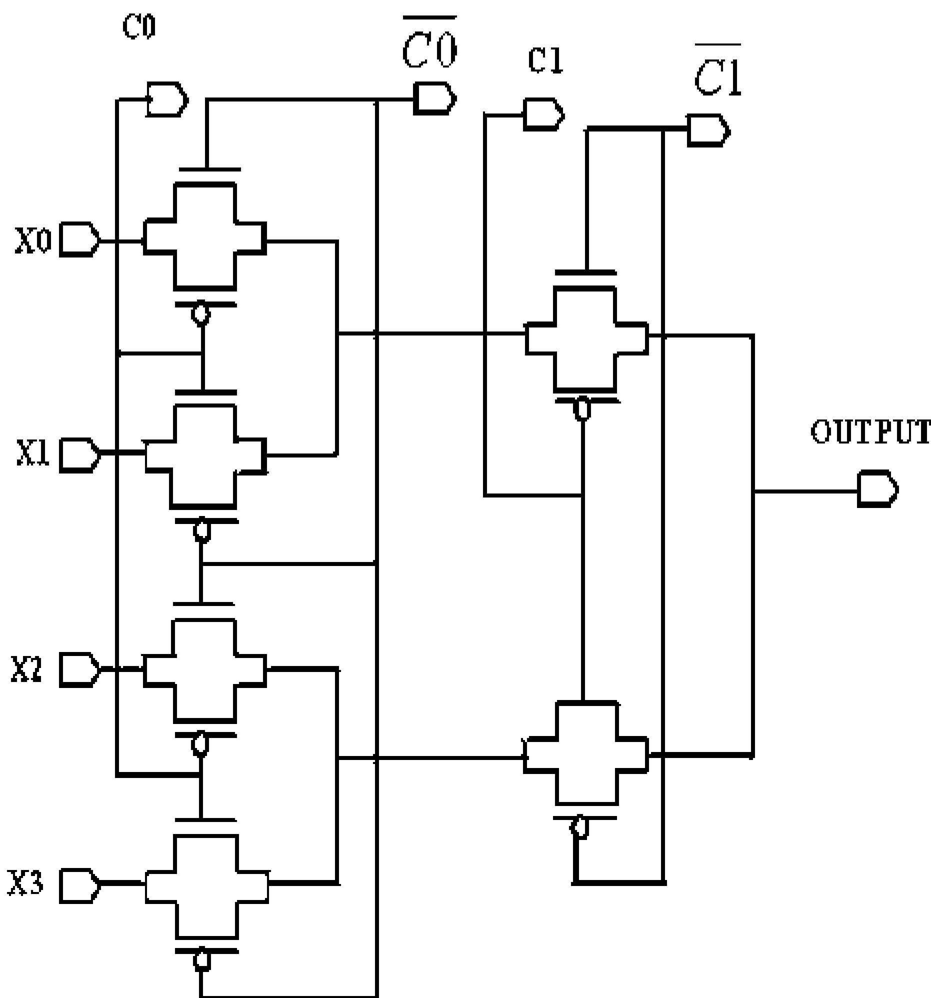
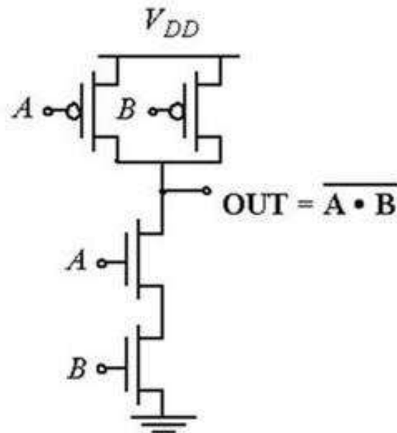


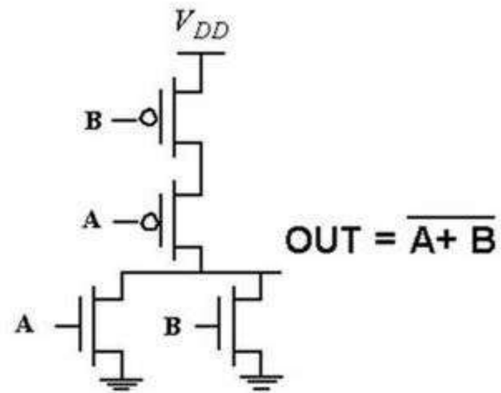
Fig: Transmission gate of 4:1 MUX using 2:1 MUX

Realize a 2-input NOR gate, NAND gate, XOR gate, XNOR gate using static CMOS logic.[Apr/May 2022]

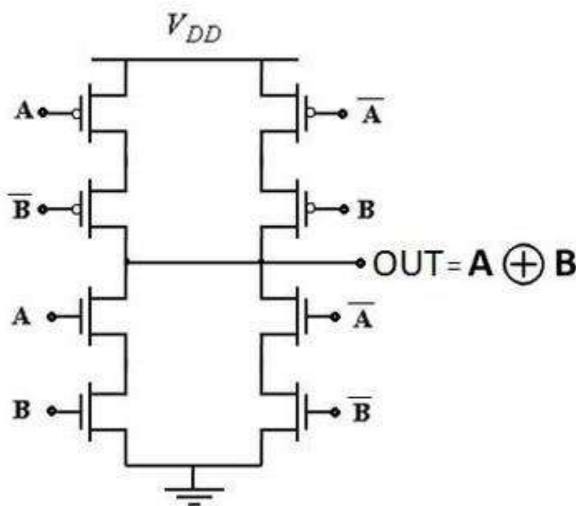
Realize a 2-input NOR gate using static CMOS logic, Domino logic and Complementary pass transistor logic. Analyze the hardware complexity in terms of transistor count. [Nov/Dec 2022]



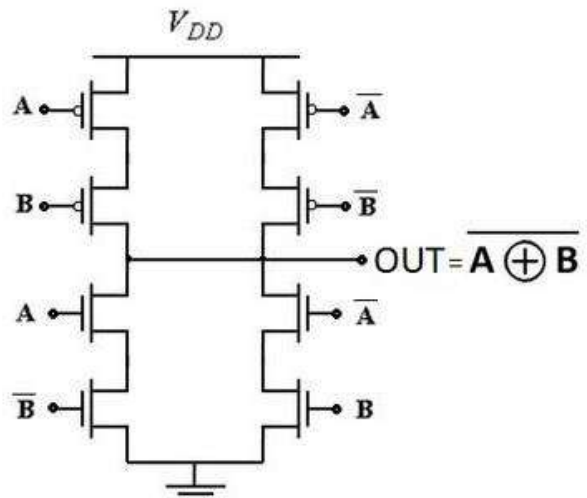
2 input NAND gate (Static Logic)



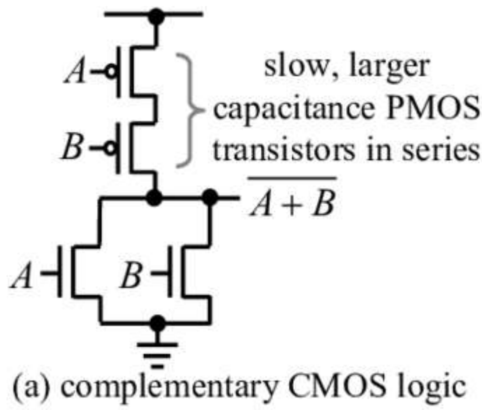
2 input NOR gate (Static Logic)



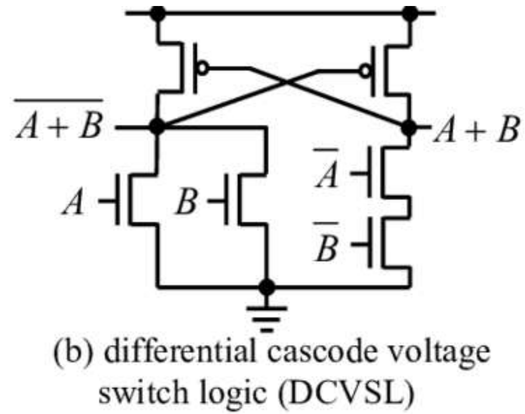
2 input XOR gate (Static logic)



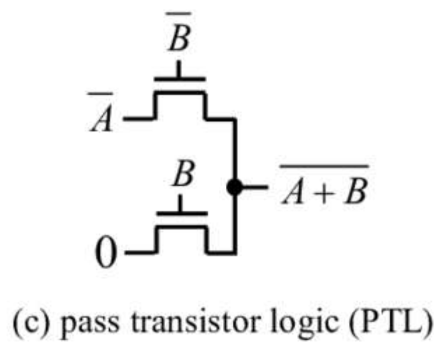
2 input XNOR gate (Static logic)



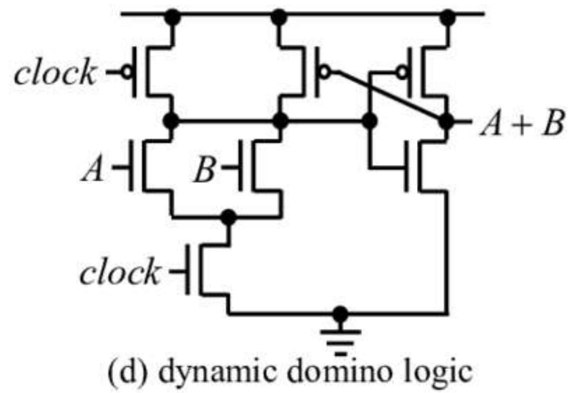
(a) complementary CMOS logic



(b) differential cascode voltage switch logic (DCVSL)



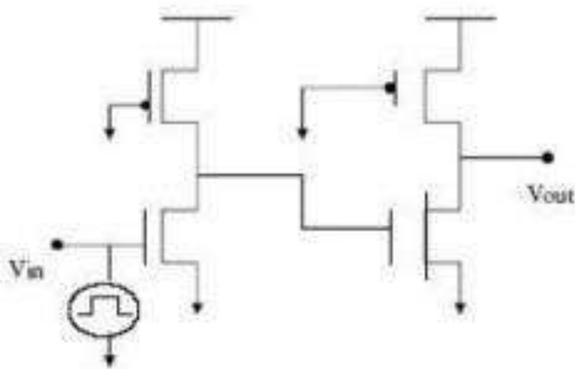
(c) pass transistor logic (PTL)



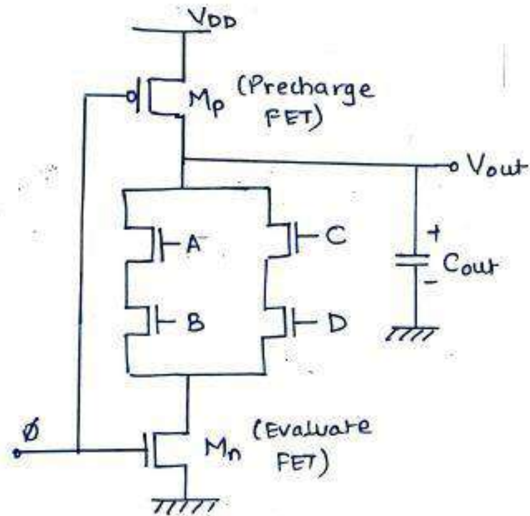
(d) dynamic domino logic

Sketch a combinational function Y=(AB+CD)'.
(i) Pseudo-nMOS logic
(ii) Domino logic
(iii) Cascode voltage switch logic. [Nov/Dec 2022] [April / May 2023]

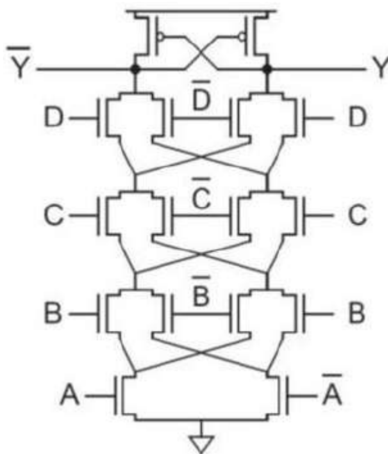
Pseudo-nMOS logic



Domino logic



Cascode voltage switch logic



70. List out the advantages and disadvantages of Pass Transistor Logic. [April/May 2023]

The **advantages** of pass-transistor logic are the simple design, the reuse of already available signals, and the low contribution to static power.

The **disadvantage** of PTL is that the output voltage is lower than the input and it does not allow series connection of a large number of transistors.

71. List any two types of layout design rules. (Nov 2008, Nov 2009, May 2010)

Two types of layout design rules:

a. Lambda design rules

b. Micron rules

72. What are design rules?
What is the need for design rules? (NOV.2014)

Design rules are a set of geometrical specifications that dictate the design of the layout masks.

Design rules are used to produce workable mask layouts from which the various layers in silicon will be formed or patterned.

73. Define the lambda layout rules. (May 2013)
What is meant by lambda layout design rules?
Discuss any two layout design rules. (MAY 2014, APRIL2015, Nov 2015, Nov 2008, Nov 2009, May 2010)

Lambda-based rules are round up dimensions of scaling to an integer multiple of scalable parameter λ. Lambda rules make scaling layout as small. The same layout can be moved to a new process by specifying a new value of λ.

Micron rules can result in as much as a 50% size reduction over lambda rules. Industry usually uses the micron design rules for layouts.

74. By what factor, gate capacitance must be scaled if constant electric field scaling is employed? (April 2019)

Gate capacitance is scaled by scaling factor $\frac{1}{S}$ in constant electric field scaling.

75. What are stick diagrams?

Stick diagrams are used to convey layer information through the use of a color code. A stick diagram is a cartoon of a chip layout. The stick diagram represents the rectangles with lines which represent wires and component symbols.

76. What are the uses of Stick diagram?

Uses of stick diagram:

- It can be drawn much easier and faster than a complex layout.
- These are especially important tools for layout built from large cells.

77. Give the various color coding used in stick diagram?

Various color coding used in stick diagram:

- Green – n-diffusion, Red- polysilicon, Blue –metal, Yellow- implant and Black-contact areas.

78. Why does interconnect increase the circuit delay? [Nov/Dec-2011]

Interconnect is defined by its resistance value and capacitance with neighbor. Delay is calculated from resistance and capacitance value.

79. What is transistor sizing problem?(MAY 2014)

Transistor sizing is carried out by equating the maximum on resistances of the logic circuit with inverter one.

80. What is CMOS latchup? How do you prevent Latch up problem? (Nov 2008) (or)
What is Latch up problem in CMOS circuits? (May 2008, April 2016)

Latch up is a condition in which the parasitic components give rise to the establishment low resistance conducting paths between V_{DD} and V_{SS} with disastrous results.

Careful control during fabrication is necessary to avoid this problem.

The remedies for the latch-up problem include:

(i) An increase in substrate doping levels.

(ii) Reducing Rnwell.
(i) By introducing guard rings.
(ii) By introducing SOI (Silicon On Insulator)

81. What is BiCMOS Gate?
Compare CMOS and BiCMOS technology. (NOV. 2013)

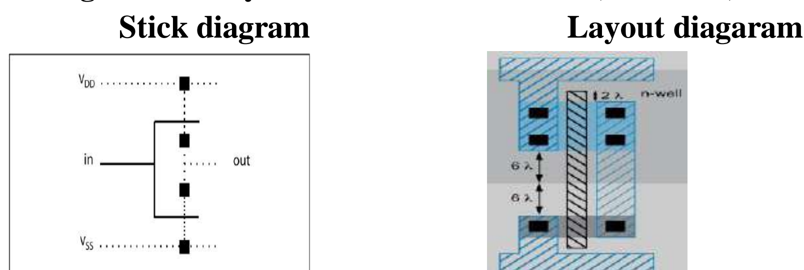
When bipolar and MOS technology are merged, the resulting circuits are referred to as biCMOS circuits. It improves bandwidth and current gain.

82. What is the need of demarcation line? (Nov 2017)

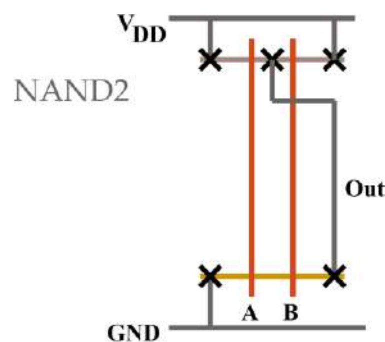
In CMOS, a demarcation line is drawn to avoid touching of p-diff with n-diff. All pMOS must lie on one side of the line and all nMOS must lie on other side.

83. Draw the stick diagram and layout for CMOS inverter. (Nov 2016)

Stick diagram **Layout diagaram**



84. Draw the stick diagram of static CMOS 2-input NAND gate. (April 2018)



85. What are simulations available for VLSI circuits?

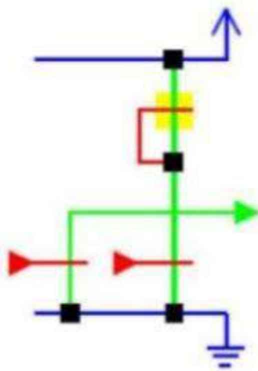
In VLSI the following simulations are available

- Path simulation, Monte Carlo simulation and Interconnect simulation

86. Why nMOS transistor is selected as pull down network? (Nov 2017)

Pull-up and pull-down networks in CMOS circuits are never both conducting and are never both opened at the same time. This is the reason that **nMOS** transistors are used in the **pull-down** network and pMOS in the **pull**-up network of a CMOS gate.

87. Draw the stick diagram of NMOS NOR gate. [Nov 2019]



88. How do you describe the term device modeling? [May/June-2013]

The device modeling describes how to model diffusion capacitance and how to run simulations in various process corners.

89. What is Elmore's delay model? (or) Give the expression for Elmore delay and state the various parameters associated with it. (NOV. 2014, April 2016, 2017, 2018, Nov 2017) [April/May – 2023]

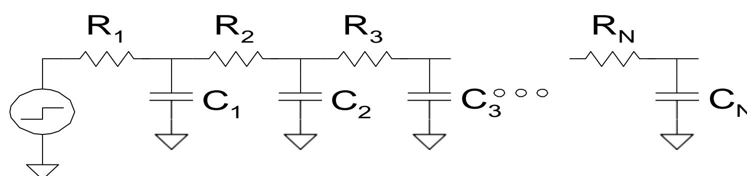
The Elmore delay model estimates the delay from a source switching to one of the leaf nodes. Delay is summing over each node i of the capacitance C_i on the node multiplied by the effective resistance R.

Propagation delay time:

$$t_{pd} \approx \sum_{\text{nodes } i} R_{i-to-source} C_i$$

$$= R_1C_1 + (R_1 + R_2)C_2 + ... + (R_1 + R_2 + ... + R_N)C_N$$

RC delay equivalent for series of transistors:



90. Define logical effort and give logical effort value of inverter.

Logical effort (g) is defined as the ratio of the input capacitance of a gate to the input capacitance of an inverter delivering the same output current.

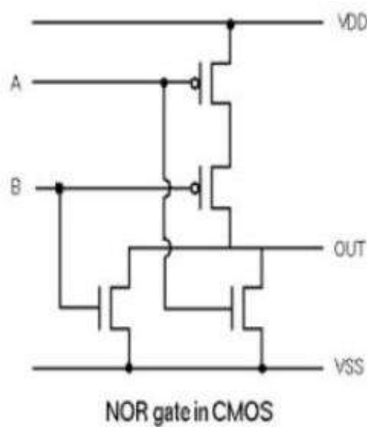
An inverter has a logical effort of 1.

91. Write the general expression of logical effort for n inputs NAND and NOR gate?

Expression of logical effort for n inputs NAND is (n+2)/3.

Expression of logical effort for n inputs NOR is (2n+1)/3. Where, n – no. of inputs.

92. Draw a 2- input CMOS NOR Gate. [Nov/Dec-2022]



NOR gate in CMOS

93. Write the expression for parasitic delay and logical effort of an N-input NAND gate. [April/May-2022]

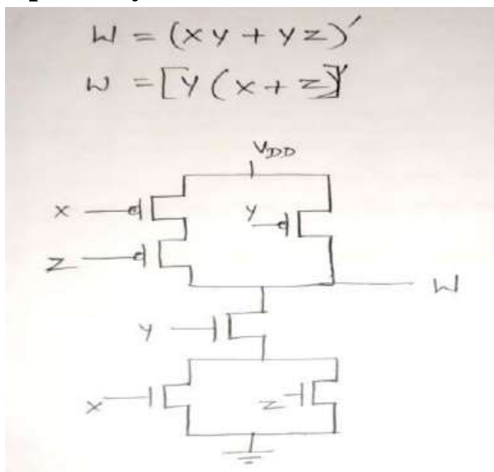
Solution :

Parasitic delay for N-input NAND gate = n

Logical effort of an N-input NAND gate = (n+2)/3

94. Sketch a complementary CMOS gate computing W = (XY+YZ)'. [April/May-2022]

Sketch a complementary CMOS gate computing Y = (AB+BC)'. [Nov/Dec-2020, April/May-2021]



95. What is body effect. [Nov/Dec-2020, April/May-2021]

V_t is not constant with respected to voltage difference between substrate and source of MOS transistor. This is known as body effect. Its other name is substrate-bias effect.

96. What is velocity saturation effect?

The velocity of charge carriers is linearly proportional to the eclectic field and the proportionally constant mobility of carrier.

When we increase the electric field beyond certain velocity called as the **thermal velocity or saturation velocity.**

Electron attains in the presence of very high electric fields.

At high electric field carriers fail to follow this linear model.

97. Define propagation delay of a CMOS inverter.

It expresses the delay experienced by a signal when passing through a gate. It is measured between the **50%** transition points of the input and output waveforms, as shown in Figure for an inverting gate. Because a gate displays different response times for rising or falling input waveforms, two definitions of the propagation delay are necessary. The $\mathbf{t_{pLH}}$ defines the response time of the gate for a low to high (or positive) output transition, while $\mathbf{t_{pHL}}$ refers to a high to low (or negative) transition. The propagation delay $\mathbf{t_p}$ is defined as the average of the two.

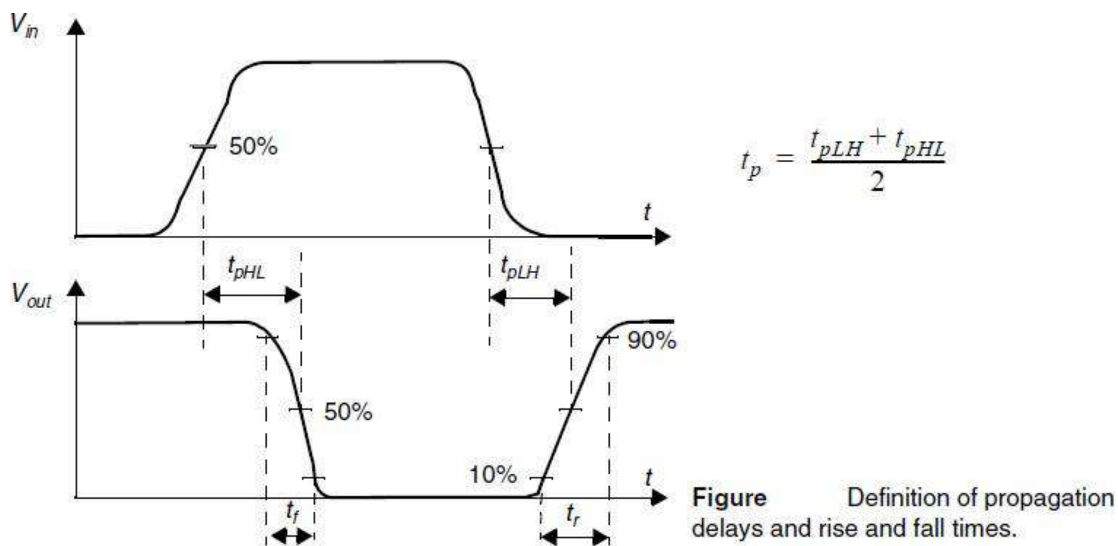


Figure Definition of propagation delays and rise and fall times.

98. Why NMOS device conducts strong zero and weak one?

Reason:

- Greater switching speed.
- Weak transistor is used to generate a high output voltage level
- The gate is "high" and channel is 'low'.
- Nmos is turn on. Vg is at V_{DD} V_S charging towards V_{DD}.

99. What is Intrinsic and Extrinsic Semiconductor?

The pure Silicon is known as Intrinsic Semiconductor. When impurity is added with pure Silicon, its electrical properties are varied. This is known as Extrinsic Semiconductor.

100. State the channel length modulation. Write the equation for describing channel length modulation effect in NMOS transistor.

Channel length is varied due to changes in $\mathbf{V_{ds}}$ (drain to source voltage). In saturation region, channel length is decreased when (W/L) ratio is increased. So β is increased and drain voltage is increased.

101. What is latch up? How is Prevent latch up?

Latch up is the condition occurs in the circuit manufactured using bulk CMOS technology. When IC is the state of "Latch Up".

Latch Up Prevention in two ways:

- Latch up resistant CMOS processes
- Layout technique.

102. What are the different MOS layers?

- n-diffusion
- p-diffusion
- Polysilicon
- Metal

103. If two CMOS inverters are cascaded with an aspect ratio of 1:1 then determine the inverter pair delay. [Nov/Dec-2022]

Solution :

Logical Effort of the inverter , g = 1

Here, single identical load. So, the electrical effort , h =1

Parasitic delay of an inverter , P_{inv} =1

Then, the delay of each stage is expressed as,

d = gh + p

= 1(1)+1

= 2

104. Differentiate static and dynamic latches and registers. [Nov/Dec-2020., April/May-2021]

Difference between Static latches & registers and Dynamic latches & registers:

Static Latches & registers	Dynamic Latches & registers
A latch is an essential component in the construction of an edge triggered register. It is low sensitive circuit that passes the D input to the Q output when clock signal is High.	Dynamic latches storage in a static sequential circuit relies on the concept that a cross coupled inverter pair.
This latch is said to be in transparent mode	It produces a bistable element and can thus be used to memorize binary values.
A static latches stores its output in a static state.	A dynamic latch uses temporary capacitance to store its state.
This register rarely or never changes.	This register are changeable

UNIT – III

SEQUENTIAL LOGIC CIRCUITS AND CLOCKING STRATEGIES

Static Latches and Registers, Dynamic Latches and Registers, Pipelines, Non-bistable Sequential Circuits, Timing classification of Digital Systems, Synchronous Design, Self-Timed Circuit Design .

3.1 Static Latches and Registers:

- **Discuss in detail various static latches and registers. (Nov 2016)**
- **Explain the methodology of sequential circuit design of Latches. (May 2014)**
- **Discuss the operation of a CMOS latch. (Nov 2007)**

3.1.1 The Bi-stability Principle

- Static memories use positive feedback to create a bistable circuit. A bistable circuit has two stable states that represent 0 and 1.
- The basic idea is shown in Figure 3.1a, which shows two inverters connected in cascade along with a voltage-transfer characteristic (VTC).
- The output of the second inverter Vo2 is connected to the input of the first V i1, as shown by the dotted lines in Figure 3.1a.
- The resulting circuit has only three possible operation points (A, B, and C).
- A and B are stable operation points, and C is a metastable operation point.

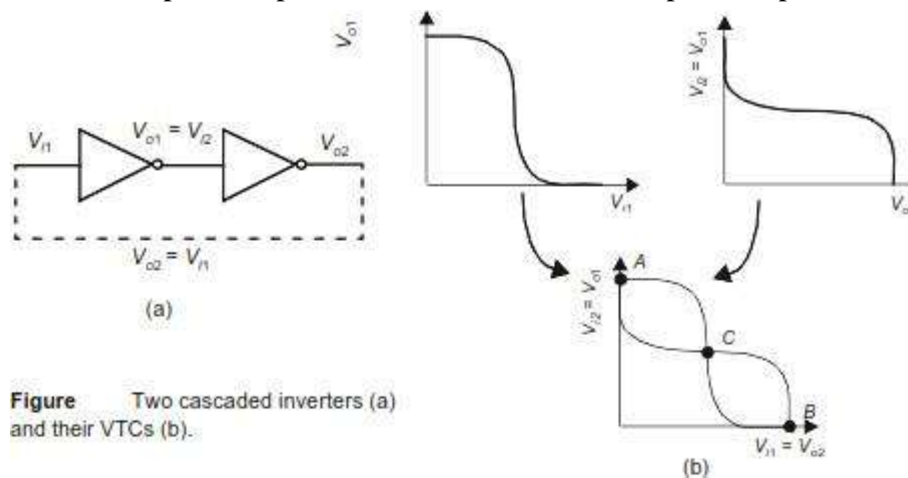
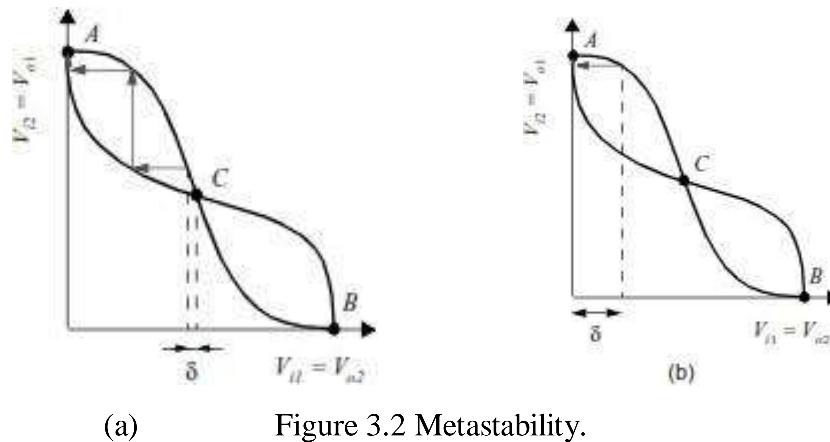


Figure 3.1: a. Two inverters connected in cascade b. VTCs

- Cross-coupled inverter pair is biased at point C. It is amplified and regenerated around the circuit loop.
- The bias point moves away from C until one of the operation points A or B is reached.
- C is an unstable operation point. Every deviation causes the operation point to run away from its original bias. Operation points with this property are termed as metastable.



(a) Figure 3.2 Metastability.

- A bistable circuit has two stable states. In absence of any triggering, the circuit remains in a single state.
- A trigger pulse must be applied to change the state of the circuit.
- Common name for a bistable circuit is flip-flop.

3.1.2 SR Flip-Flops

- The SR or set-reset flip-flop implementation is shown in Figure (a) below.
- This circuit is similar to the cross-coupled inverter pair with NOR gates replacing the inverters.
- The second input of the NOR gates is connected to the trigger inputs (S and R), that make it possible to force the outputs Q and Q_{bar}.
- These outputs are complimentary (except for the SR = 11 state).
- When both S and R are 0, the flip-flop is in a quiescent state and both outputs retain their value.
- If a positive (or 1) pulse is applied to the S input, the Q output is forced into the 1 state (with Qbar going to 0).
- Vice versa, a 1 pulse on R resets the flip-flop and the Q output goes to 0.

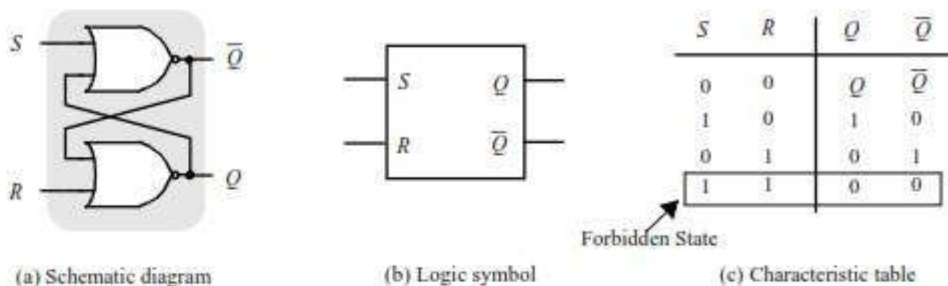


Figure 3.3

- When both S and R are high, both Q and Q_{bar} are forced to zero. This input mode is considered to be forbidden.
- An SR flip-flop can be implemented using a cross-coupled NAND structure as shown in Figure 3.4

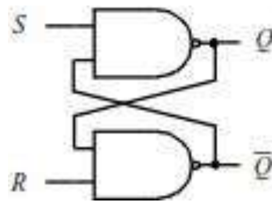


Figure 3.4 NAND based SR flip-flop

Clocked SR flip-flop:

- Clocked SR flip-flop (a level-sensitive positive latch) is shown in Figure 3.5.
- It consists of a cross-coupled inverter pair, plus 4 extra transistors to drive the flip-flop from one state to another and to provide clocked operation.
- Consider the case where Q is high and R pulse is applied.
- The combination of transistors M_4, M_7, and M_8 forms a ratioed inverter.
- In order to make the latch switch, we must succeed in bringing Q below the switching threshold of the inverter M_1-M_2.
- Once this is achieved, the positive feedback causes the flip-flop to invert states. This requirement forces to increase the sizes of transistors M_5, M_6, M_7, and M_8.

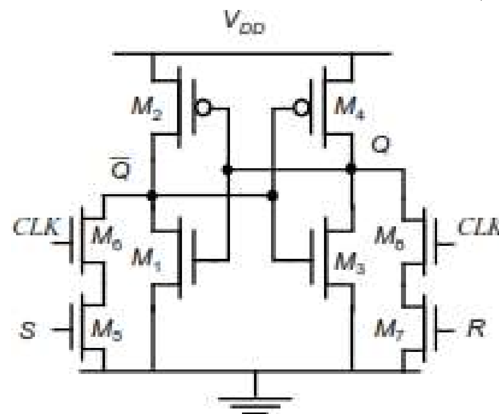


Figure 3.5 CMOS clocked SR flip-flop

- The clocked SR flip-flop does not consume any static power.

3.1.3 Multiplexer Based Latches:

- Multiplexer based latches can provide similar functionality to the SR latch.
- But sizing of devices only affects performance and is not critical to the functionality.
- Figure 3.6 shows an implementation of static positive and negative latches based on multiplexers.
- For a negative latch, when the clock signal is low, the input 0 of the multiplexer is selected, and the D input is passed to the output.
- When the clock signal is high, the input 1 of the multiplexer connected to the output of the latch.
- The feedback holds the output stable while the clock signal is high.

- Similarly in the positive latch, the D input is selected when clock is high and the output is held (using feedback) when clock is low.

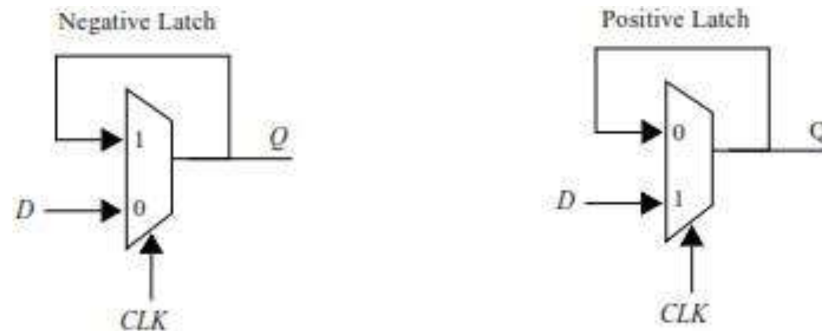


Figure 3.6 Negative and positive latches based on multiplexers.

- ❖ **Design a d-latch using transmission gate. (May 2015)**
- ❖ **Design a D-latch using transmission gate. Using which realize a two phase non – overlapping master slave negative edge triggered D flip-flop. (April 2019-13M)**
- ❖ **Illustrate the circuit designs for basic latches, then build the flip-flops and pulsed latches. [May 2021][Nov/Dec 2022]**

- A transistor level implementation of a positive latch is shown in Figure 3.7.
- When CLK is high, the bottom transmission gate is ON and the latch is transparent – i.e, the D input is copied to the Q output.
- During this phase, the feedback loop is open due to the top transmission gate is OFF.

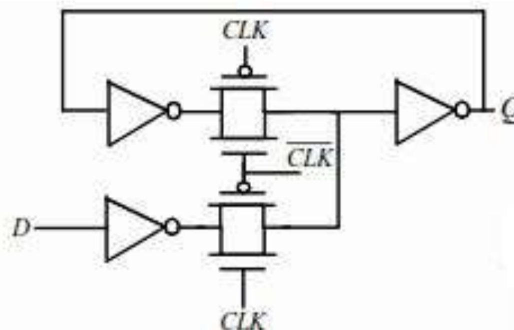


Figure 3.7 Transistor level implementation of a positive latch built using transmission gates.

- To reduce the clock load, implement a multiplexer based NMOS latch using two pass transistors as shown in Figure 3.8.
- The advantage of this approach is the reduced clock load of only two NMOS devices.
- When CLK is high, the latch samples the D input, while a low clock-signal enables the feedback-loop and puts the latch in the hold mode.

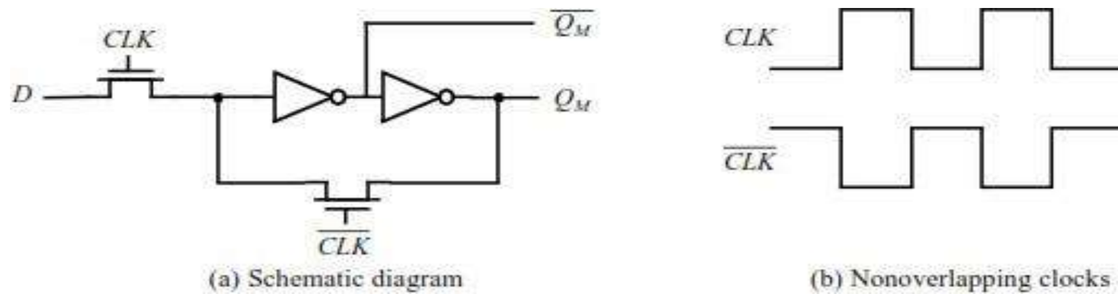


Figure 3.8 Multiplexer based NMOS latch using NMOS only pass transistors for multiplexers.

- **Explain the operation of master-slave based edge triggered register. (May 2016)**
- **Draw and explain the operation of conventional CMOS, pulsed and resettable latches. (Nov 2012)**
- **Discuss about CMOS register concept and design master slave triggered register, explain its operation with overlapping periods. (April 2018, NOV 2018)**
- **Realize a negative level sensitive latch using which realize an edge triggered master slave D-Flip flop. Explain its working. (Nov 2019) [April / May 2023]**

3.1.4 Master-Slave Based Edge Triggered Register:

- An edge-triggered register is to use a master-slave configuration as shown in Figure 3.9.
- The register consists of cascading a negative latch (master stage) with a positive latch (slave stage).
- A multiplexer based latch is used to realize the master and slave stages.
- On the low phase of the clock, the master stage is transparent and the D input is passed to the master stage output, Q_M.
- During this period, the slave stage is in the hold mode, keeping its previous value.
- On the rising edge of the clock, the master slave stops sampling the input and the slave stage starts sampling.
- During the high phase of the clock, the slave stage samples the output of the master stage (Q_M), while the master stage remains in a hold mode.
- A negative edge-triggered register can be constructed using the same principle by simply switching the order of the positive and negative latch (i.e., placing the positive latch first).

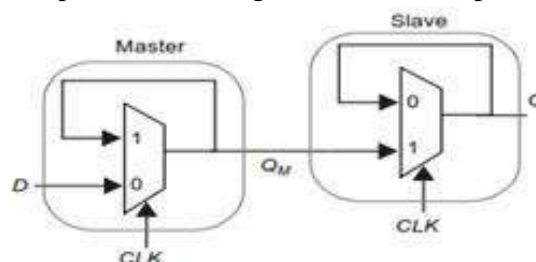


Figure 3.9: Positive edge-triggered register based on a master-slave configuration.

- A complete transistor level implementation of the master-slave positive edge-triggered register is shown in Figure 3.10.
- When clock is low (CLK bar = 1), T_1 is ON and T_2 is OFF and the D input is sampled onto node Q_M.
- During this period, T_3 is OFF and T_4 is ON and the cross-coupled inverters (I_5 & I_6) hold the state of the slave latch.

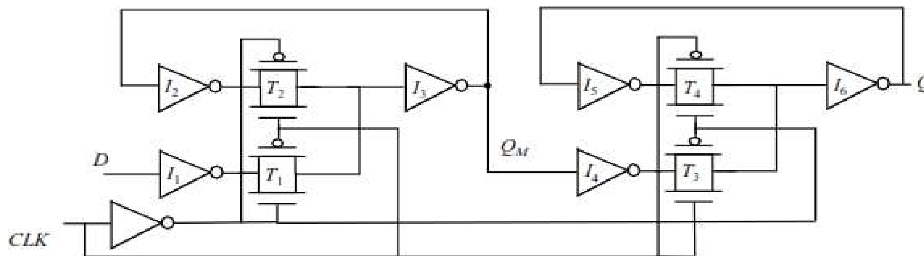


Figure 3.10 Transistor-level implementation of a master-slave positive edge-triggered register using multiplexers.

- When the clock goes high, the master stage stops sampling the input and goes into a hold mode.
- T_1 is OFF and T_2 is ON and the cross coupled inverters I_3 and I4 hold the state of Q_M. Also T_3 is ON and T_4 is OFF and Q_M is copied to output Q.

3.1.5 Non-ideal clock signals:

- We have assumed that $\overline{CLK}$ is a perfect inversion of CLK.
- Even if this was possible, this would still not be a good assumption.
- Variations can exist in the wires. It is used to route the two clock signals or the load capacitances can vary based on data stored in the connecting latches.
- This effect, known as clock skew is a major problem and causes the two clock signals to overlap as is shown in Figure 3.11 b.
- Clock-overlap can cause two types of failures, as illustrated for the NMOS-only negative master-slave register of Figure 3.11 a.

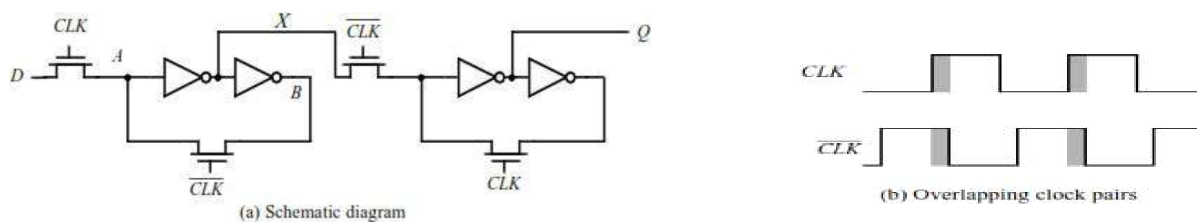


Figure 3.11 Master-slave register based on NMOS-only pass transistors.

- When the clock goes high, the slave stage should stop sampling the master stage output and go into a hold mode.
- However, since CLK and $\overline{CLK}$ are both high for a short period of time (the overlap period).

- Both sampling pass transistors conduct and there is a direct path from the D input to the Q output.
- As a result, data at the output can change on the rising edge of the clock, which is undesired for a negative edge triggered register.
- The is known as a race condition in which the value of the output Q is a function of whether the input D arrives at node X before or after the falling edge of $\overline{CLK}$.
- If node X is sampled in the metastable state, the output will switch to a value determined by noise in the system.
- Those problems can be avoided by using two non-overlapping clocks PHI and PH2.
- By keeping the non-overlap time $t_{non_overlap}$ between the clocks large enough, such that no overlap occurs even in the presence of clock-routing delays.
- During the non-overlap time, the FF is in the high-impedance state.
- Leakage will destroy the state, if this condition holds for too long a time.
- Hence the name ***pseudostatic***: the register employs a combination of static and dynamic storage approaches depending upon the state of the clock.

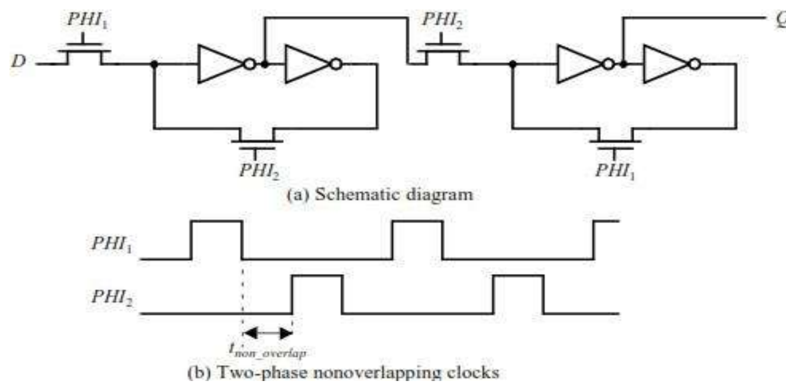


Figure 3.12 Pseudostatic two-phase D register.

3.1.6 Low-Voltage Static Latches:

- The scaling of supply voltages is critical for low power operation.
- At very low power supply voltages, the input to the inverter cannot be raised above the switching threshold, resulting in incorrect evaluation.
- Scaling to low supply voltages, hence requires the use of reduced threshold devices.
- The shaded inverters and transmission gates are implemented in low-threshold devices.
- The low threshold inverters are gated using high threshold devices to eliminate leakage.
- During normal mode of operation, the sleep devices are tuned on.
- During idle mode, the high threshold devices in series with the low threshold inverter are turned OFF (the SLEEP signal is high), eliminating leakage.

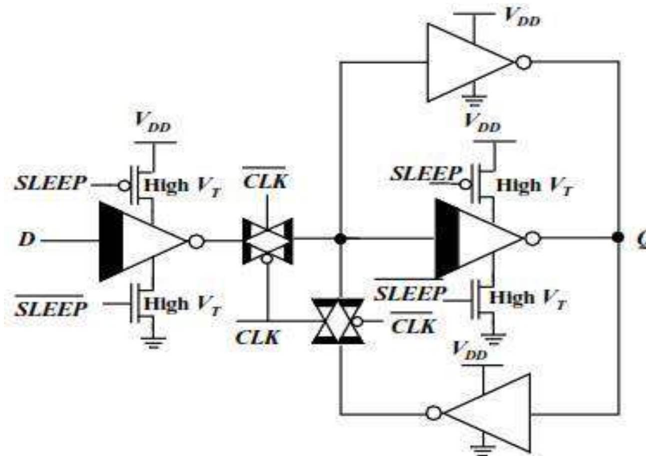


Figure 3.13: One solution for the leakage problem in low-voltage operation using MTCMOS.
****

3.2 Dynamic Latches and Registers:

- ❖ **Discuss about the design of sequential dynamic circuits. (Nov 2012, Nov 2017)**
- ❖ **Explain the methodology of sequential circuit design of flip-flop. (May 2014)**

- A stored value remains valid as long as the supply voltage is applied to the circuit, hence the name static.
- The major disadvantage of the static gate is, its complexity.
- Registers are used in computational structures that are constantly clocked, such as pipelined data path.
- The requirement that the memory should hold state for extended periods of time.
- This results in circuits, based on temporary storage of charge on parasitic capacitors.
- The principle is identical to the dynamic logic. In dynamic logic, logic signal is a charge, stored on a capacitor.
- The absence of charge denotes as logic 0 and presence of charge denotes as logic 1.
- A stored value can be kept for a limited amount of time (range of milliseconds).
- A periodic refresh of its value is necessary.

3.2.1 Dynamic Transmission-Gate Based Edge-triggered Registers:

❖ **Design a d-flipflop using transmission gate. (Nov 2016)**

- A dynamic positive edge-triggered register based on the master-slave concept is shown in Figure 3.14.
- When CLK = 0, the input data is sampled on storage node 1. It has an equivalent capacitance of C1 consisting of the gate capacitance of I_1, the junction capacitance of T_1, and the overlap gate capacitance of T_1.

- During this period, the slave stage is in a hold mode with node 2 in a high-impedance state.
- On the rising edge of clock, the transmission gate T_2 turns on. The value is sampled on node1 before the rising edge propagates to the output Q.
- Node 2 stores the inverted version of node 1.
- The reduced transistor provides high-performance and low-power systems.

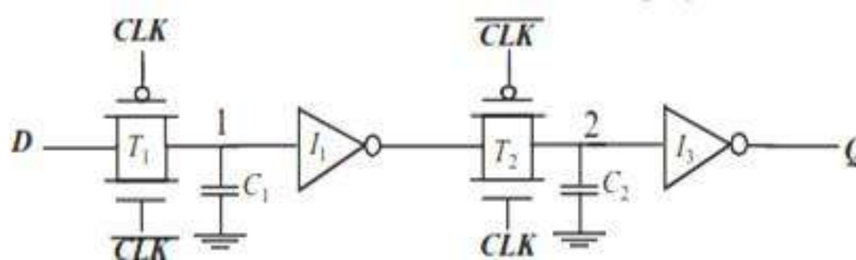


Figure 3.14 Dynamic edge-triggered register.

- The set-up time of this circuit is the delay of the transmission gate and time takes node 1 to sample the D input.
- The hold time is approximately zero, since the transmission gate is turned OFF.
- The propagation delay (t_{c-q}) is equal to two inverter delays plus the delay of the transmission gate T_2.
- Dynamic register has storage nodes (i.e., the state). A node has to be refreshed at periodic intervals to prevent a loss. Loss due to charge leakage and diode leakage.
- Clock overlap is an important for register. Consider the clock waveforms shown in Figure 3.15.
- During the 0-0 overlap period, the NMOS of T_1 and the PMOS of T_2 are simultaneously on.
- It is creating a direct path for data to flow from the D input of the register to the Q output. This is known as a race condition.
- The output Q can change on the falling edge, if the overlap period is large.
- Overlap period constraint is given as: $t_{overlab0-0} < t_{T1} + t_{I1} + t_{T2}$
- Similarly, the constraint for the 1-1 overlap is given as: $t_{hold} > t_{overlab1-1}$

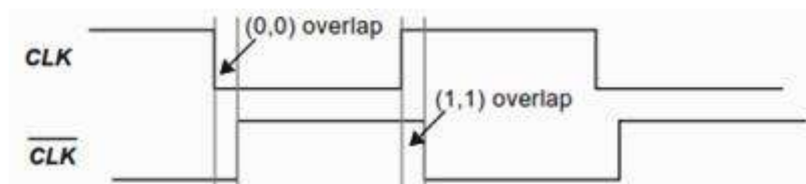


Figure 3.15 Impact of non-overlapping clocks

3.2.2 C²MOS Dynamic Register:

The C²MOS Register

- Figure 3.16 shows positive edge-triggered register based on the master-slave concept, which is insensitive to clock overlap. This circuit is called the C^2MOS (Clocked CMOS) register.

1. CLK = 0 ($\overline{CLK}$ = 1):
    - The first tri-state driver is turned ON. The master stage acts as an inverter, sampling the inverted of D on the internal node X.
    - The master stage is, in evaluation mode. The slave section is, in hold mode.
    - Both transistors M_7 and M_8 are OFF, decoupling the output from the input. The output Q retains its previous value stored on the output capacitor C_{L2}.

2. CLK = 1 ($\overline{CLK}$ = 0):
    - The master stage section is in hold mode (M_3-M_4 off), while the second section evaluates (M_7-M_8 on).
    - The value stored on C_{L1} propagates to the output node through the slave stage, which acts as an inverter.
    - The overall circuit operates as a positive edge-triggered master-slave register.

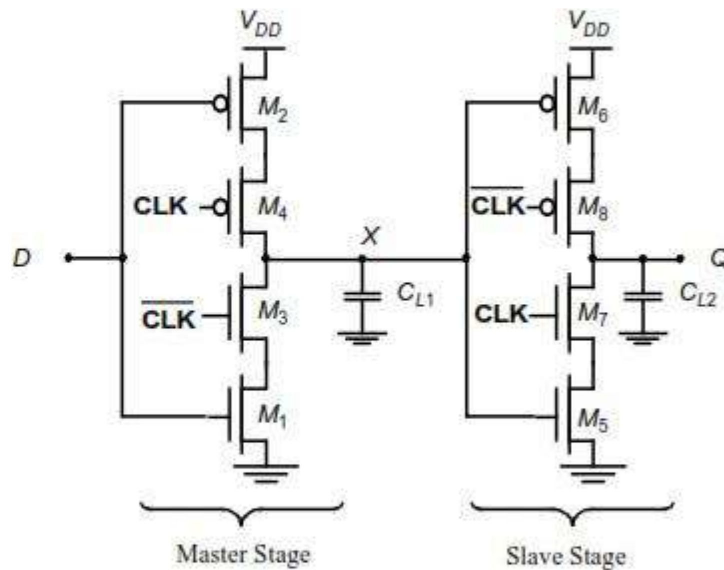


Figure 3.16 C^2MOS master-slave positive edge-triggered register.

- It is similar to the transmission-gate based register, presented earlier. However, there is an important difference.
- *A C^2MOS register with CLK-CLK clocking is insensitive to overlap, as long as the rise and fall times of the clock edges are small.*

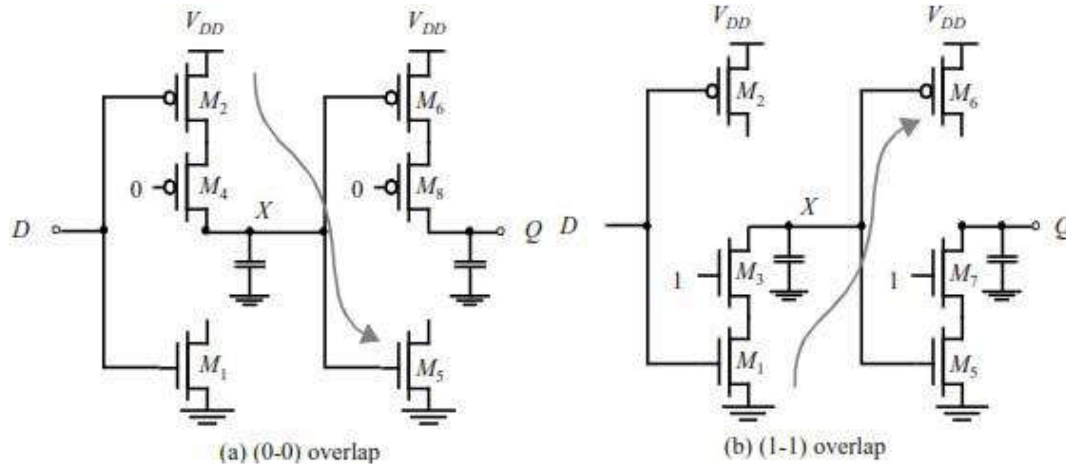


Figure 3.17 C^2MOS D FF during overlap periods.

3.2.3 True Single-Phase Clocked Register (TSPCR):

Explain the operation of True Single Phase Clocked Register. (Nov 2016, April 2017)

- In the two-phase clocking schemes, care must be taken in routing the two clock signals to ensure that overlap is minimized.
- While the C^2MOS provides a skew-tolerant solution, it is possible to design registers that only use a single phase clock.
- The True Single-Phase Clocked Register (TSPCR) uses a single clock without an inverse clock.
- Figure 3.19 shows positive and negative latch concept.
- For the positive latch, when CLK is high, the latch is in the transparent mode and propagates the input to the output. Latch has two cascaded inverters, so latch is non-inverting.
- When CLK = 0, both inverters are disabled and the latch is, in hold-mode.
- Only the pull-up networks are still active, while the pull-down circuits are deactivated.
- As a result of the dual-stage approach, no signal can ever propagate from the input to the output.
- For the negative latch, when CLK is low, the latch is in the transparent mode and propagates the input to the output.
- When CLK = 1, both inverters are disabled and the latch is in hold-mode.
- A register can be constructed by cascading positive and negative latches.
- The main advantage is the use of a single clock phase.
- The disadvantage is, increase in the number of transistors (12 transistors are required).

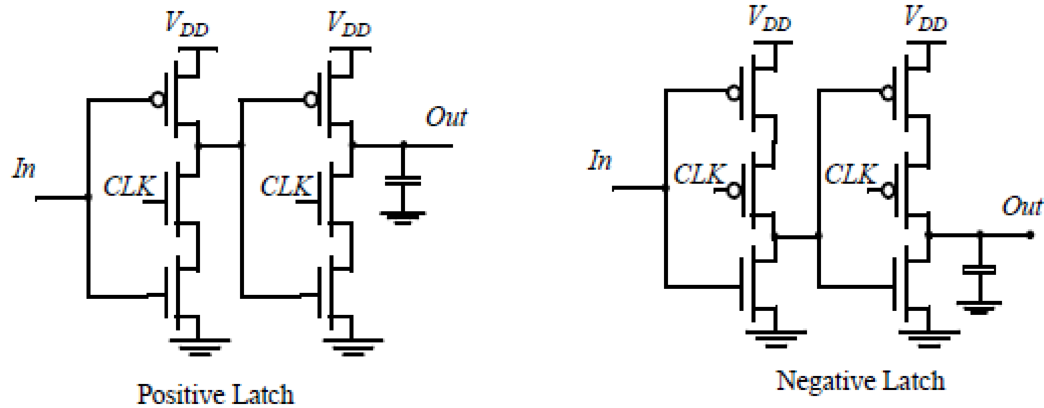


Figure 3.19 TSPC approach.

- The TSPC latch circuits can be reduced, as in Figure 3.20, where only the first inverter is controlled by the clock.
- Number of transistors are reduced and clock load is reduced by half.

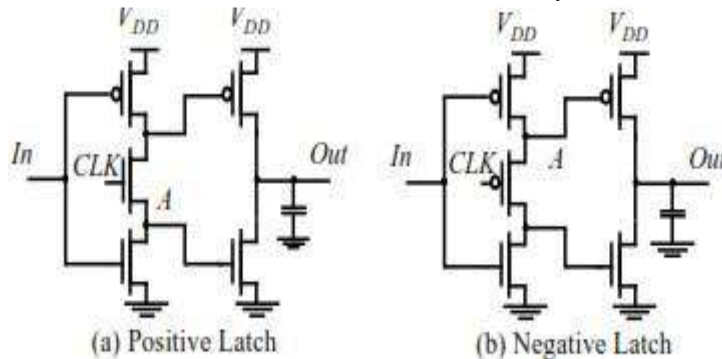


Figure 3.20 Simplified TSPC latch (also called split-output).

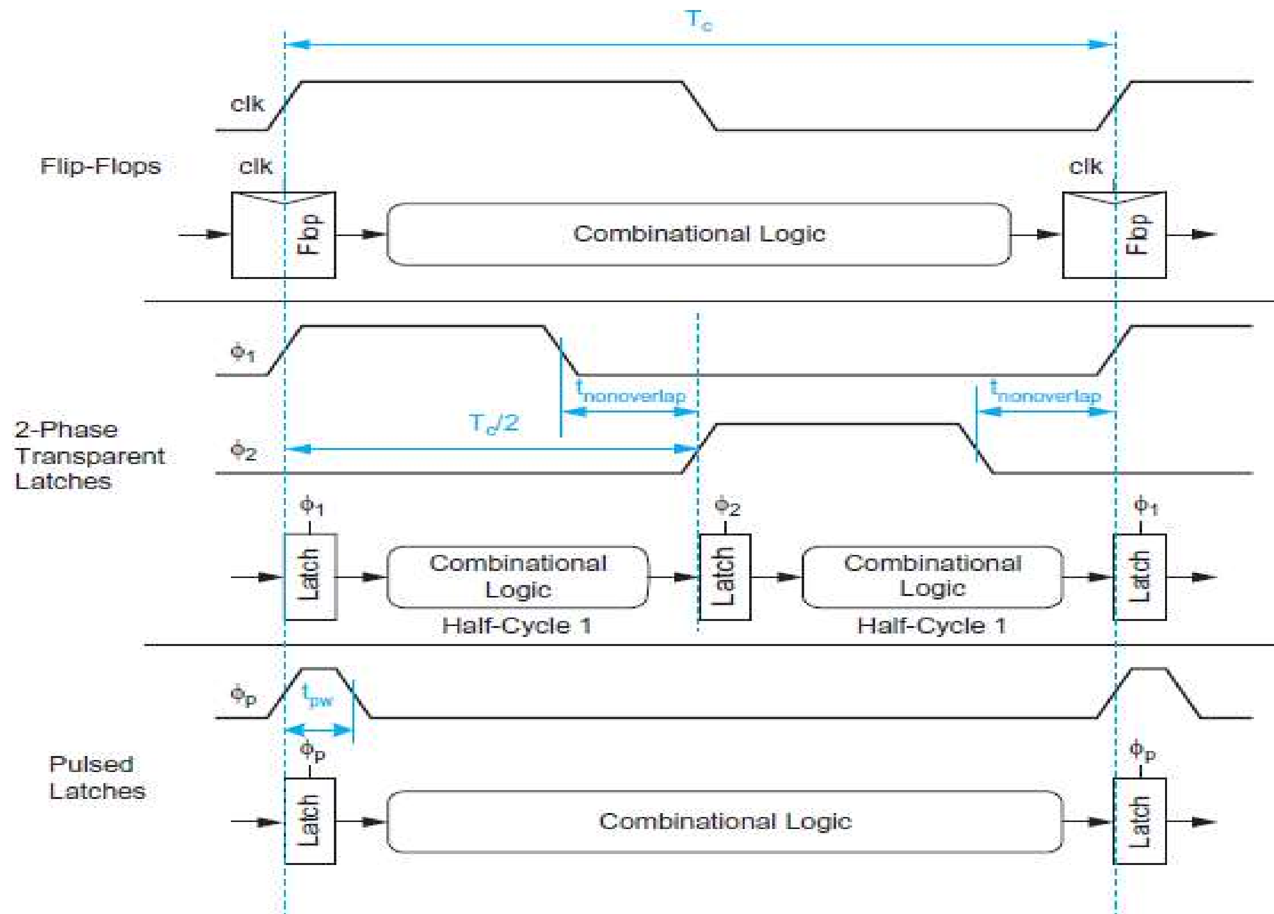
****

3.3 Timing Issues:

- ❖ **Explain in detail about timing issues needed for a logic operation. (April 2017)**
- ❖ **Explain the timing basics in synchronous design in detail. (Nov 2017)[April/May 2023]**
- ❖ **Discuss the timing parameters that characterize the timing of sequential circuit. (NOV 2021)**

(A) Sequencing methods:-

- Three methods of sequencing block of combinational logic are possible, as shown in figure below.
- In flip-flop based system, one flip flop use one cycle boundary.
- Token (data) advances from one cycle to the next on the rising edge. If a token arrives too early, it waits at the flip flop until next cycle.
- In 2-phase system, phases may be separated by $t_{nonoverlap}$. [$t_{nonoverlap}>0$]
- In pulsed system, pulse with is t_{pw}.



- In 2-phase system, full cycle of combinational logic is divided into two phases, sometimes called "half-cycles". Two latch clocks are called $\varphi 1$ and φ 2.
- Flip flop can be viewed as, a pair of back to back latches using clk and its complements.
- Table shows delay and timing notations of combinational and sequencing elements. These delays may differ for rising (with suffix 'r') and falling (with suffix 'f').

TERM	*NAME*
T_{pd}	logic propagation delay
T_{cd}	logic contamination delay
T_{pcq}	latch flop clock-Q propagation delay
T_{ccq}	latch flop clock- to Q contamination delay
T_{pdq}	latch flop D –to Q propagation delay
T_{cdq}	latch flop clock D to Q contamination delay
T_{setup}	latch flop setup time
T_{hold}	latch flop hold time

- The delay with timing diagram for all three sequencing elements are, as shown in figure below.
- In combinational logic, input A changing to another value, output Y cannot change instantaneously. After the contamination delay {t_{cd}}, Y may begin to change (or) glitch.
- Output Y settles to a value in propagation delay {t_{pd}} .
- Input D in flip flop must have settled by some setup time {t_{setup}} before the rising edge of clock and should not change again until, a hold time {t_{hold}} after the clock edge.

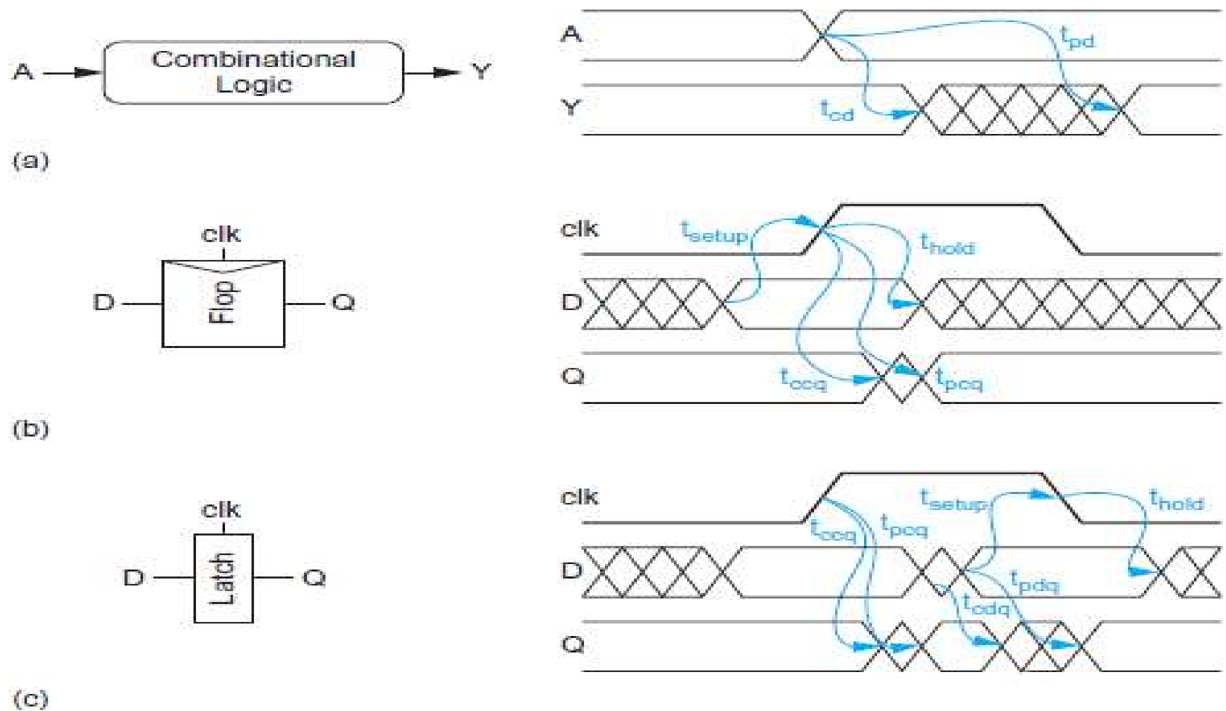
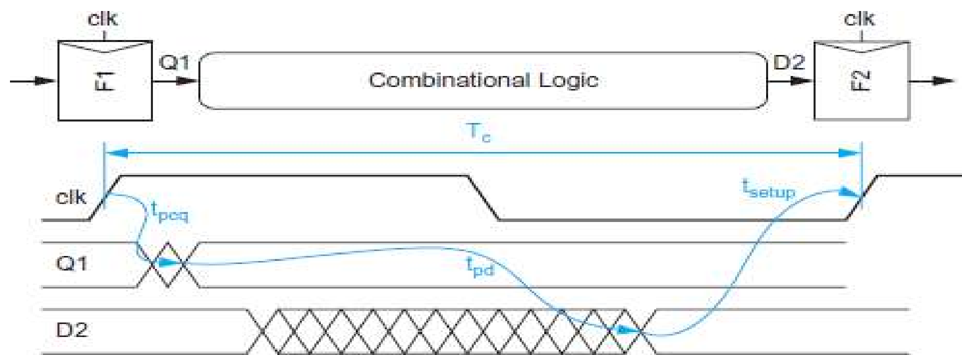


Figure: Timing diagrams

- The output begins to change after a clock-to-contamination delay {t_{ccq}} and completely settles after clock to-Q propagation delay {t_{pcq}}.

(B) Max Delay Constraints:-

- Ideally, the entire clock cycle will be available for computation in the combinational logic.
- If the combination logic delay is too high, the receiving element {next flop/latch} will miss its setup time and sample the wrong value.
- This is called as “setup-time failure” or “max-delay failure”.
- It can be solved by redesigning the logic to be faster (or) by increasing the clock period.



- The clock period must be $Tc \geq tpcq + tpd + tsetup$

(or)

$$Tpd \leq tc - (tsetup + tpcq)$$

Where, tsetup+tpcq – sequencing overhead.

(C) Min-delay constraints:-

- Sequencing elements can be placed back to back without intervening combinational logic and still function correctly.
- If the hold time is large and contamination delay is small, data can incorrectly propagate through successive elements, on one clock edge.
- This corrupt the state of the system called, race condition (or) hold time failure (or) min-delay failure.
- It can be fixed by redesigning the logic and not by slowing the clock.

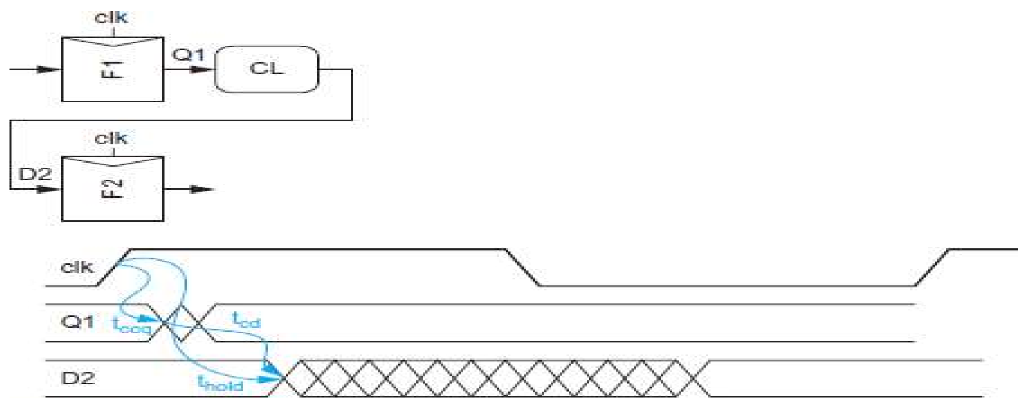


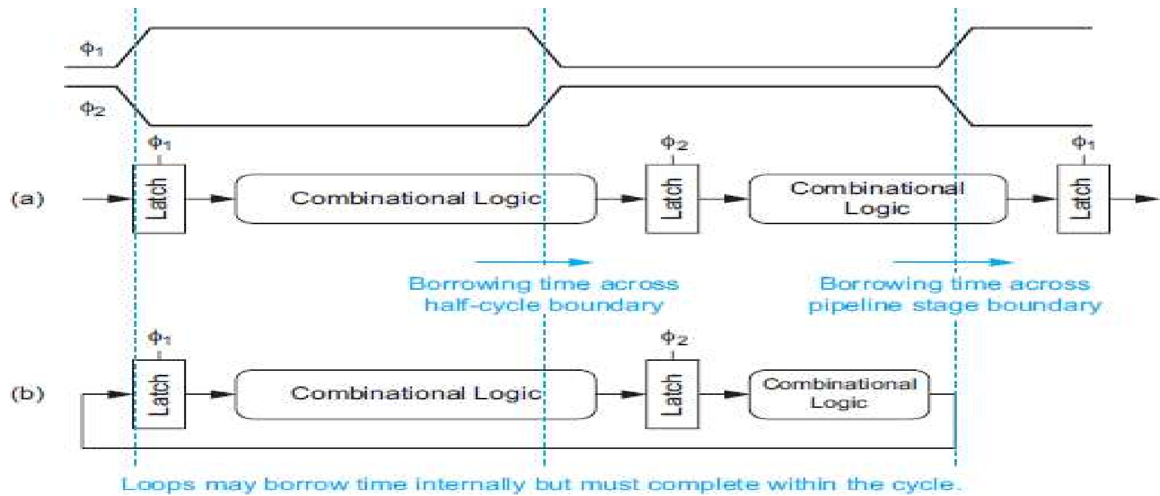
FIGURE Flip-flop latch min-delay constraint

$$tcd \geq thold - tccq$$

(D) Time Borrowing:

- In flip-flop, dada departs the first flip-flop on the rising edge of the clock and must set up at the second flip-flop before the next rising edge of the clock.
- If data arrives late, produces wrong result.
- If data arrives early, it is blocked until the clock edge arrives and the remaining time goes unused and clock imposes a “hard edge”.

- If one half cycle (or) stage of a pipeline has too much logic, it can borrow time in half-cycle (or) stage.
- This is called as "Time borrowing", which can accumulate across multiple cycles.



$$tborrow \leq \frac{Tc}{2} - (tsetup - tnonoverlap)$$

***(E)* Clock Skew**

<u>Analyze the impact of spatial variations of clock signal on edge-triggered sequential logic circuits. (NOV 2018)</u>

- The spatial variation in arrival time of a clock transition in an integrated circuit is referred as clock skew.
- The clock skew between two points i and j on a IC is given by $\delta(i,j)=t_i-t_j$, where t_i and t_j are the position of the rising edge of the clock with respect to a reference.
- The clock skew can be positive or negative, depending upon the routing direction and position of the clock source.
- The timing diagram for the case with positive skew, is shown in figure.
- In the figure, the rising clock edge is delayed by a positive δ at the second register.

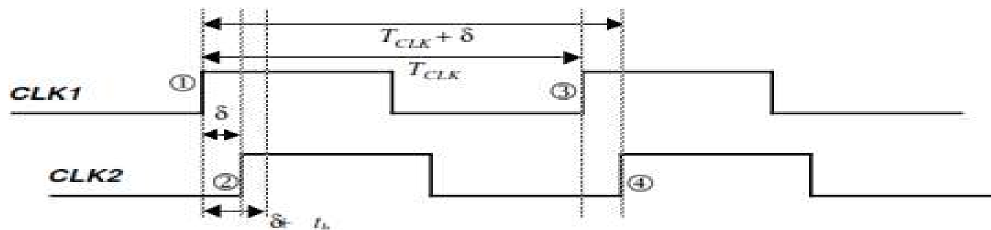


Figure: Timing diagram to study the impact of clock skew on performance and functionality. In this sample timing diagram, δ >0.

(F) Clock Jitter:

- Clock jitter is the temporal variation of the clock period at a given point. The clock period can reduce or expand on a cycle-by-cycle basis. It is a temporal uncertainty measure.
- Cycle-to-cycle jitter refers to time varying deviation of a single clock period.
- For a given spatial location, i is given as $T_{jitter, i}(n) = T_{i,n+1} - T_{i,n} - T_{CLK}$.

Where Ti,n is the clock period for period n, $T_{i,n+1}$ is clock period for period n+1, and T_{CLK} is the nominal clock period.

- Jitter directly impacts the performance of a sequential system.
- Figure shows the nominal clock period as well as variation in period.

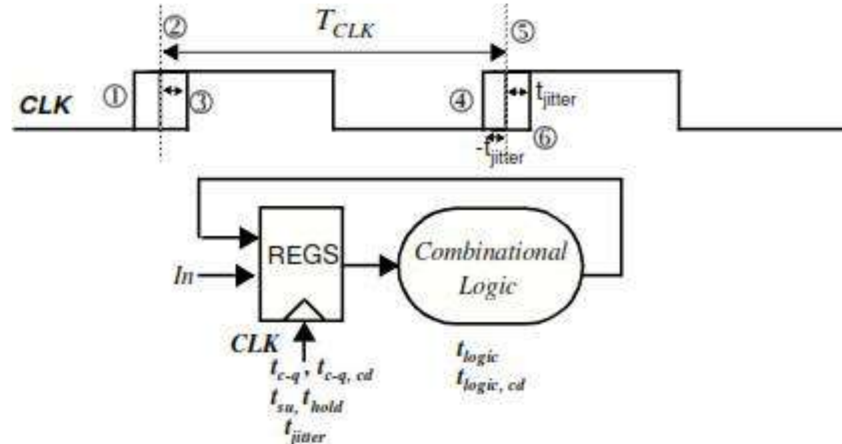


Figure: Circuit for studying the impact of jitter on performance.

****

3.4 Pipelining:

- ❖ **Explain in detail about pipelining structure needed for a logic operation. (April 2017, Nov 2017)**
- ❖ **Discuss in detail various pipelining approaches to optimize sequential circuits. (May 2013, 2016, May 2021)[Apr/May 2022] [April / May 2023]**

- Pipelining is a design technique used to accelerate the operation of the datapaths in digital processors.
- The idea is explained with Figure 3.22a.
- The goal of the circuit is to compute log(|a - b|), where both a and b represent streams of numbers.
- The minimal clock period T_{min} necessary to ensure correct evaluation is given as:

$$T_{\min} = t_{c-q} + t_{pd,\text{logic}} + t_{su}$$

Where, t_{c-q} and t_{su} are the propagation delay and the set-up time of the register respectively.

- Registers are edge-triggered D registers.
- The term $t_{pd,logic}$ stands for the worst-case delay path through the combinatorial network, which consists of the adder, absolute value and logarithm functions.
- In conventional systems, the delay is larger than the delays associated with the registers and dominates the circuit performance.
- Assume that each logic module has an equal propagation delay.
- Each logic module is then, active for only 1/3 of the clock period.
- Pipelining is a technique to improve the resource utilization and increase the functional throughput.
- Introduce registers between the logic blocks, as shown in Figure 3.22b.

- This causes the computation for one set of input data to spread over a number of clock periods, as shown in Table 1.
- The result for the data set (a1, b1) only appears at the output after three clock-periods.

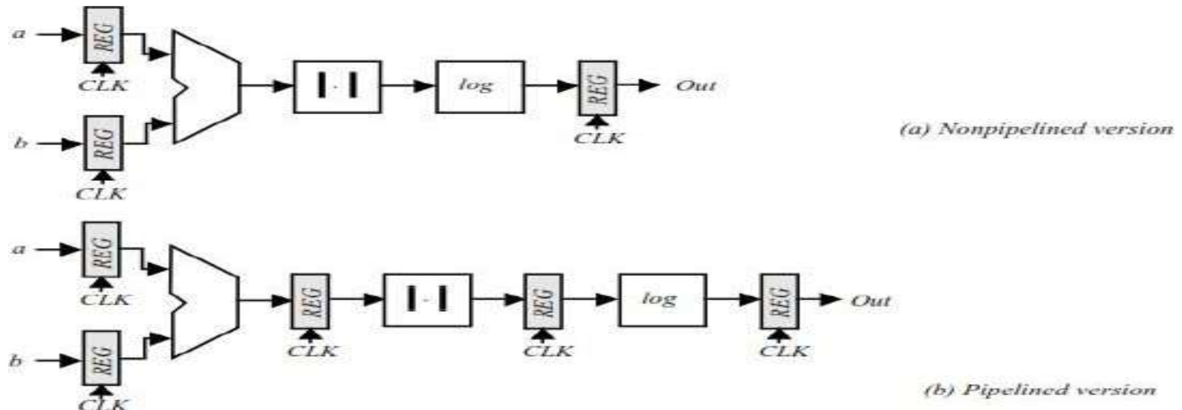


Figure 3.22 Data path for the computation of log(|a + b|).

- At that time, the circuit has already performed parts of the computations for the next data sets, (a2, b2) and (a3, b3).
- The computation is performed in an assembly-line fashion, hence the name pipeline.
- The combinational circuit block has been partitioned into three sections, each of which has a smaller propagation delay than the original function.
- This reduces the value of the minimum allowable clock period:

$$T_{\min,pipe} = t_{c-q} + \max(t_{pd,add}, t_{pd,abs}, t_{pd,\log})$$

- Suppose all logic blocks have same propagation delay and that the register overhead is small with respect to the logic delays.

Clock Period	Adder	Absolute Value	Logarithm
1	$a_1 + b_1$		
2	$a_2 + b_2$	$\lvert a_1 + b_1 \rvert$	
3	$a_3 + b_3$	$\lvert a_2 + b_2 \rvert$	$\log(\lvert a_1 + b_1 \rvert)$
4	$a_4 + b_4$	$\lvert a_3 + b_3 \rvert$	$\log(\lvert a_2 + b_2 \rvert)$
5	$a_5 + b_5$	$\lvert a_4 + b_4 \rvert$	$\log(\lvert a_3 + b_3 \rvert)$

Table 1: Example of pipelined computations.

- The pipelined network performs the original circuit by a factor of three, under these assumptions $T_{min,pipe}=T_{min}/3$.
- The increased performance comes at the relatively small cost of two additional registers, and an increased latency.
- Pipelining is implemented for very high-performance datapaths.

3.4.1 NORA-CMOS—A Logic Style for Pipelined Structures:

Discuss about the NORA–CMOS structure. (Nov 2016)

- The latch-based pipeline circuit can also be implemented using C^2MOS latches, as shown in Figure 3.24.
- This topology has one additional property:
  - *A C^2MOS-based pipelined circuit is race-free as long as all the logic functions F (implemented using static logic) between the latches are noninverting.*

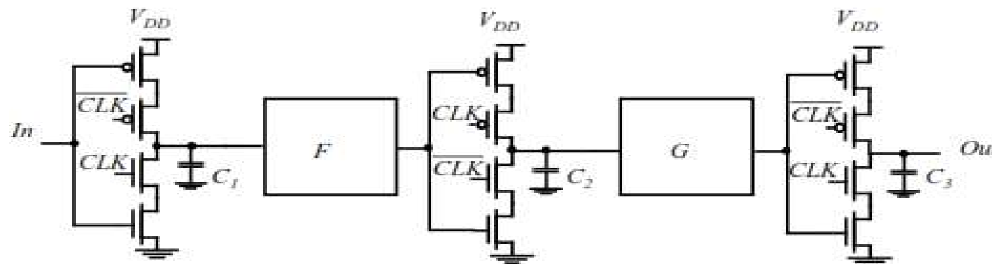


Figure 3.24 Pipelined datapath using C^2MOS latches

- The only way, a signal can race from stage to stage under this condition is, when the logic function F is inverting, as in Figure 3.25.
- Here F is replaced by a single, static CMOS inverter. Similar considerations are valid for the (1-1) overlap.
- It combines C^2MOS pipeline registers and NORA dynamic logic function blocks.
- Each module consists of a block of combinational logic, that can be a mixture of static and dynamic logic, followed by a C^2MOS latch.

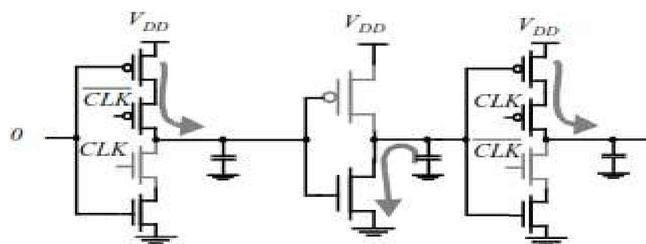


Figure 3.25 Potential race condition during (0-0) overlap in C^2MOS-based design.

- Logic and latch are clocked, in such a way that both are simultaneously in either evaluation or hold (precharge) mode.
- A block that is, in evaluation during CLK = 1 is called a CLK-module, while the inverse is called a $\overline{CLK}$ -module.
- The operation modes of the modules are summarized in Table 2.

	CLK block		$\overline{CLK}$ block	
	Logic	**Latch**	**Logic**	**Latch**
$CLK = 0$	Precharge	Hold	Evaluate	Evaluate
$CLK = 1$	Evaluate	Evaluate	Precharge	Hold

Table 2: Operation modes for NORA logic modules.

3.4.2 Latch- vs. Register-Based Pipelines:

Compare Latch and register based pipelines.

- Pipelined circuits can be constructed using level-sensitive latches instead of edge-triggered registers.
- Consider the pipelined circuit of Figure 3.23.
- The pipeline system is implemented based on pass-transistor-based positive and negative latches instead of edge triggered registers.
- Here logic is introduced between the master and slave latches of a master-slave system.
- When the clocks CLK and $\overline{CLK}$ are non-overlapping, correct pipeline operation is obtained.
- Input data is sampled on C_1 at the negative edge of CLK and the computation of logic block F starts.
- The result of the logic block F is stored on C_2 on the falling edge of $\overline{CLK}$ and the computation of logic block G starts.
- The non-overlapping of the clocks ensures correct operation.
- The value stored on C_2 at the end of the CLK low phase, is the result of passing the previous input through the logic function F.
- When overlap exists between CLK and $\overline{CLK}$, the next input is already being applied to F, and its effect might propagate to C_2 before $\overline{CLK}$ goes low.

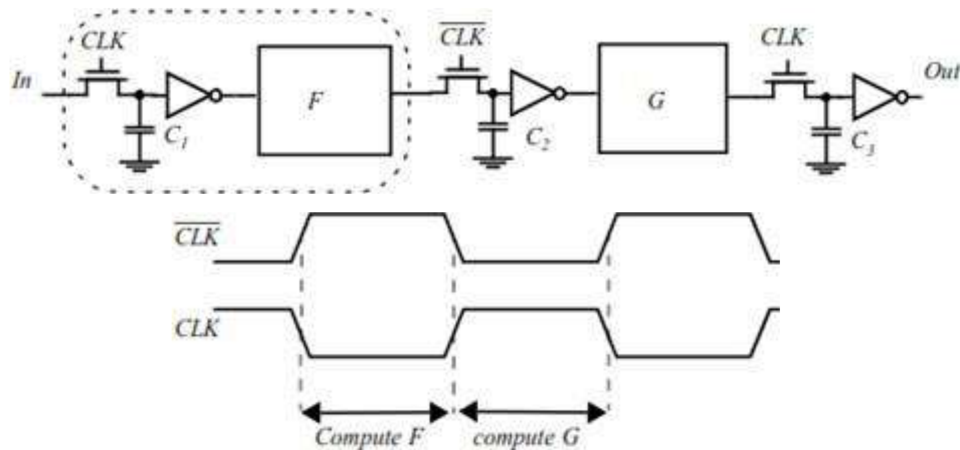


Figure 3.23 Operation of two-phase pipelined circuit using dynamic registers.

****

3.5 Choosing a Clocking Strategy:

Discuss about strategy required for choosing a clock signal.

- Choosing the right clocking scheme affects the functionality, speed and power of a circuit.
- The simple clocking scheme is the two-phase master-slave design.
- The predominant approach is use the multiplexer-based register and to generate the two clock phases locally, by simply inverting the clock.
- High-performance CMOS VLSI design is using simple clocking schemes, even at the expense of performance.

3.5.1 Static sequencing element methodology:-

- Many issues are related to static sequencing element methodology.

(a) Choice of element:-

Flip-flop: Flip-flop has high sequencing overhead. It is simple and easy to understand the operation of flip-flop.

Pulsed latches:-

- Faster than flip-flop.
- Provides some time borrowing option.
- Consumes law power.

Transparent Latch:-

- It has low sequencing overhead compared with flip-flop.
- It allows almost half cycle of time borrowing and it is good choice.

(b) Low power sequential design:-

- Pulsed latches are power efficient.
- Flip-flop consumes more power.
- Clock gating can be used to reduce power.

(c) Two-phase Timing types:-In this type, the signal can belong to phase 1 (or) phase 2. In each phase, 3 different clocks are stable, valid and qualified clock.

Stable clock:-

A signal is stable (in$\varphi 1$), if it settles to a value before rises and remains constant until after, $\varphi 1$falls.

Valid clock: - A signal is valid (in$\varphi 1$), if it settles to a value before $\varphi 1$falls and remains at that value after $\varphi 1$ falls.

Qualified signal: -A signal is said to be in qualified clock in$\varphi 1$, if it either rises and falls like $\varphi 1$(or) remains low for the entire cycle.

****

3.6 Synchronous and Asynchronous circuits-Timing issues:

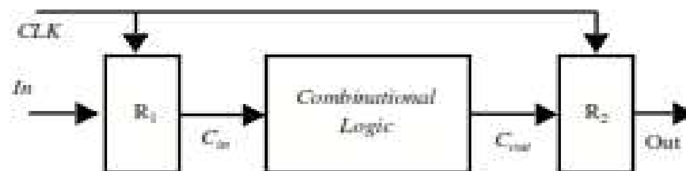
- A synchronous system approach is one, in which all memory elements in the system are simultaneously updated using a globally distributed periodic synchronization signal.
- Functionality is ensured by imposing some strict constraints on the generation of the clock signals and their distribution to the memory elements.
- Analyze the impact of spatial variations of the clock signal, called clock skew and temporal variations of the clock signal, called clock jitter.
- Asynchronous design avoids the problem of clock uncertainty by eliminating the need for globally-distributed clocks.
- The important issues of synchronization, which is required when interfacing different clock domains or when sampling an asynchronous signal.

Classification of Digital Systems:

- In digital systems, signals can be classified depending on, how they are related to a local clock.
- Signals that transition only at predetermined periods in time can be classified as synchronous, mesochronous and plesiochronous with respect to a system clock.
- A signal that can transition at arbitrary times is considered asynchronous.

3.6.1 Synchronous Interconnect

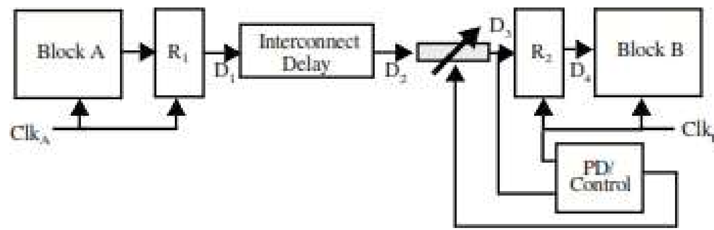
- A synchronous signal has the exact same frequency and a known fixed phase offset with respect to the local clock.
- In such a timing methodology, the signal is "synchronized" with the clock and the data can be sampled directly without any uncertainty.
- In digital logic design, synchronous systems are straight forward type of interconnect, where the flow of data in a circuit proceeds with the system clock as shown below.



3.6.2 Mesochronous interconnect:

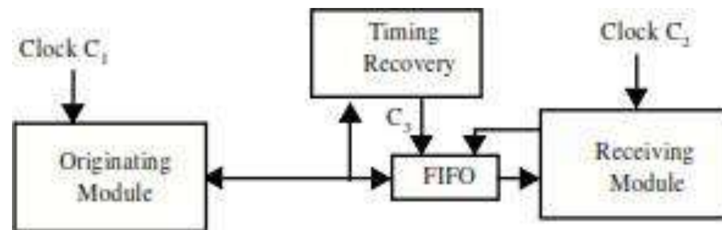
- A mesochronous signal has the same frequency but an unknown phase offset with respect to the local clock ("meso" from Greek is middle).
- For example, if data is being passed between two different clock domains, then the data signal transmitted from the first module can have an unknown phase relationship to the clock of the receiving module.
- In such a system, it is not possible to directly sample the output at the receiving module because of the uncertainty in the phase offset.
- A (mesochronous) synchronizer can be used to synchronize the data signal with the receiving clock as shown below.

- The synchronizer serves to adjust the phase of the received signal to ensure proper sampling.



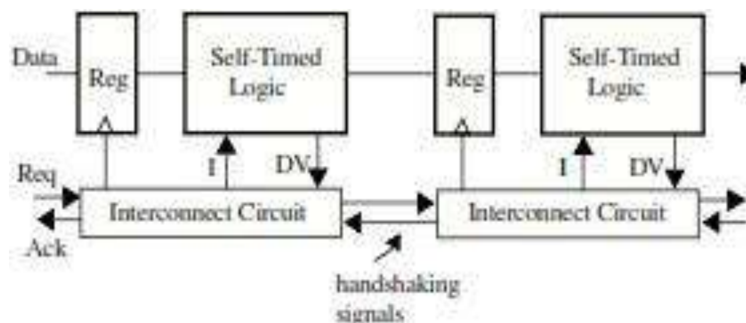
3.6.3 Plesiochronous Interconnect

- A plesiochronous signal has nominally the same, but slightly different frequency as the local clock (“plesio” from Greek is near).
- This scenario can easily arise when two interacting modules have independent clocks generated from separate crystal oscillators.
- Since the transmitted signal can arrive at the receiving module at a different rate than the local clock, one need to utilize a buffering scheme to ensure, all data is received.
- A possible framework for plesiochronous interconnect is shown in Figure below.



3.6.4 Asynchronous Interconnect:

- Asynchronous signals can transition at any arbitrary time and are not slaved to any local clock.
- As a result, it is not to map these arbitrary transitions into a synchronized data stream.
- Asynchronous signals are used to eliminate the use of local clocks and utilize a self-timed asynchronous design approach.
- In this approach, communication between modules is controlled through a handshaking protocol.



****

3.7 Clock-Distribution Techniques

- ❖ **Explain the clock distribution techniques in synchronous design in detail. (Nov 2017)**
- ❖ **Design a clock distribution network based on H tree model for 16 nodes. (April 2018)**

- Clock skew and jitter are major issues in digital circuits and they limit the performance of a digital system.
- It is necessary to design a clock network, that minimizes skew and jitter.
- Another important consideration in clock distribution is the power dissipation.
- In most high-speed digital processors, a majority of the power is dissipated in the clock network.
- To reduce power dissipation, clock networks must support clock conditioning, the ability to shut down parts of the clock network.
- Unfortunately, clock gating results in additional clock uncertainty.

Fabrics for clocking:

- Clock networks include a network that is used to distribute a global reference to various parts of the chip.
- A final stage is responsible for local distribution of the clock, while considering the local load variations.
- Most clock distribution schemes use the absolute delay from a central clock source to the clocking elements.
- Therefore one common approach to distributing a clock is, to use balanced paths (or called trees).
- The most common type of clock primitive is, the H-tree network (named for the physical structure of the network) in figure, where a 4x4 array is shown.
- In this scheme, the clock is routed to a central point on the chip and balanced paths.
- Include both matched interconnect as well as buffers, are used to distribute the reference to various leaf nodes.
- If each path is balanced, the clock skew is zero. It takes multiple clock cycles for a signal to propagate from the central point to each leaf node. The arrival times are equal at every leaf node.
- The H-tree configuration is particularly useful for regular-array networks, in which all elements are identical and the clock can be distributed as a binary tree.

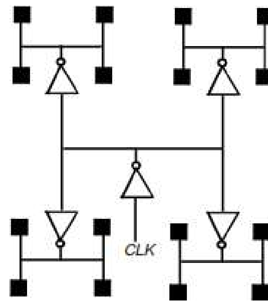


Figure: Example of an H-tree clock-distribution network for 16 leaf nodes.

Latch-Based Clocking:

- The use of a latch based methodology (in Figure) enables more flexible timing, allowing one stage to pass slack to or steal time from following stages.
- This flexibility allows an overall performance increase.
- In this configuration, a stable input is available to the combinational logic block A (CLB_A) on the falling edge of CLK1 (at edge2).
- On the falling edge of CLK2 (at edge3), the output CLB_A is latched and the computation of CLK_B is launched.
- CLB_B computes on the low phase of CLK2 and the output is available on the falling edge of CLK1 (at edge4).
- This timing appears, equivalent to having an edge-triggered system where CLB_A and CLB_B are cascaded and between two edge-triggered registers.
- In both cases, it appears that the time available to perform the combination of CLB_A and CLB_B are T_{CLK}.

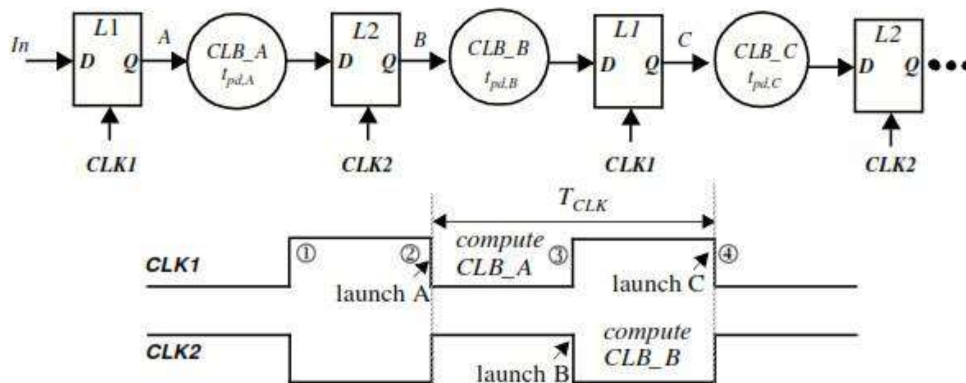


Figure: Latch-based design in which transparent latches are separated by combinational logic.

****

3.8 Self-Timed Circuit Design

Discuss about asynchronous design in logic design.

3.8.1 Self-Timed Logic : An Asynchronous Technique

- A more reliable and robust technique is the self-timed approach, which presents a local solution to the timing problem.
- Figure uses a pipelined datapath to illustrate how this can be accomplished.
- The computation of a logic block is initiated by asserting a Start signal.
- The combinational logic block computes on the input data.
- This signaling ensures the logical ordering of the events and can be achieved with the aid of an extra Ack(nowledge) and Req(uest) signal.

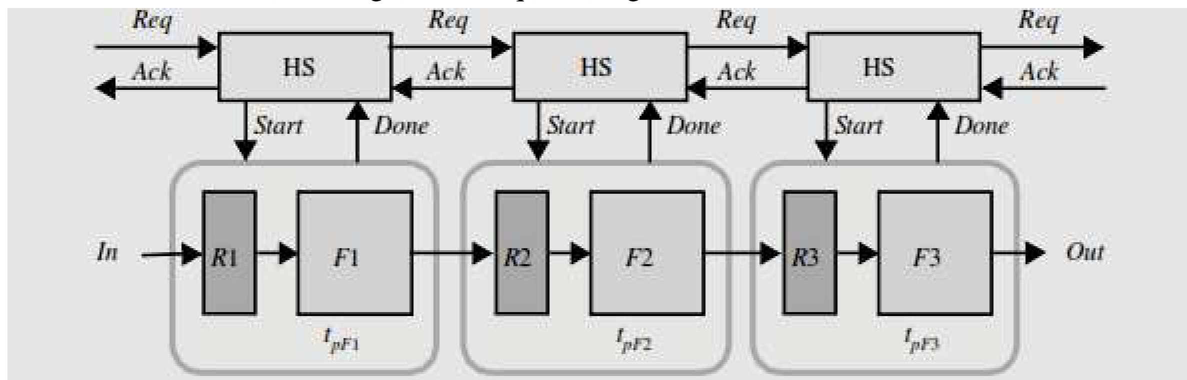


Figure: Self-timed, pipelined datapath.

In the case of the pipelined datapath, the scenario could proceed as follows.

1. An input word arrives, and a Req(uest) to the block F1 is raised. If F1 is inactive at that time, it transfers the data and acknowledges this fact to the input buffer.
2. F1 is enabled by raising the Start signal. After a certain amount of time, dependent upon the data values, the Done signal goes high indicating the completion of the computation.
3. A Req(uest) is issued to the F2 module. If this function is free, an Ack(nowledge) is raised, the output value is transferred and F1 can go ahead with its next computation.

3.8.2 A simple synchronizer

How do eliminates metastability problem in sequential circuit and explain?

- A synchronizer accepts an input D and a clock ϕ. It produces an output Q that should be valid for some bounded delay after the clock.
- The synchronizer has an aperture, defined by a setup and hold time around the rising edge of the clock.

- If the data is stable during the aperture, Q should equal D. If the data changes during the aperture, Q can be chosen arbitrarily.

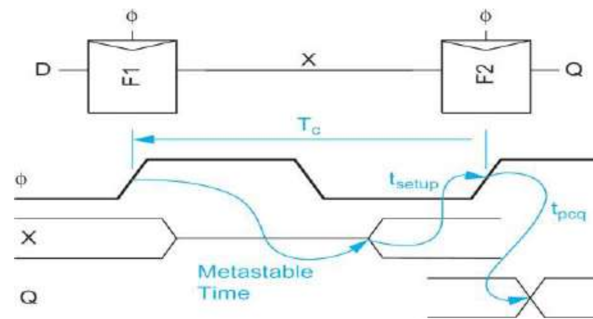


Figure: Simple Synchronizer

- Figure shows a simple synchronizer built from a pair of flip-flops. F1 samples the asynchronous input D.
- The output X may be metastable for some time, but will settle to a good level with high probability, if we wait long enough.
- F2 samples X and produce an output Q, that should be a valid logic level and be aligned with the clock.
- The synchronizer has a latency of one clock cycle Tc .

3.8.3 Communicating between asynchronous clock domains

- A common application of synchronizers is in communication between asynchronous clock domains, i.e., blocks of circuits that do not share a common clock.
- Suppose System A is controlled by clkA that needs to transmit N-bit data words to System B, which is controlled by clkB, as shown in Figure.
- The systems can represent separate chips or separate units within a chip using unrelated clocks.
- System A must guarantee that the data is stable, while the flip-flops in System B sample the word.
- It indicates when new data is valid by using a request signal (Req), so System B receives the word exactly once rather than zero or multiple times.
- System B replies with an acknowledge signal (Ack), when it has sampled the data, so System A knows when the data can safely be changed.
- If the relationship between clkA and clkB is completely unknown, a synchronizer is required at the interface.

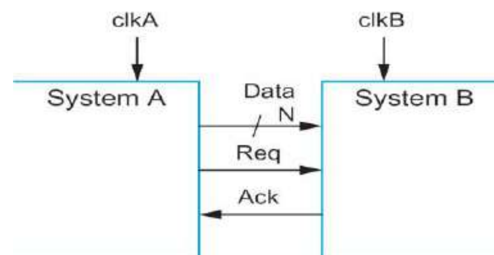


Figure: Communication between asynchronous systems

3.8.4 Arbiter

- The arbiter of Figure (a) is related to the synchronizer. It determines which of two inputs arrived first.
- If the spacing between the inputs exceeds some aperture time, the first input should be acknowledged.
- If the spacing is smaller, exactly one of the two inputs should be acknowledged, but the choice is arbitrary.
- For example, in a television game show, two contestants may hit buttons to answer a question.
- If one presses the button first, should be acknowledged. If both presses the button at times too close to distinguishes, the host may choose one of the two contestants arbitrarily.

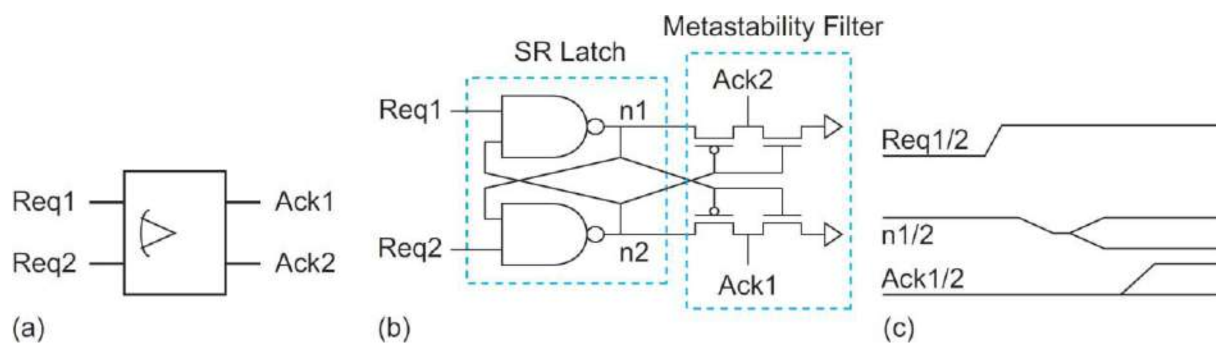


Figure: Arbiter

- Figure (b) shows an arbiter built from an SR latch and a four-transistor metastability filter.
- If one of the request inputs arrives well before the other, the latch will respond appropriately.
- If they arrive at nearly the same time, the latch may be driven into metastability, as shown in Figure (c).
- The filter keeps both acknowledge signals low, until the voltage difference between the internal nodes n1 and n2 exceeds Vt , indicating that a decision has been made.
- Such an asynchronous arbiter will never produce metastable outputs.

3.8.5 Synchronous versus Asynchronous Design:

Compare synchronous and asynchronous design.

- The self-timed approach offers a potential solution to the growing clock-distribution problem.
- It translates the global clock signal into a number of local synchronization problems.

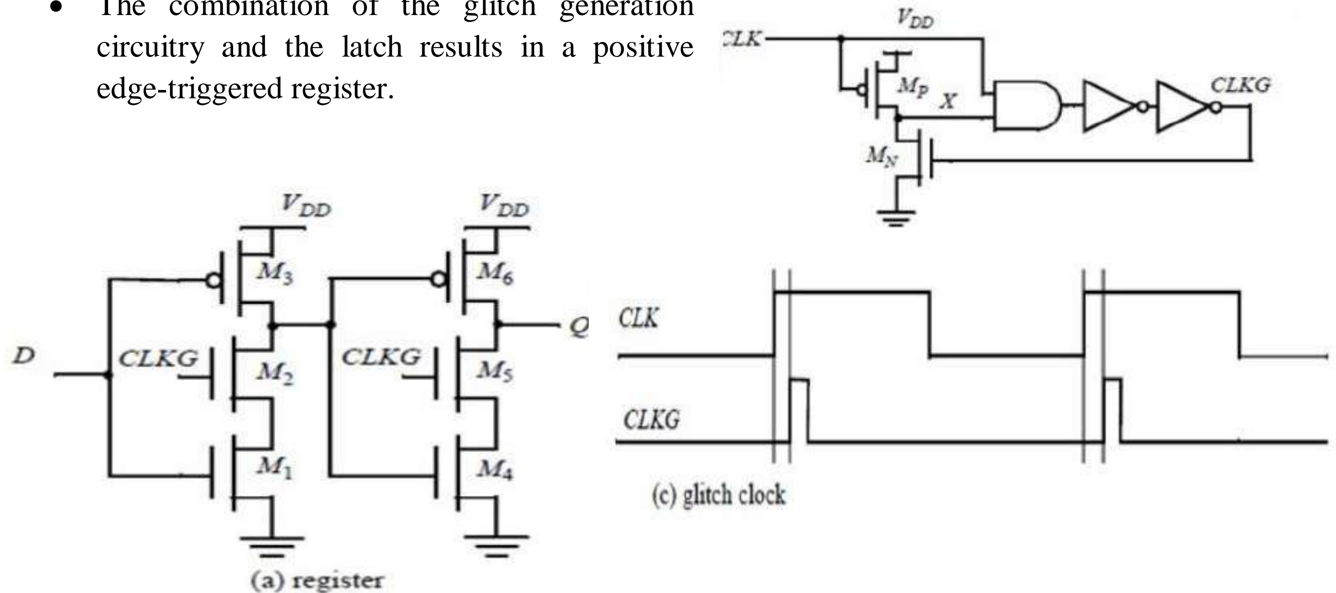
- Handshaking logic is needed to ensure the logical ordering of the circuit events and to avoid race conditions.
- In general, synchronous logic is both faster and simpler since the overhead of completion-signal generation and handshaking logic is avoided.
- Skew management requires extensive modeling and analysis, as well as careful design.
- It will not be easy to extend this methodology into the next generation of designs.
- This observation is already reflected in the fact that the routing network for the latest generation of massively parallel supercomputers is completely implemented using self-timing.
- For self timing to become a mainstream design technique however (if it ever will), further innovations in circuit and signaling techniques and design methodologies are needed.

****

3.9 Pulse Registers

Design the pulse registers suitable for sequential CMOS circuits. [May 2021][Nov/Dec 2022]

- A fundamentally different approach for constructing a register uses pulse signals.
- The idea is to construct a short pulse around the rising (or falling) edge of the clock.
- This pulse acts as the clock input to a latch, sampling the input only in a short window.
- Race conditions are thus avoided by keeping the opening time (i.e, the transparent period) of the latch very short.
- The combination of the glitch generation circuitry and the latch results in a positive edge-triggered register.



(a) register

(c) glitch clock

- **Figure b shows** an example circuit for constructing a short intentional glitch on each rising edge of the clock.
- When CLK = 0, node X is charged up to VDD (MN is off since CLKG is low).
- On the rising edge of the clock, there is a short period of time when both inputs of the AND gate are high, causing CLKG to go high.

- This in turn activates M_N, pulling X and eventually CLKG low.
- The length of the pulse is controlled by the delay of the AND gate and the two inverters.
- Note that there exists also a delay between the rising edges of the input clock (CLK) and the glitch clock (CLKG) — also equal to the delay of the AND gate and the two inverters.
- If every register on the chip uses the same clock generation mechanism, this sampling delay does not matter.
- However, process variations and load variations may cause the delays through the glitch clock circuitry to be different.
- This must be taken into account when performing timing verification and clock skew analysis.

Waveform

- If set-up time and hold time are measured in reference to the rising edge of the glitch clock, the set-up time is essentially zero, the hold time is equal to the length of the pulse (if the contamination delay is zero for the gates), and the propagation delay (tc-q) equals two gate delays.

Advantage

- The reduced clock load and the small number of transistors required.
- The glitch-generation circuitry can be amortized over multiple register bits.

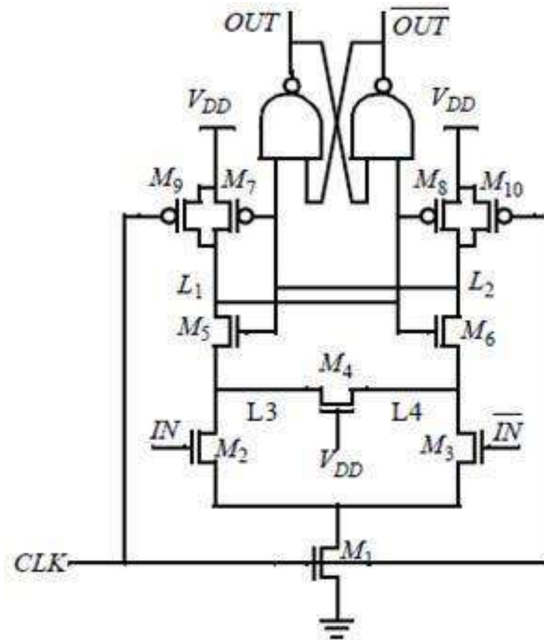
Disadvantage

- A substantial increase in verification complexity.
- This has prevented a wide-spread use.

3.10 Sense-Amplifier Based Registers

Write short notes on Sense – Amplifier Based Registers. [Nov/Dec 2022][April/May 2023]

- A sense amplifier structure to implement an edge- triggered register.
- Sense amplifier circuits accept small input signals and amplify them to generate rail-to-rail swings.
- There are many techniques to construct these amplifiers, with the use of feedback (e.g., cross-coupled inverters).



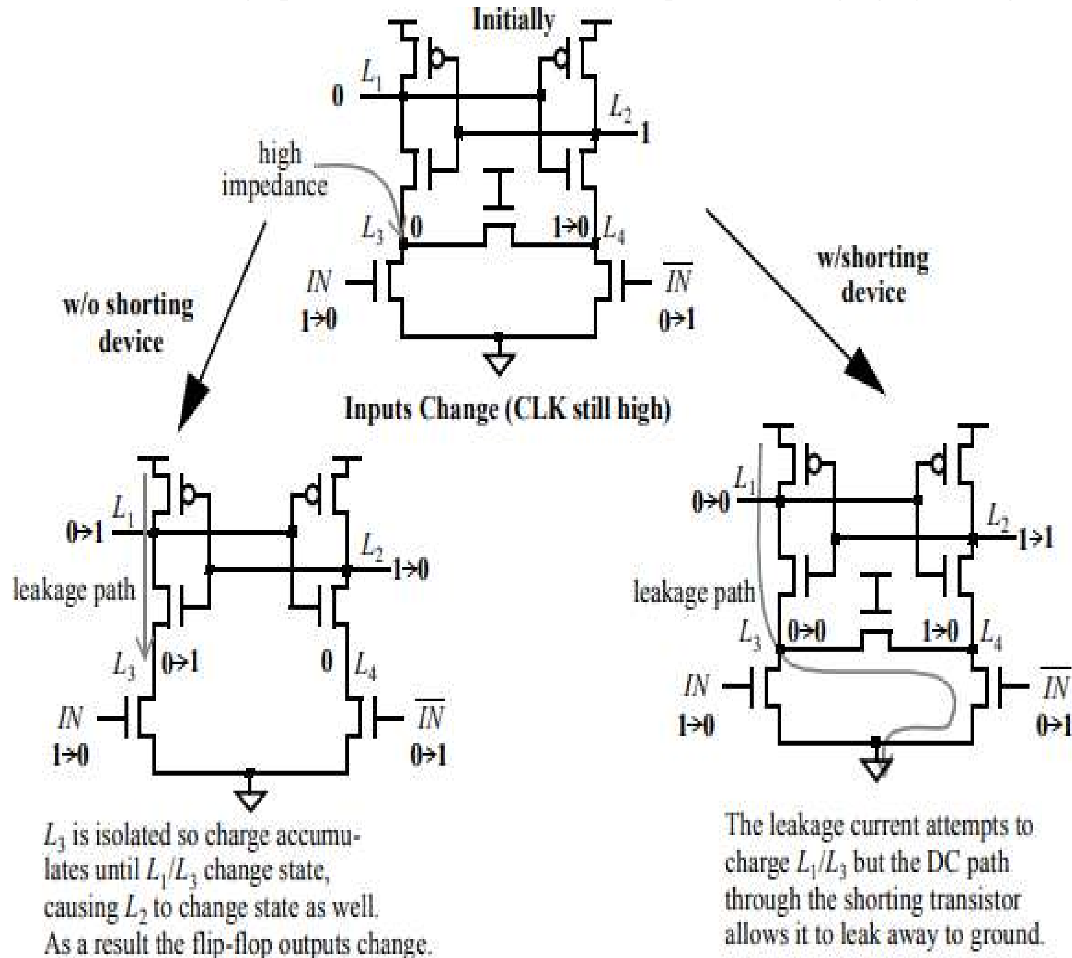
Positive edge-triggered register based on sense-amplifier

- The circuit uses a precharged front-end amplifier that samples the differential input signal on the rising edge of the clock signal.
- The outputs of front-end are fed into a NAND cross- coupled SR FF that holds the data and guarantees that the differential outputs switch only once per clock cycle.
- The differential inputs in this implementation don't have to have rail-to-rail swing and hence this register can be used as a receiver for a reduced swing differential bus.

Operation

- The core of the front-end consists of a cross-coupled inverter (M5-M8) whose outputs (L1 and L2) are precharged using devices M9 and M10 during the low phase of the clock.
- As a result, PMOS transistors M7 and M8 to be turned off and the NAND FF is holding its previous state.
- Transistor M1 is similar to an evaluate switch in dynamic circuits and is turned off ensuring that the differential inputs don't affect the output during the low phase of the clock.
- On the rising edge of the clock, the evaluate transistor turns on and the differential input pair (M2 and M3) is enabled, and the difference between the input signals is amplified on the output nodes on L1 and L2.
- The cross-coupled inverter pair flips to one of its the stable states based on the value of the inputs.
- For example, if IN is 1, L1 is pulled to 0, and L2 remains at VDD. Due to the amplifying properties of the input stage, it is not necessary for the input to swing all the way up to VDD and enables the use of low swing signaling on the input wires.

- The shorting transistor, M4, is used to provide a DC leakage path from either node L3, or L4, to ground.
- This is necessary to accommodate the case where the inputs change their value after the positive edge of CLK has occurred, resulting in either L3 or L4 being left in a high-impedance state with a logical low voltage level stored on the node.
- Without the leakage path that node would be susceptible to charging by leakage currents.



3.11 Schmitt Trigger

Explain the circuit and working of CMOS implementation of Schmitt Trigger. (NOV 2021) [Nov/Dec 2022]

- A Schmitt trigger is a device with two important properties:

  1. It responds to a slowly changing input waveform with a fast transition time at the output.
  2. The voltage-transfer characteristic of the device displays different switching thresholds for positive- and negative-going input signals.

A typical voltage-transfer characteristic of the Schmitt trigger is shown (and its schematics symbol). The switching thresholds for the low-to-high and high to-low transitions are called V_{M+} and V_{M-}, respectively. The hysteresis voltage is defined as the difference between the two.

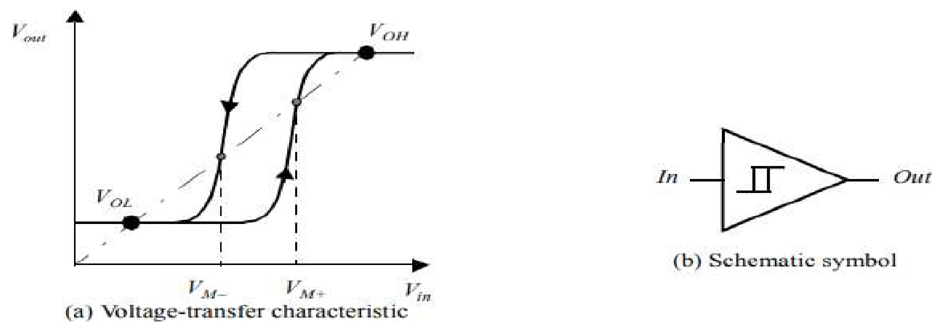
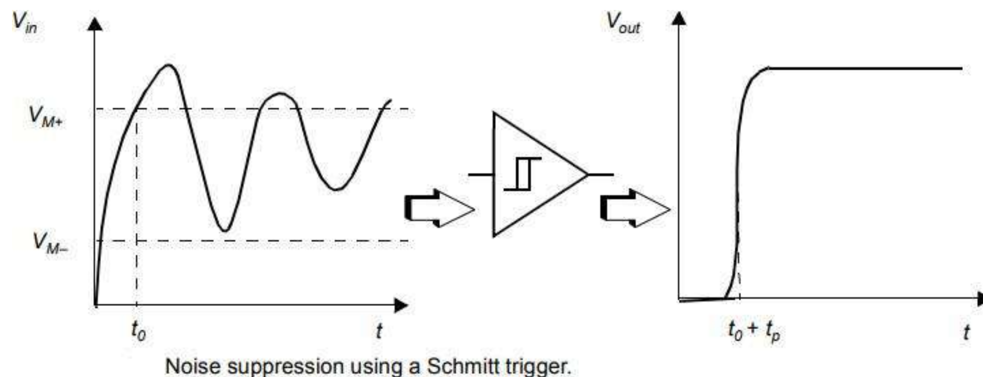


Fig: Non-inverting Schmitt trigger.

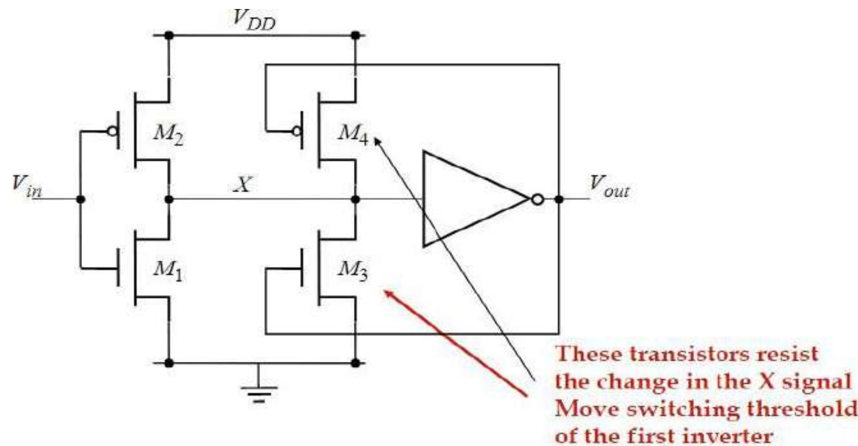
- One of the main uses of the Schmitt trigger is to turn a noisy or slowly varying input signal into a clean digital output signal.
- Notice how the hysteresis suppresses the ringing on the signal.
- At the same time, the fast low-to-high (and high-to-low) transitions of the output signal should be observed.
- For instance, steep signal slopes are beneficial in reducing power consumption by suppressing direct-path currents.
- The “secret” behind the Schmitt trigger concept is the use of positive feedback.



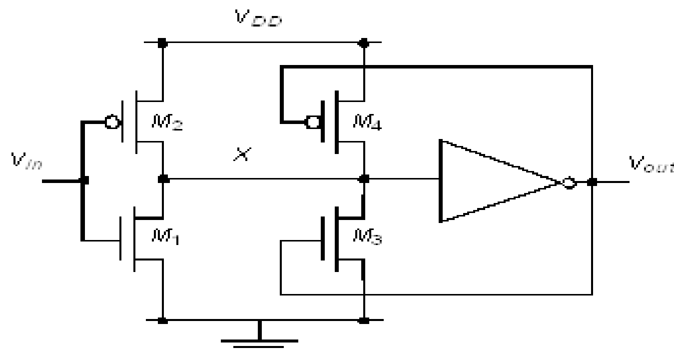
Noise suppression using a Schmitt trigger.

CMOS Implementation

- Increasing k_n/k_p ratio decreases the logical switching threshold
- If V_{in} =0 the V_{out} (connected to M_4) is also zero So effectively the input is connected to M_2 and M_4 in parallel This increases k_p and the switching threshold



- If V $_{in}$ =0 the situation is reversed and k_n increases reducing the switching threshold



- Suppose that V_{in} is initially equal to 0, so that $V_{out} = 0$ as well.
- The feedback loop biases the PMOS transistor M_4 in the conductive mode while M_3 is off.
- The input signal effectively connects to an inverter consisting of two PMOS transistors in parallel (M_2 and M_4) as a pull-up network, and a single NMOS transistor (M_1) in the pull-down chain.
- This modifies the effective transistor ratio of the inverter to $k_{M1}/(k_{M2}+k_{M4})$, which moves the switching threshold upwards.
- Once the inverter switches, the feedback loop turns off M_4, and the NMOS device M_3 is activated.
- This extra pull-down device speeds up the transition and produces a clean output signal with steep slopes.
- A similar behavior can be observed for the high-to-low transition.
- In this case, the pull-down network originally consists of M_1 and M_3 in parallel, while the pull-up network is formed by M_2.
- This reduces the value of the switching threshold to V_{M-}.

3.12 Multivibrator Circuits

Explain about Monostable and Astable Sequential Circuits. [Apr/May 2022] [Nov/Dec 2022] Sketch and explain the Monostable sequential circuits based on CMOS logic. (May 2021) [April / May 2023]

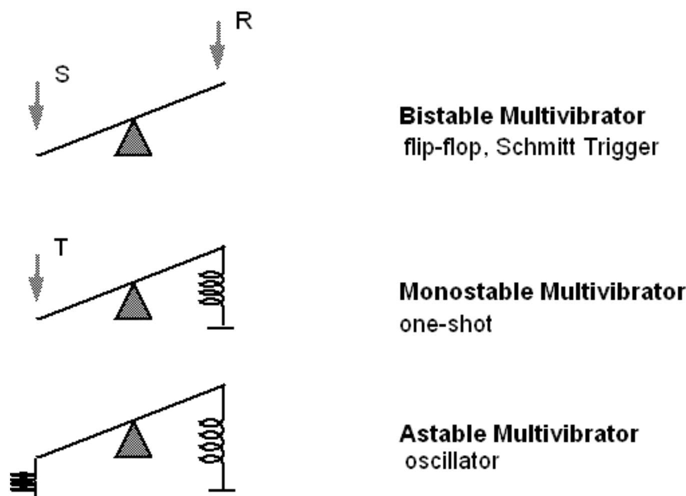
Sequential circuits are of three types −

Bistable − Bistable circuits have two stable operating points and will be in either of the states. Example − Memory cells, latches, flip-flops and registers.

Monostable − Monostable circuits have only one stable operating point and even if they are temporarily perturbed to the opposite state, they will return in time to their stable operating point. Example: Timers, pulse generators.

Astable − circuits have no stable operating point and oscillate between several states.

Example − Ring oscillator.



Transition-Triggered Monostable Sequential Circuits

- A monostable element is a circuit that generates a pulse of a predetermined width every time the quiescent circuit is triggered by a pulse or transition event.
- It is called monostable because it has only one stable state (the quiescent one).
- A trigger event, which is either a signal transition or a pulse, causes the circuit to go temporarily into another quasi-stable state.
- This means that it eventually returns to its original state after a time period determined by the circuit parameters.
- This circuit, also called a one-shot, is useful in generating pulses of a known length.

- This functionality is required in a wide range of applications.
- We have already seen the use of a one-shot in the construction of glitch registers.
- Another notorious example is the address transition detection (ATD) circuit, used for the timing generation in static memories.
- This circuit detects a change in a signal, or group of signals, such as the address or data bus, and produces a pulse to initialize the subsequent circuitry.
- The most common approach to the implementation of one-shots is the use of a simple delay element to control the duration of the pulse.
- In the quiescent state, both inputs to the XOR are identical, and the output is low.
- A transition on the input causes the XOR inputs to differ temporarily and the output to go high.
- After a delay t_d (of the delay element), this disruption is removed, and the output goes low again.
- A pulse of length td is created.
- The delay circuit can be realized in many different ways, such as an RC-network or a chain of basic gates.

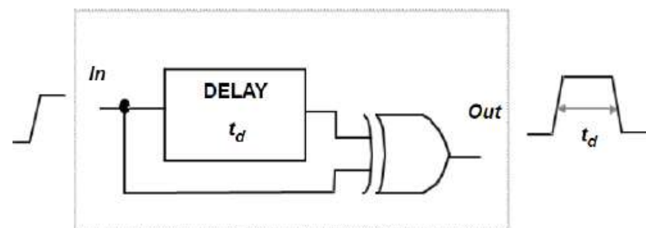


Fig: Transition-triggered one-shot.

3.13 Astable Sequential Circuits

- An astable circuit has no stable states.
- The output oscillates back and forth between two quasi-stable states with a period determined by the circuit topology and parameters (delay, power supply, etc.).
- One of the main applications of oscillators is the on-chip generation of clock signals.
- This application is discussed in detail in a later chapter (on timing). The ring oscillator is a simple, example of an astable circuit.
- It consists of an odd number of inverters connected in a circular chain.
- Due to the odd number of inversions, no stable operation point exists, and the circuit oscillates with a period equal to $2 \times t_p \times N$, with N the number of inverters in the chain and tp the propagation delay of each inverter.
- The ring oscillator composed of cascaded inverters produces a waveform with a fixed oscillating frequency determined by the delay of an inverter in the CMOS process.

- In many applications, it is necessary to control the frequency of the oscillator.
- An example of such a circuit is the voltage-controlled oscillator (VCO), whose oscillation frequency is a function (typically non-linear) of a control voltage.
- The standard ring oscillator can be modified into a VCO by replacing the standard inverter with a current-starved inverter as shown in Figure.
- The mechanism for controlling the delay of each inverter is to limit the current available to discharge the load capacitance of the gate.
- In this modified inverter circuit, the maximal discharge current of the inverter is limited by adding an extra series device.
- Note that the low-to-high transition on the inverter can also be controlled by adding a PMOS device in series with M2.
- The added NMOS transistor M3, is controlled by an analog control voltage Vcntl, which determines the available discharge current.
- Lowering Vcntl reduces the discharge current and, hence, increases tpHL.
- The ability to alter the propagation delay per stage allows us to control the frequency of the ring structure.
- The control voltage is generally set using feedback techniques.
- Under low operating current levels, the current-starved inverter suffers from slow fall times at its output. This can result in significant short-circuit current.
- This is resolved by feeding its output into a CMOS inverter or better yet a Schmitt trigger.
- An extra inverter is needed at the end to ensure that the structure oscillates.

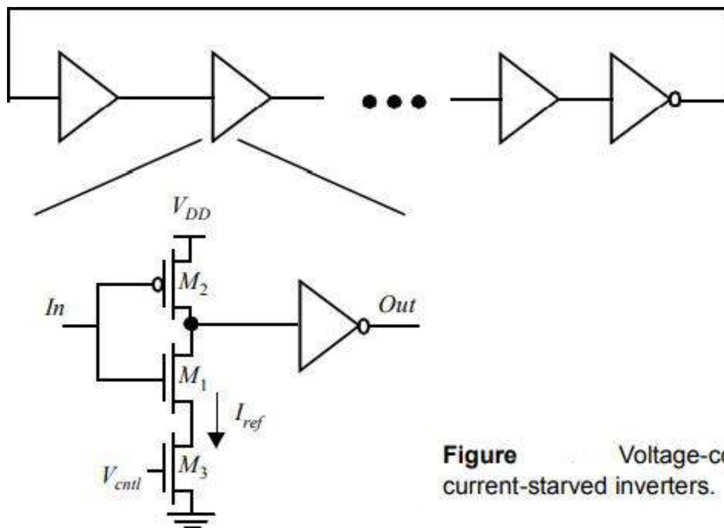
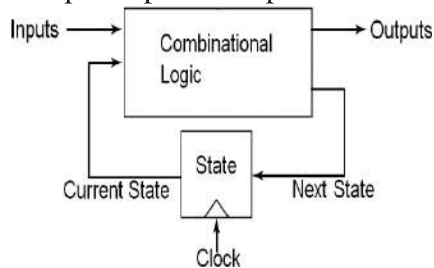


Figure Voltage-controlled oscillator based on current-starved inverters.

UNIT – III
SEQUENTIAL LOGIC CIRCUITS AND CLOCKING STRATEGIES

1. What is a Sequential circuit?

In sequential circuits, the output depends on previous as well as current inputs.



2. What are sequencing methods available in sequential circuit design? (NOV/DEC-2012)

The sequencing methods available in sequential circuit design are,

i. Flipflop
ii. Two phase transparent latches
iii. Pulsed latches

3. What is meant by maximum delay or setup time failure and how to avoid?

If the combinational logic delay is too high, the receiving element will miss its setup time and sample the wrong value. This is called a *setup time failure* or *max-delay failure*. Max-delay can be solved by redesigning the logic to be faster or by increasing the clock period.

4. Define sequencing overhead.

Sequencing overhead is defined as an additional delay to Tokens (Data) that are already critical, decreasing the performance of the system. This extra delay is called *sequencing overhead*.

5. What is meant by Min-delay failure and how to avoid?

If the hold time is large and the contamination delay is small, data can incorrectly propagate through two successive elements on one clock edge, corrupting the state of the system. This is called a *race condition*, *hold-time failure*, or *min-delay failure*.

Min-delay can only be fixed by redesigning the logic, not by slowing the clock.

6. Define time borrowing.

Time borrowing is defined as if one half-cycle or stage of a pipeline has too much logics, it can borrow time into the next half-cycle or stage.

Time borrowing can accumulate across multiple cycles.

7. Define clock skew. (April 2018, April 2019, NOV 2021)[Apr/May 2022]

Clocks have some uncertainty in their arrival times that can cut into the time available for useful computation.

8. How to design CMOS flip-flop?

Dynamic inverting flip-flop built from a pair of back-to-back dynamic latches.

9. How to design Semidynamic Flip-flop?

Semidynamic flip-flop (SDFF) is a cross coupled between a pulsed latch and a flip-flop.

10. What is meant by semidynamic flip-flop(SDFF)?

The SDFF accepts rising inputs slightly after the rising clock edge. Like a flip-flop, falling inputs must set up before the rising clock edge called as *semidynamic*. It combines the dynamic input stage with static operation.

11. What is meant by Differential flip-flop?

Differential flip-flops accept true and complementary inputs and produce true and complementary outputs. They are built from a clocked sense amplifier so that they can respond to small differential input voltages.

12. What is meant by True Single Phase Clock (TSPC) Latch or flip-flop?

True Single Phase clock Latch or flipflop avoids complementary clock pulse.

13. What are sequencing dynamic circuits?

Sequencing dynamic circuits are

i. Traditional domino circuit

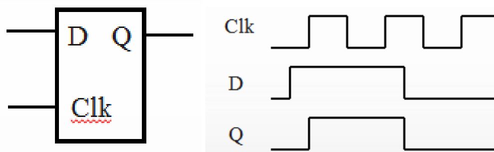
ii. Skew tolerant domino circuit

14. Define Synchronizer. (MAY/JUE-2014)

Synchronizer is defined as a circuit that accepts an input that can change at arbitrary times and produces an output aligned to the synchronizer's clock.

15. What is a Latch? (NOV/DEC-2014)

Latch is a bistable device. i.e., it has two stable states (0 and 1). It is the level triggering method.



16. What is a flip-flop? (NOV/DEC-2014)

Flip-flop is a bistable device. i.e., it has two stable states (0 and 1). It is the edge triggering method.

17. What is meant by Bistability and metastability? (NOV 2021) [Apr/May 2022]

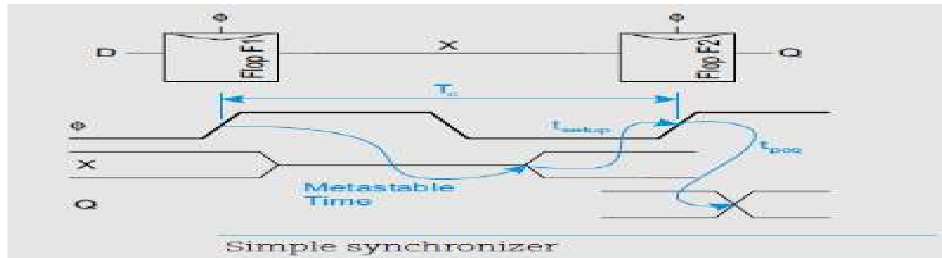
A latch is a bistable device. i.e., it has two stable states (0 and 1). Latch can enter a metastable state in which the output is at an indeterminate level between 0 and 1.

18. Define aperture.

Aperture is defined as a setup and hold time around the rising edge of the clock.

19. How to design simple synchronizer circuit.

Simple synchronizer built from a pair of flip-flops. *F*1 samples the asynchronous input *D*. The output *X* may be metastable for some time, but will settle to a good level with high probability.



Simple synchronizer

20. What is meant by Arbiter?

The *arbiter* is closely related to the synchronizer. It determines which of two inputs arrived first.

If the spacing between the inputs exceeds some aperture time, the first input should be acknowledged.

If the spacing is smaller, exactly one of the two inputs should be acknowledged, but the choice is arbitrary.

21. What are the advantages of differential Flip flop? (Nov 2011)

The advantages of differential Flip flops are

a. Reduce the parasitic delay of the pull down networks.
b. Lower electric fields across the pull down networks.
c. It reduces the channel length of the transistors.

22. State the reasons for the speed advantages of CVSL family. (Nov 2012, Nov 2014)

CVSL has a potential speed advantage because all of the logic is performed with nMOS transistors, thus reducing the input capacitance.

23. Enumerate the features of synchronizers. (May 2013)

The features of synchronizers are,

a. A good synchronizer should have a feedback loop with high gain bandwidth product.
b. It can produce metastable output.

24. What are the disadvantages of using a pseudo nMOS gate instead of a full CMOS gate? (May 2012)

Ratioed circuits dissipate power continually in certain states and have poor noise margin than complementary circuits. Ratioed circuits used in situations where smaller area is needed.

25. What is Klass-semidynamic flip-flop?

Klass-semidynamic is a single-input single-output positive edge-triggered flip-flop. It is domino-style from end allows for efficient embedded combinational logic and reduces the load on the data network.

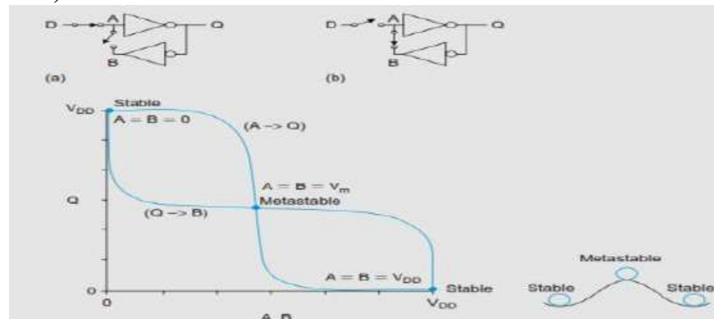
26. What are the different phases of VLSI design flow?

The different phases of VLSI design flows are**,**

- Function Verification and testing, logic synthesis/Timing verification
- Logical verification and testing

- Floor planning automatic place and route
- Layout verification
- Implementation

27. Draw the circuit diagram of a CMOS bistable element and its time domain behavior. (APRIL/MAY-2011)



28. What is CMOS clocked SR flipflop?

CMOS Clocked SR flip-flop consists of a cross-coupled inverter pair, plus 4 extra transistors to drive the flip-flop from one state to another and to provide clocked operation.

29. What do mean by multiplexer based latches?

Multiplexer based latches can provide similar functionality to the SR latch, but has the important added advantage that the sizing of devices only affects performance and is not critical to the functionality.

30. What is Master-Slave Based Edge Triggered Register?

The register consists of cascading a negative latch (master stage) with a positive latch (slave stage).

31. What is the advantage of multiplexer based latch?

The advantage of the multiplexer-based register is the feedback loop is open during the sampling period, and sizing of devices is not critical to functionality.

32. What is pseudostatic?

The register employs a combination of static and dynamic storage approaches depending upon the state of the clock.

33. What is called Clocked CMOS Register? (May 2016)

C^2MOS is a positive edge-triggered register based on the master-slave concept which is insensitive to clock overlap. The register operates in two phases.

34. What is meant by Dual-edge Triggered Register? Give it advantage.

Dual-edge triggered register is a design of sequential circuits that sample the input on both edges. The advantage of this scheme is that a lower frequency clock.

35. What is True Single-Phase Clocked Register (TSPCR)?

The True Single-Phase Clocked Register (TSPCR) uses a single clock (without an inverse clock).

36. What is the advantage and disadvantage of True Single-phase clocked register?

The main advantage is the use of a single clock phase. The disadvantage is the increase the number of transistors. 12 transistors are required.

37. What is pipelining? (Dec. 2016, April 2017)

Pipelining is a popular design technique used to accelerate the operation of the datapaths in digital processors.

38. What is necessary of non-overlapping clocks?

The non-overlapping of the clocks ensures correct operation. When CLK and $\overline{CLK}$ signals are non-overlapping, correct pipeline operation is obtained.

39. What is topology for NORA-CMOS?(Nov 2017)

The latch-based pipeline circuit implemented using C^2MOS latch is known as NORA-CMOS circuit. A NORA CMOS circuit is race-free as long as all the logic functions between the latches are non-inverting.

40. Tabulate the operation modules of NORA-CMOS circuit.

	CLK block		$\overline{CLK}$ block	
	Logic	**Latch**	**Logic**	**Latch**
$CLK = 0$	Precharge	Hold	Evaluate	Evaluate
$CLK = 1$	Evaluate	Evaluate	Precharge	Hold

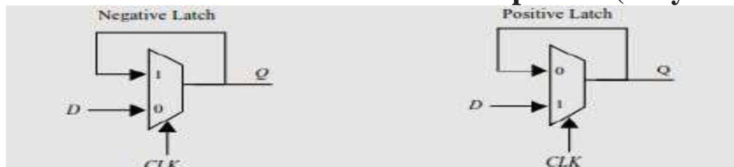
41. Define the dynamic-logic rule.

Inputs to a dynamic CLKn (CLKp) block are only allowed to make a single 0→1 (1 → 0) transition during the evaluation period.

42. Define C^2MOS design rule.

C^2MOS design rule is defining to avoid races, the number of static inversions between C^2MOS latches should be even.

43. Draw the switch level schematic of multiplexer. (May 2016)



.

44. What is known as H-tree clock distribution?

The H-tree configuration is useful for regular-array networks in which all elements are identical and the clock can be distributed as a binary tree. For example, arrays of identical tiled processors.

45. Define clock jitter. (Nov 2017)[Apr/May 2022]

Clock jitter refers to the temporal variation of the clock period at a given point. i.e, the clock period can reduce or expand on a cycle-by-cycle basis.

46. What is meant by Latch based clocking?

Use of registers in sequential circuits enables a robust design methodology. Advantage is combinational logic is separated by transparent latches.

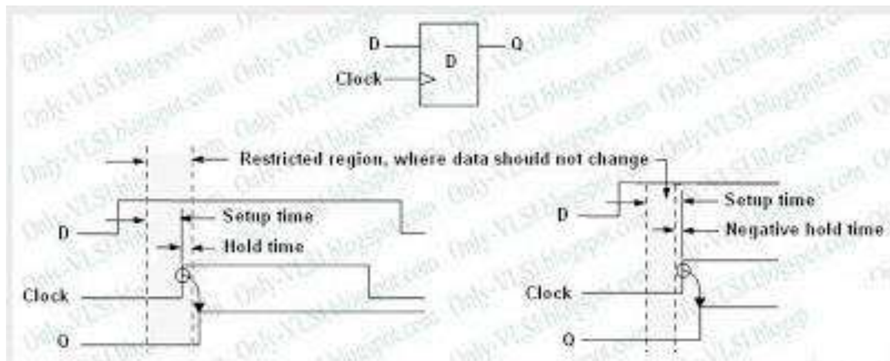
47. Compare synchronous and asynchronous design. (April 2017)

Synchronous design	Asynchronous design
Synchronous logic is both faster and simpler.	Asynchronous logic is slow and complex.
Distributing a clock at high speed becomes exceedingly difficult.	Distributing a clock at high speed becomes easy.

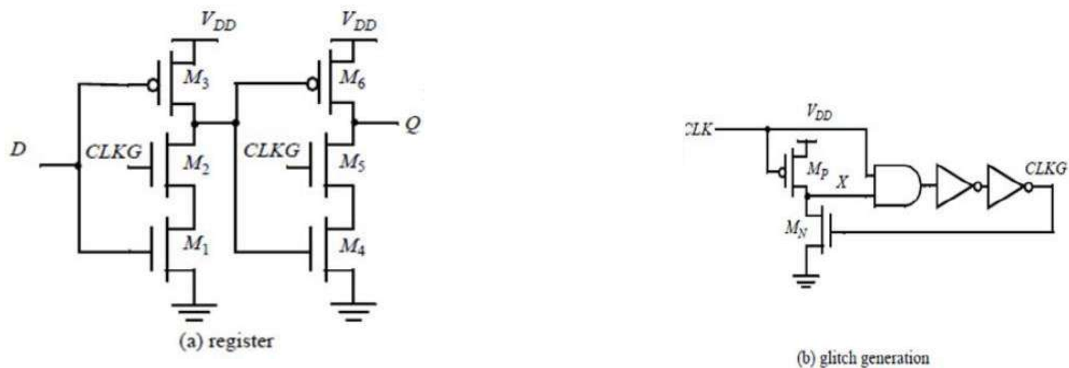
48. Define Set up time and hold time. (Nov 2019)

The **setup time** is the interval before the clock where the data must be held stable.

The **hold time** is the interval after the clock where the data must be held stable. Hold time can be negative, which means the data can change slightly before the clock edge and still be properly captured. Most of the current day flip-flops has zero or negative hold time.

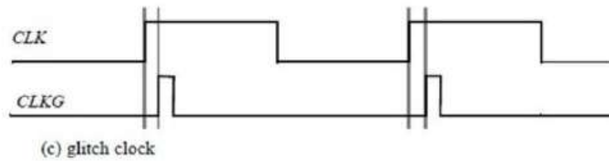


49. Draw the circuit and wave form for Pulse Registers.



(a) register

(b) glitch generation



(c) glitch clock

50. Define Pulse Registers.

The idea is to construct a short pulse around the rising (or falling) edge of the clock.
This pulse acts as the clock input to a latch, sampling the input only in a short window.
Race conditions are thus avoided by keeping the opening time (i.e, the transparent period) of the latch very short.
The combination of the glitch generation circuitry and the latch results in a positive edge-triggered register.

51. List the advantage and disadvantage of Pulse Registers.

Advantage

- The reduced clock load and the small number of transistors required.
- The glitch-generation circuitry can be amortized over multiple register bits.

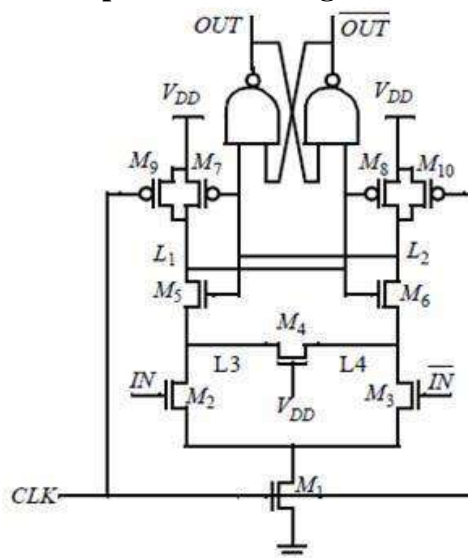
Disadvantage

- A substantial increase in verification complexity.
- This has prevented a wide-spread use.

52. What is Sense – Amplifier Based Registers?

- A sense amplifier structure to implement an edge- triggered register.
- Sense amplifier circuits accept small input signals and amplify them to generate rail-to-rail swings.
- There are many techniques to construct these amplifiers, with the use of feedback (e.g., cross-coupled inverters).

53. Draw the circuit of Sense – Amplifier Based Registers.



54. Differentiate between latch and flip-flop. (Nov 2015) (Nov 2019) [May 2021]
Compare register and latch. (April 2018) [April / May 2023]

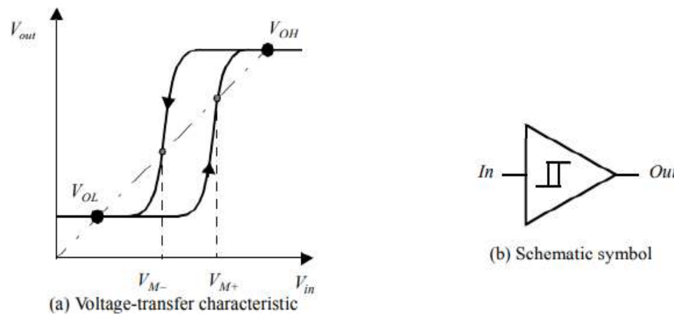
Latch is a bistable device. i.e., it has two stable states. It is the level triggering method.
Flip-flop is a bistable device. i.e., it has two stable states. It is the edge triggering method.
Register has number of flipflops.

55. Define Schmitt trigger.

A Schmitt trigger is a device with two important properties:

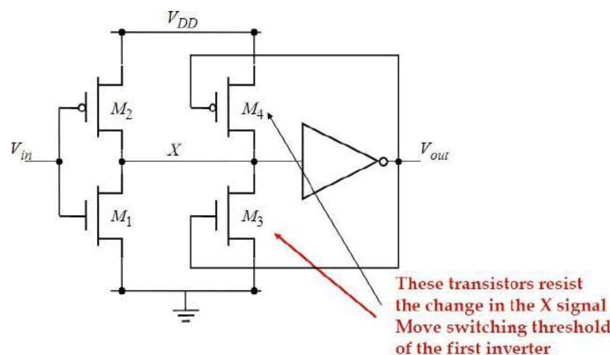
1. It responds to a slowly changing input waveform with a fast transition time at the output.
2. The voltage-transfer characteristic of the device displays different switching thresholds for positive- and negative-going input signals.

56. Draw the symbol, circuit and voltage transfer characteristics of Schmitt trigger.



(a) Voltage-transfer characteristic

(b) Schematic symbol



57. Define Astable sequential circuits. Draw the circuit also. [April / May 2023]

- An astable circuit has no stable states.
- The output oscillates back and forth between two quasi-stable states with a period determined by the circuit topology and parameters (delay, power supply, etc.).
- One of the main applications of oscillators is the on-chip generation of clock signals.

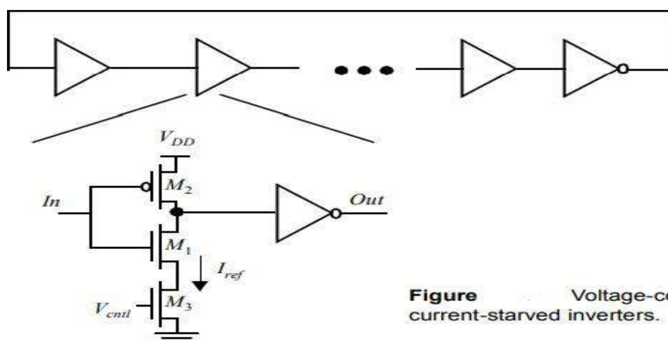
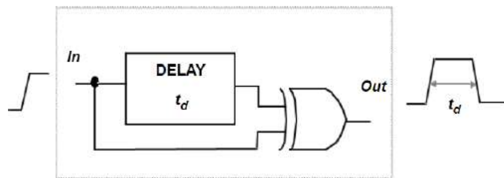


Figure Voltage-controlled oscillator based on current-starved inverters.

58. Define Monostable sequential circuits. [April / May 2023]

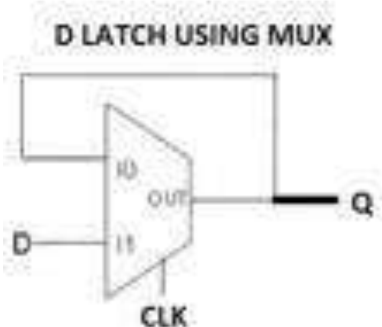
Monostable − Monostable circuits have only one stable operating point and even if they are temporarily perturbed to the opposite state, they will return in time to their stable operating point. Example: Timers, pulse generators.



59. State the use of Schmitt trigger.[Nov/Dec 2022]

Schmitt triggers can be used to change a sine wave into a square wave, clean up noisy signals, and convert slow edges to fast edges.

60. Draw a MUX based negative level sensitive D-latch. [Nov/Dec 2022]



61. List the timing classification of Digital system. [May 2021]

Synchronous, mesochronous and plesiochronous with respect to a system clock.

UNIT – IV
INTERCONNECT, MEMORY ARCHITECTURE AND ARITHMETICCIRCUITS

Interconnect Parameters – Capacitance, Resistance, and Inductance, Electrical Wire Models, **Sequential digital circuits:** adders, multipliers, comparators, shift registers. Logic Implementation using Programmable Devices (ROM, PLA, FPGA), Memory Architecture and Building Blocks, Memory Core and Memory Peripherals Circuitry.

4.1. Design of Data path circuits:

Discuss about data path circuits.

- Data path circuits are meant for passing the data from one segment to other segment for processing or storing.
- The datapath is the core of processors, where all computations are performed.
- It is generally defined with general digital processor. It is shown in figure.

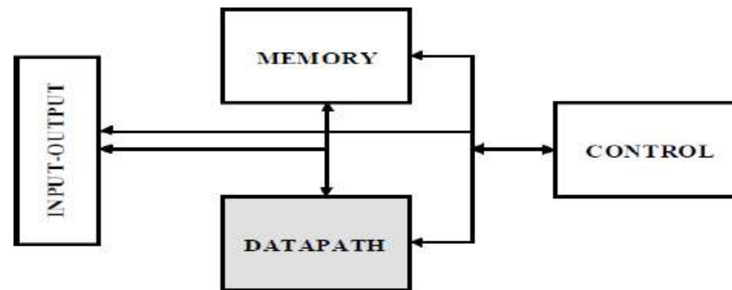
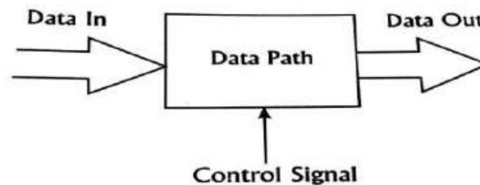
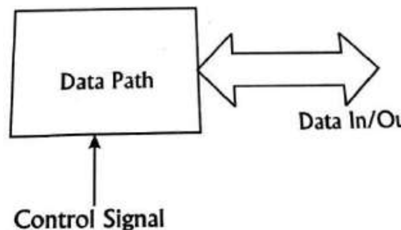


Figure: General digital processor

- If only data path and its communication is shown as



- In this, data is applied at one port and data output is obtained at second port.



- Data path block consists of arithmetic operation, logical operation, shift operation and temporary storage of operands.
- Datapaths are arranged in a bit sliced organization.
- Instead of operating on single bit digital signals, the data in a processor are arranged in a word based fashion.
- Bit slices are either identical or resemble a similar structure for all bits.

- The datapath consists of the number of bit slices (equal to the word length), each operating on a single bit. Hence the term is ***bit-sliced.***

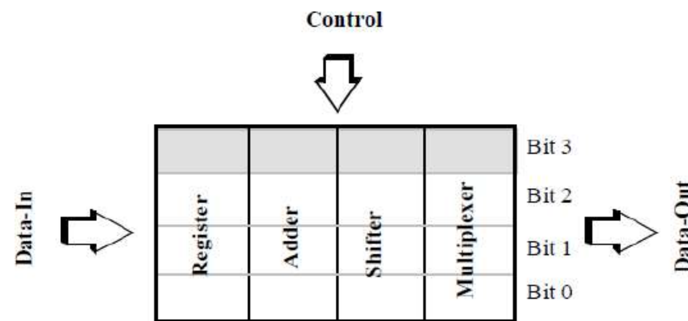


Figure: Bit-sliced datapath organization

****

4.2. Ripple Carry Adder:

- **Draw the structure of ripple carry adder and explain its operation. (Nov 2017)**
- **Explain the operation of a basic 4 bit adder. (Nov 2016)**
- **Realize a 1-bit adder using static CMOS logic. Optimize the Boolean expressions of sum and carryout and realize a 1-bit adder using static CMOS logic. Also realize a 1-bit adder using transmission gate. Compare all the three cases from hardware perspective. (Nov 2019)**

Architecture of Ripple Carry Adder:

- AOI Full adder circuit (AND OR INVERT)
- An AOI algorithm for static CMOS logic circuit can be obtained by using the equation.

$$C_{i+1} = a_i b_i + c_i.(a_i + b_i)$$
$$\overline{S_i} = (a_i + b_i + c_i)\overline{c_i} + (a_i.b_i.c_i)$$

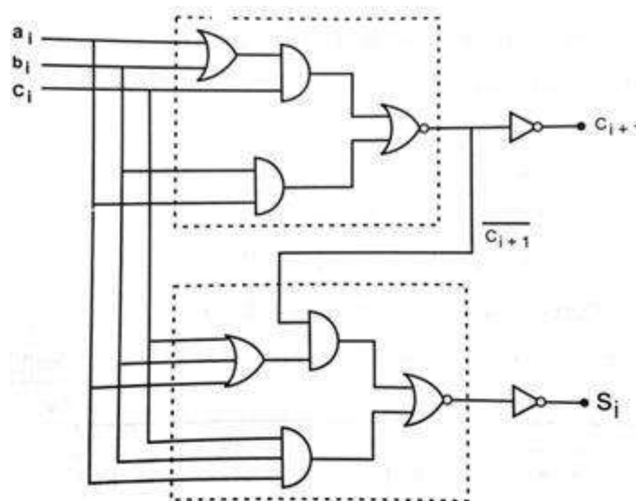


Figure: AOI Full adder

- If n bits are added, then we can get n-bit sum and carry of C_n.

C_i= Carry bit from the previous column.

- N bit ripple carry adder needs n full adders with C_{i+1} carry out bit.

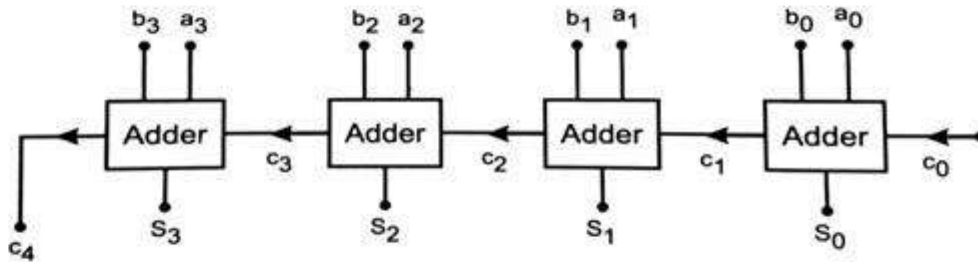


Figure: Ripple carry adder

- The overall delay depends on the characteristics of full adder circuit. Different CMOS implementation can produce different delay parts.
- t_{di}- worst case delay through the ith stage. We can calculate the total delay using the following equation

$$t_{4b} = t_{d3}+t_{d2}+t_{d1}+t_{d0} \text{ and } t_{d0} = t_d(a_0,b_0 \rightarrow c_1)$$

- This is the time for the input to produce the carry out bit.

$$t_{d1}= t_{d2} = t_d(c_{in} \rightarrow c_{out})$$

$$t_{d3}= t_d(c_{in} \rightarrow S_3)$$

$$t_{4b} = t_d(c_{in} \rightarrow S_3) + 2t_d(c_{in} \rightarrow c_{out}) + t_d(a_0,b_0 \rightarrow c_1)$$

- If it is extend to n-bit, then the worst case delay is

$$t_{n\text{-}bit} = t_d(c_{in} \rightarrow S_{n-1}) + (n-2)t_d(c_{in} \rightarrow c_{out}) + t_d(a_0,b_0 \rightarrow c_1)$$

- Worst case delay linear with the number of bits

$$t_d = O(N)$$

$$t_{adder} = (N-1)t_{carry} + t_{sum}$$

- The figure below shows 4-bit adder/subtractor circuit.
- In this, if add/sub=0, then sum is a+b. If add/sub=1, then the output is a-b.

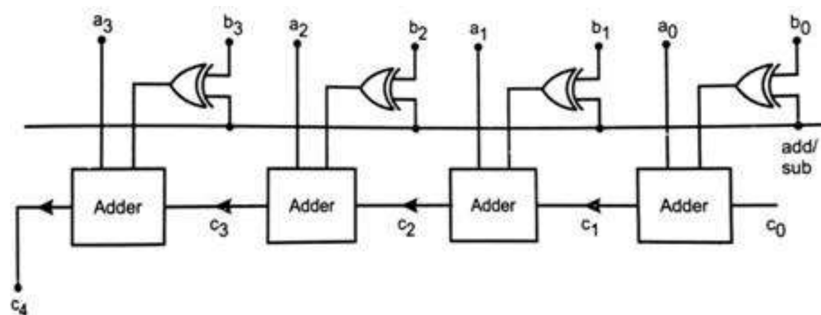


Figure:4-bit adder/subtractor circuit

- Sum and carry expressions are designed using static CMOS.
- It requires 28 transistors which lead large area and circuit is slow.

  Sum, S= $ABC_i + \overline{C_0}\ (A + B + C_i)$ and Carry, C_0= $AB + BC_i + C_iA$

Drawbacks:

- Circuit is slower.

- In ripple carry adder, carry bit is calculated along with the sum bit. Each bit must wait for calculation of previous carry.

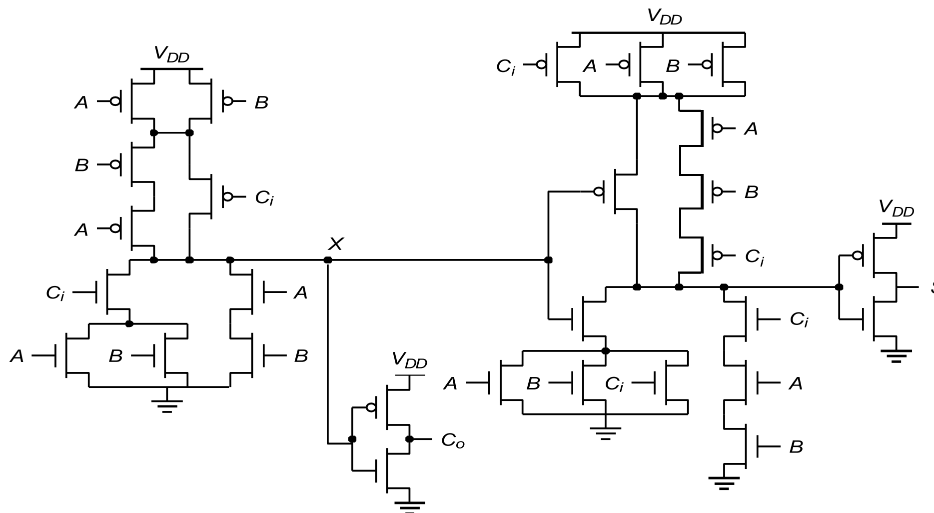


Figure:Complimentary Static CMOS Full Adder

****

4.3. Carry Look Ahead Adder (CLA):

- ❖ **Explain the operation and design of Carry lookahead adder (CLA). (May 2017, Nov 2016)[Apr/May 2022] [Nov/Dec 2022]**
- ❖ **How the drawback in ripple carry adder overcome by carry look ahead adder and discuss. (Nov 2017)**
- ❖ **Explain the concept of carry lookahead adder and discuss its types. (April 2018)**
- ❖ **Derive the necessary expressions of a 4 bit carry look ahead adder and realize the carry out expressions using dynamic CMOS logic. (April 2019-13M)**

- A carry-lookahead adder (CLA) is a type of adder used in digital circuit.
- A carry-lookahead adder improves speed by reducing the amount of time required todetermine carry bits.
- In ripple carry adder, carry bit is calculated alongwith the sum bit.
- Each bit must wait until the previous carry is calculated to begin calculating its own result and carry bits.
- The carry-lookahead adder calculates one or more carry bits before the sum, which reduces the wait time to calculate the result of the larger value bits.
- A ripple-carry adder works starting at the rightmost (LSB) digit position, the two corresponding digits are added and a result obtained. There may be a carry out of this digit position.
- Accordingly all digit positions other than LSB.Need to take into account the possibility to add an extra 1, from a carry that has come in from the next position to the right.
- Carry lookahead depends on two things:
  - ✓ Calculating, for each digit position, whether that position is going to propagate a carry if one comes in from the right.

- ✓ Combining these calculated values to be able to realize quickly whether, for each group of digits, that group is going to propagate a carry.

- Theory of operation:
  - ✓ Carry lookahead logic uses the concept of generating and propagating carry.
  - ✓ The addition of two 1-digit inputs A and B is said to generate if the addition will carry, regardless of whether there is an input carry.
- *Generate:*
  - ✓ In binary addition, A + B generates if and only if both A and B are 1.
  - ✓ If we write G(A,B) to represent the binary predicate that is true if and only if A + B generates, we have:

$$G(A,B) = A \cdot B$$

- *Propagate:*
  - ✓ The addition of two 1-digit inputs A and B is said to propagate if the addition will carry whenever there is an input carry.
  - ✓ In binary addition, A + B propagates if and only if at least one of A or B is 1.
  - ✓ If we write P(A,B) to represent the binary predicate that is true if and only if A + B propagates, we have:

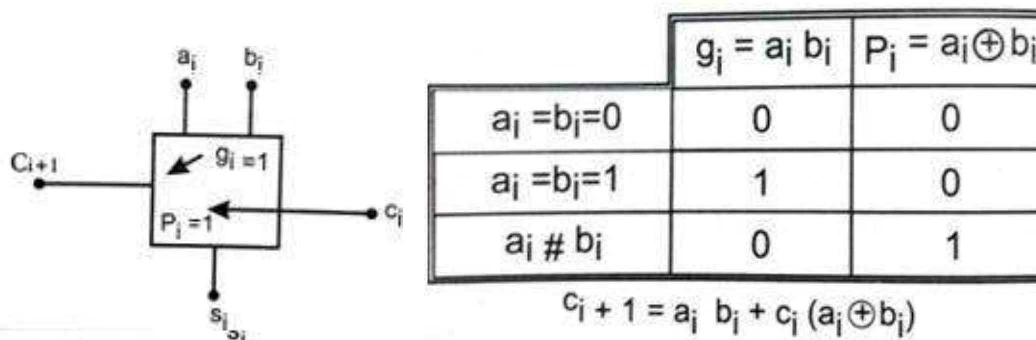
$$P(A,B) = A \oplus B$$

- These adders are used to overcome the latency which is introduced by the rippling effect of carry bits.
- Write carry look-ahead expressions in terms of the generate g_i and propagate p_i signals. The general form of carry signal c_i thus becomes

$$c_{i+1} = a_i.b_i + c_i.(a_i \oplus b_i) = g_i + c_i.p_i$$

- If $a_i.b = 1$, then $c_{i+1} = 1$, write generate term as, $g_i = a_i.b_i$
- Write the propagate term as, $p_i = a_i \oplus b_i$
- Sum and carry expression are written as,

$$S_i = a_i \oplus b_i$$
$$c_1 = g_0 + p_0.c_0$$
$$c_2 = g_1 + p_1.c_1 = g_1 + p_1.(g_0 + p_0.c_0)$$
$$c_3 = g_2 + p_2.c_2$$
$$c_4 = g_3 + p_3.c_3 = g_3 + p_3.g_2 + p_3.p_2.g_1 + p_3.p_2.p_1.g_0 + p_3.p_2.p_1.p_0.c_0$$



	$g_i = a_i b_i$	$P_i = a_i \oplus b_i$
$a_i = b_i = 0$	0	0
$a_i = b_i = 1$	1	0
$a_i \# b_i$	0	1

$c_{i+1} = a_i\ b_i + c_i\ (a_i \oplus b_i)$

Figure:Symbol and truth table of generate & propagate

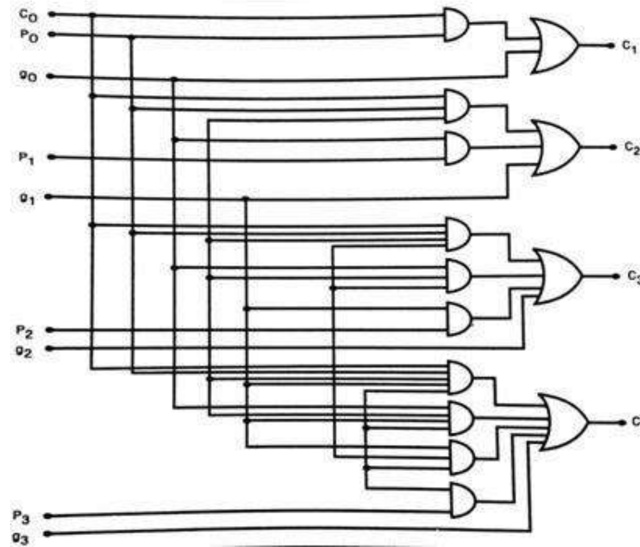


Figure – Logic network for 4-bit CLA carry bits

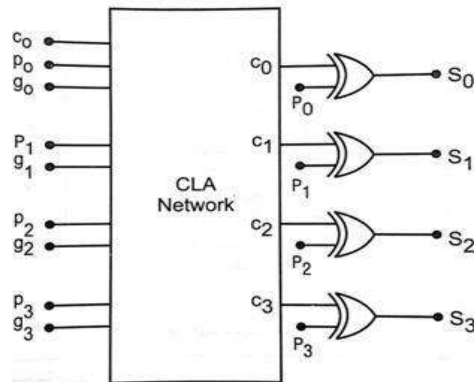


Figure – Sum calculation using the CLA network

- The symmetry in the array is shown in mirror. It allows more structured layout at the physical design level.

6

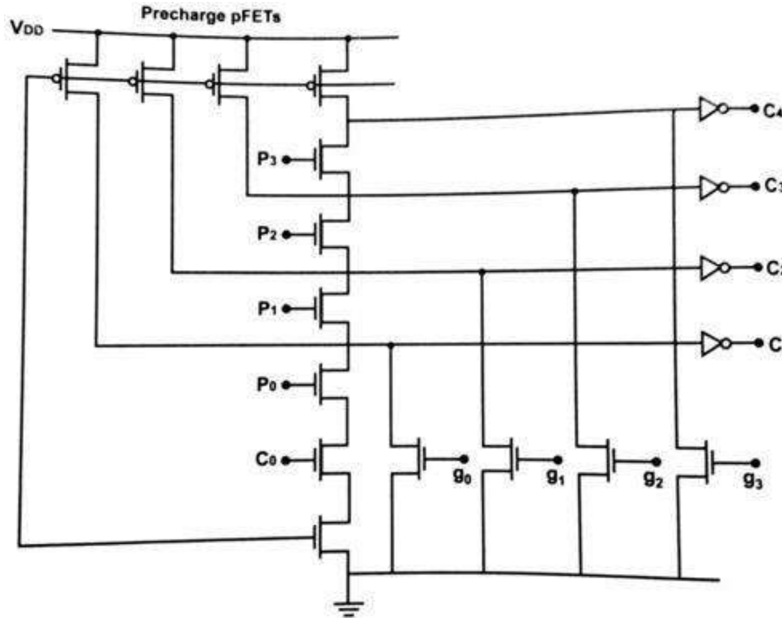
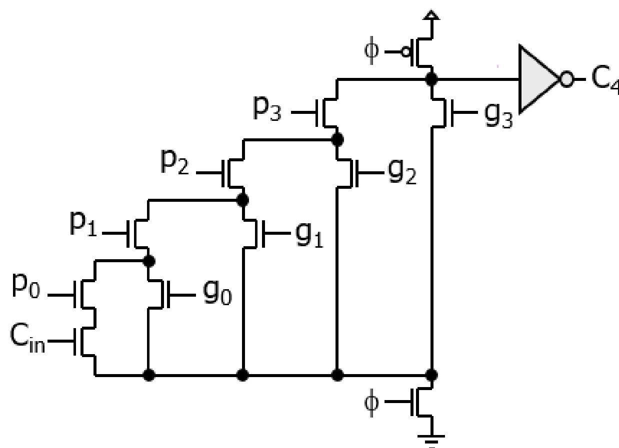


Figure – MODL carry circuit

- MODL-Multiple Output Domino Logic.
- MODL is non-inverting logic family and is a dynamic circuit technique.
- Its limitations are
  - i. Clocking in mandatory
  - ii. The output is subject to charge leakage and charge sharing.
  - iii. Series connected nFET chains can give long discharge times.

CLA: Dynamic Logic Implementation

- Dynamic gate implementation:
  - $C_4 = g_3 + p_3 \cdot (g_2 + p_2 \cdot (g_1 + p_1 \cdot (g_0 + P_0 \cdot C_{in})))$



****

4.4. Manchester Carry Chain Adder:

Discuss about Manchester Carry Chain Adder.

- The Manchester carry chain is a variation of the carry-lookaheadadder that uses shared logic to lower the transistor count.
- A Manchester carry chain generates the intermediate carries by tapping off nodes in the gate that calculates the most significant carry value.
- Dynamic logic can support shared logic, as transmission gate logic.
- One of the major drawbacksof the Manchester carry chain is increase the propagation delay.
- A Manchester-carry-chain section generally won't exceed 4 bits.
- In this adder, the basic equation is $c_{i+1} = g_i + c_i . p_i$

  Where $p_i = a_i \oplus b_i$ and $g_i = a_i . b_i$

- Carry kill bit $k_i = \overline{a_i + b_i} = \overline{a_i}.\overline{b_i}$
- If K_i=1, then p_i=0 and g_i=0. Hence, k_i is known as carry kill bit.

a_i	b_i	p_i	g_i	k_i
0	0	0	0	1
0	1	1	0	0
1	0	1	0	0
1	1	0	1	0

Table

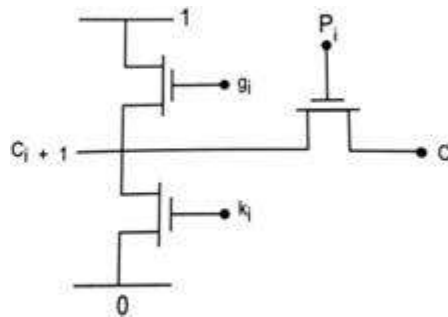
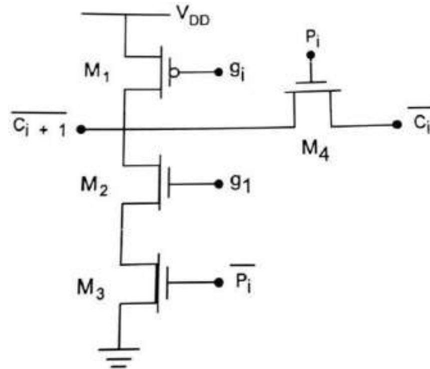


Figure – switch level circuit

- In the circuit shown below $\overline{C_i}$ is used as an input if P_i = 0, then M_3 is ON, M_4 is OFF.
- If g_i=0, then M_1 is ON, M_2 is ON
- If g_i=1, then M_2 is OFF, M_4 is ON and output equal to zero.
- If P_i=1, then this case is a complicated one.



- In dynamic circuit figure
- If $\phi = 0$, then recharge occur and output is 1
- If $\phi = 1$, then evaluation occur.

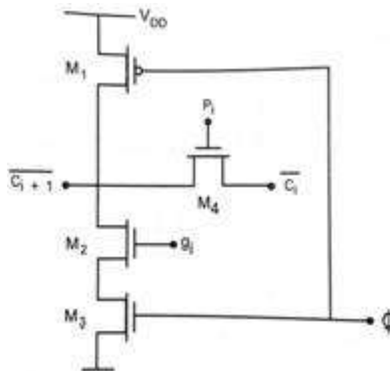
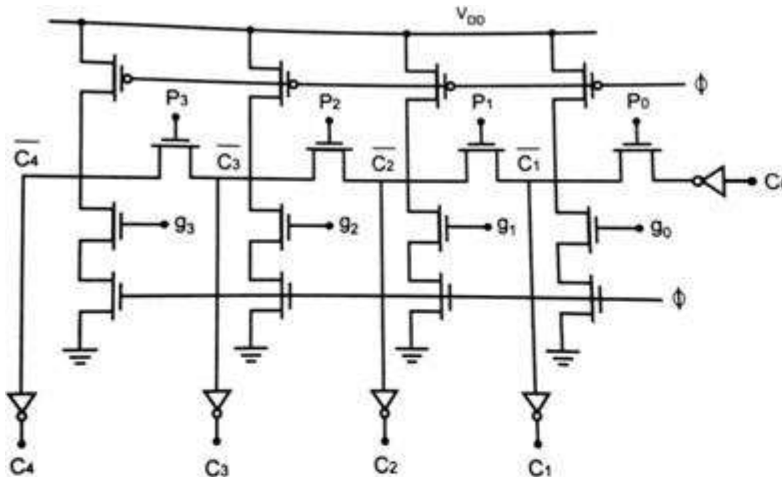


Figure dynamic circuit

- Dynamic Manchester carry chain for the carry bit upto C_4 is shown below. C_1, C_2, C_3, C_4 can be taken by using inverters. The carry input is given as C_0



****

4.4.1. HIGH SPEED ADDERS:

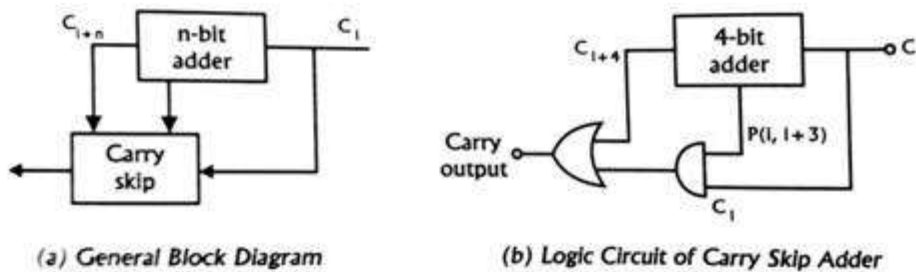
- ❖ **Discuss about different types of high speed adders. (Apr. 2016)**
- ❖ **Describe the different approaches of improving the speed of the adder. (Nov 2016)**

(i) Carry Skip(bypass) Adder:

9

Design a carry bypass adder and discuss its features. (May 2016)
Explain the carry-propagate adder and show how the generation and propagation signals are framed. [May 2021]

- It is high speed adder. It consist of adder, AND gate and OR gate.
- An incoming carry $C_{i,0}$=1 propagates through the complete adder chain and an outgoing carry $C_{0,3}$=1.
- In other words, if ($P_0P_1P_2P_3$ =1) then $C_{0,3}$= $C_{i,0}$ else either DELETE or GENERATE occurred.
- It can be used to speed up the operation of the adder, as shown in below fig (b).



(a) General Block Diagram (b) Logic Circuit of Carry Skip Adder

Figure: Carry Skip Adder.

- When BP= $P_0P_1P_2P_3$ =1, the incoming carry is forwarded immediately to the next block.
- Hence the name carry bypass adder or carry skip adder.
- Idea: if (P_0 and P_1 and P_2 and P_3 =1) the C_{03} = C_0, else "kill" or "generate".

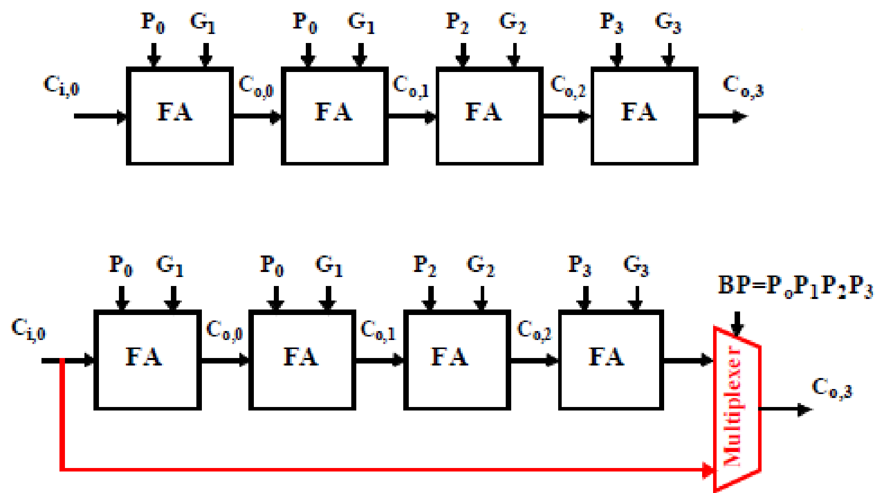
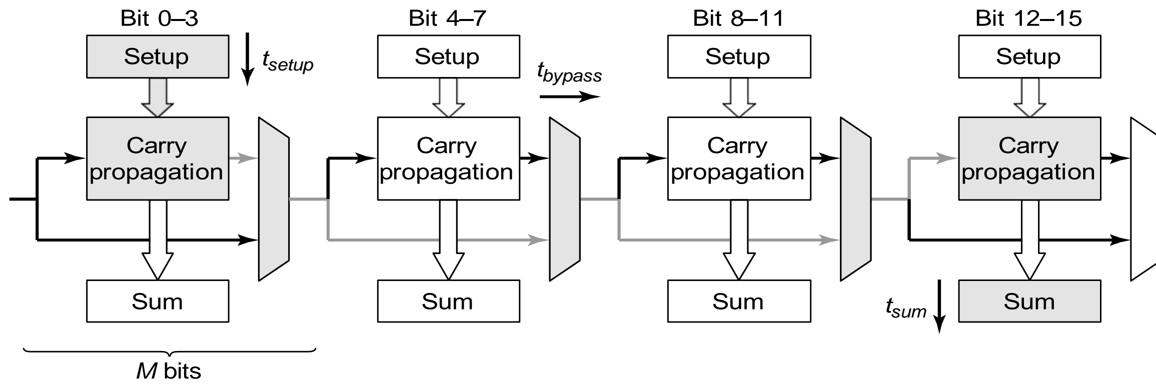


Figure: (a) Carry propagation (b) Adding a bypass

- The below figure shows n no. of bits carry skip adder.

10



$t_{adder} = t_{setup} + Mt_{carry} + (N/M-1)t_{bypass} + (M-1)\, t_{carry} + t_{sum}$(worst case)

t_{setup}: overhead time to create G, P, D signals

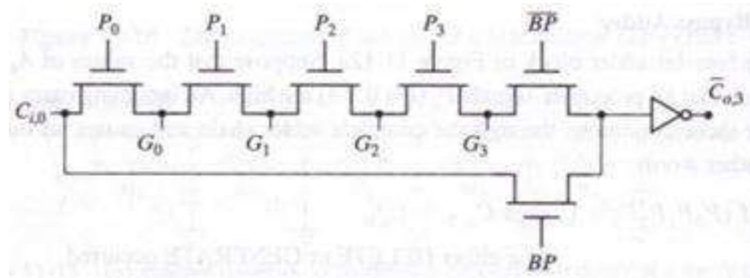


Figure: Manchester carry-chain implementation of bypass adder

(ii) Carry Select Adder:

Design a carry select adder and discuss its features. (May 2016)

- A carry-select adder is a particular way to implement an adder, which is a logic element that computes the (n+1)-bit sum of two n-bit numbers.
- The carry-select adder is simple but rather fast, having a gate level depth of O($\sqrt{n}$).
- The carry-select adder generally consists of two ripple carry adders and a multiplexer.
- Adding two n-bit numbers with a carry-select adder is done with two adders in order to perform the calculation twice.
- One time with the assumption of the carry-in being zero and the other assuming it will be one.
- After the two results are calculated (the correct sum as well as the correct carry-out),it is then selected with the multiplexer once the correct carry-in is known.
- The number of bits in each carry select block can be uniform, or variable.
- In the uniform case, the optimal delay occurs for a block size of $\sqrt{n}$.
- The O($\sqrt{n}$) delay is derived from uniform sizing, where the ideal number of full-adder elements per block is equal to the square root of the number of bits being added.
- Propagation delay, P is equal to $\sqrt{2N}$ where N = N- bit adder
- Below is the basic building block of a carry-select adder, where the block size is 4.

11

- Two 4-bit ripple carry adders are multiplexed together, where the resulting carry and sum bits are selected by the carry-in.

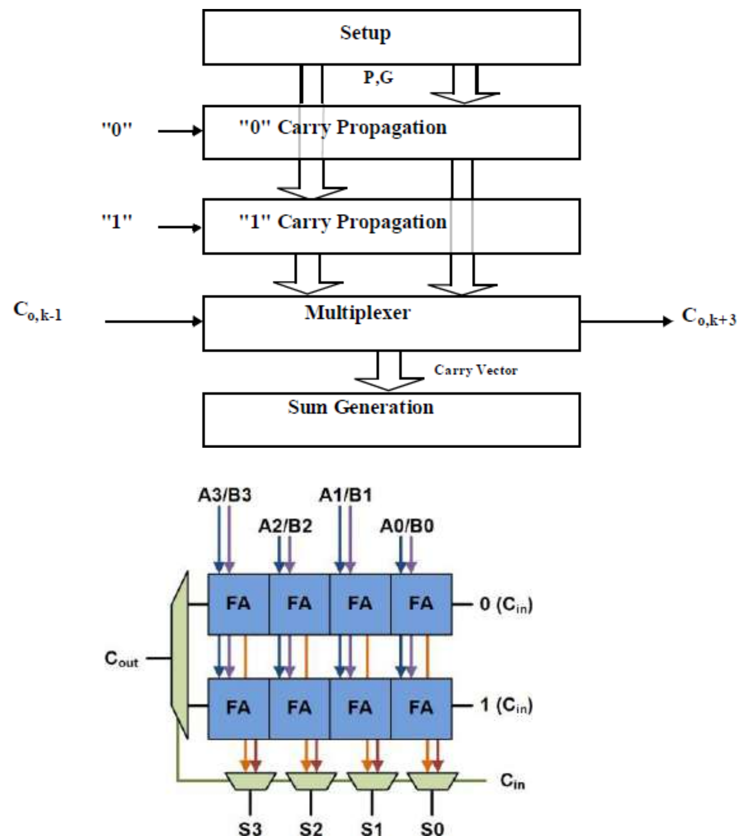


Figure: Building blocks of a carry-select adder

Uniform-sized adder:

- A 16-bit carry-select adder with a uniform block size of 4 can be created with three of these blocks and a 4-bit ripple carry adder.
- Since carry-in is known at the beginning of computation, a carry select block is not needed for the first four bits.
- The delay of this adder will be four full adder delays, plus three MUX delays.
- $t_{adder} = t_{setup} + Mt_{carry} + (N/M)t_{mux} + t_{sum}$

12

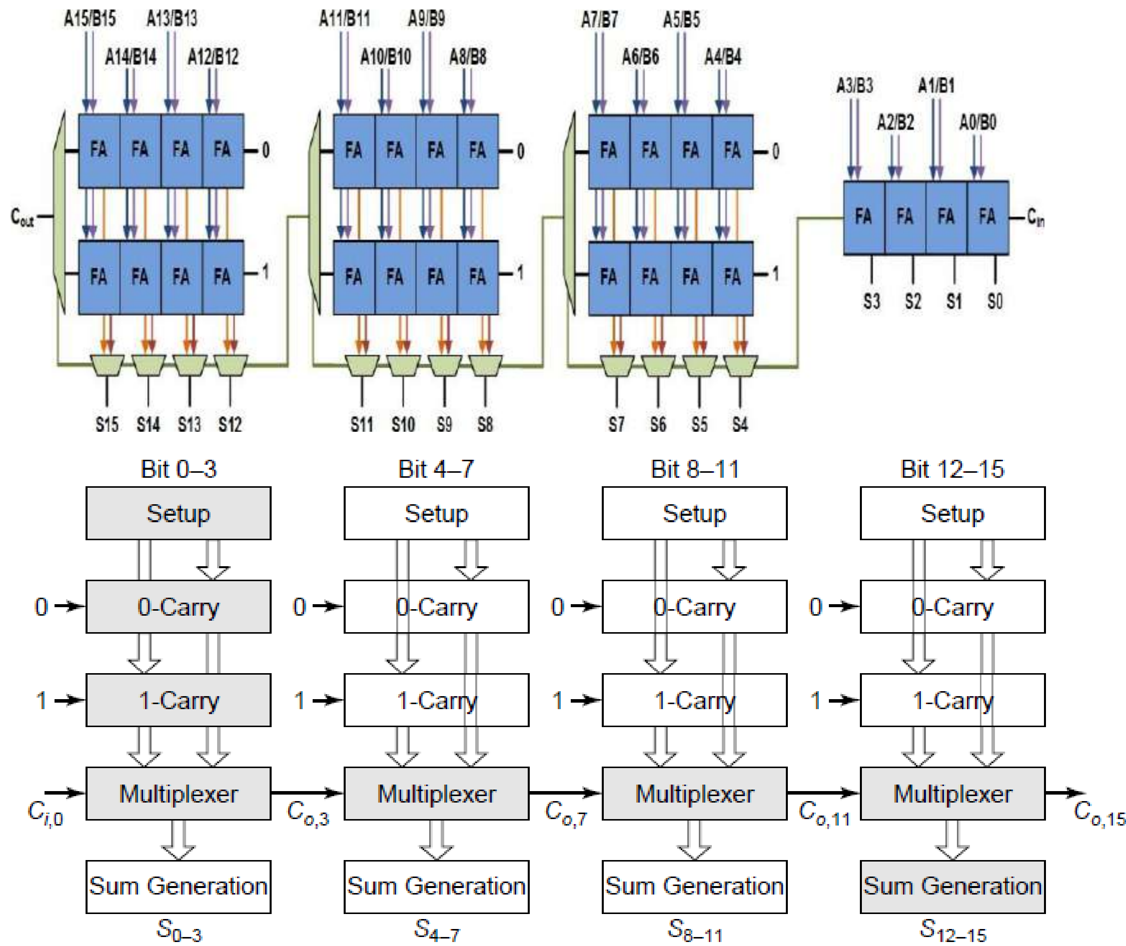
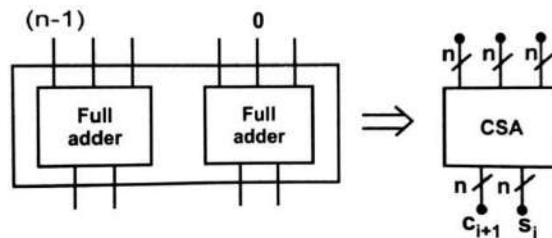


Figure: general structure of 16 bit adder

Disadvantage: hardware cost is increased.

(iii) Carry Save Adder:

- Carry save adder is similar to the full adder. It is used when adding multiple numbers.
- All the bits of a carry save adder work in parallel.
- In carry save adder, the carry does not propagate. So, it is faster than carry propagate adder.
- It has three inputs and produces 2 outputs, carry-out is saved. It is not immediately used to find the final sum value.



n-bit Carry Save Adder

13

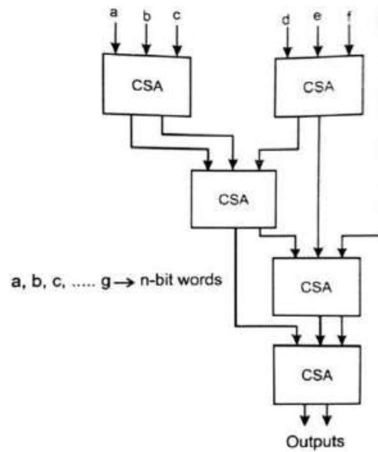


Figure: ***Carry Save Adder***

****

4.5. ALUs (ACCUMULATOR):

Briefly discuss about ALUs (accumulators).

- Accumulator acts as a part of ALU and it is identified as register A. The result of an operation performed in the ALU is stored in the accumulator.
- It is used to hold the data for manipulation (arithmetic and logical)
- Arithmetic functions are very important in VLSI. Ex: multiplication.
- Half adder circuit has two inputs and two outputs. S = $x \oplus y$, C = x.y.

$$S = x \oplus y$$
$$C = x \cdot y$$

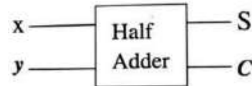
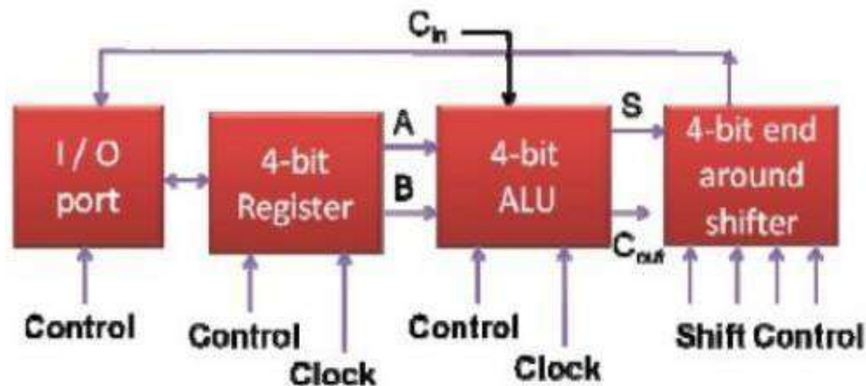


Figure: Half adder



- Full adder circuit has three inputs and two outputs

14

ai	bi	Ci	Si	Ci + 1
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

Figure : Full adder and truth table

CPL --- Complementary Pass Logic

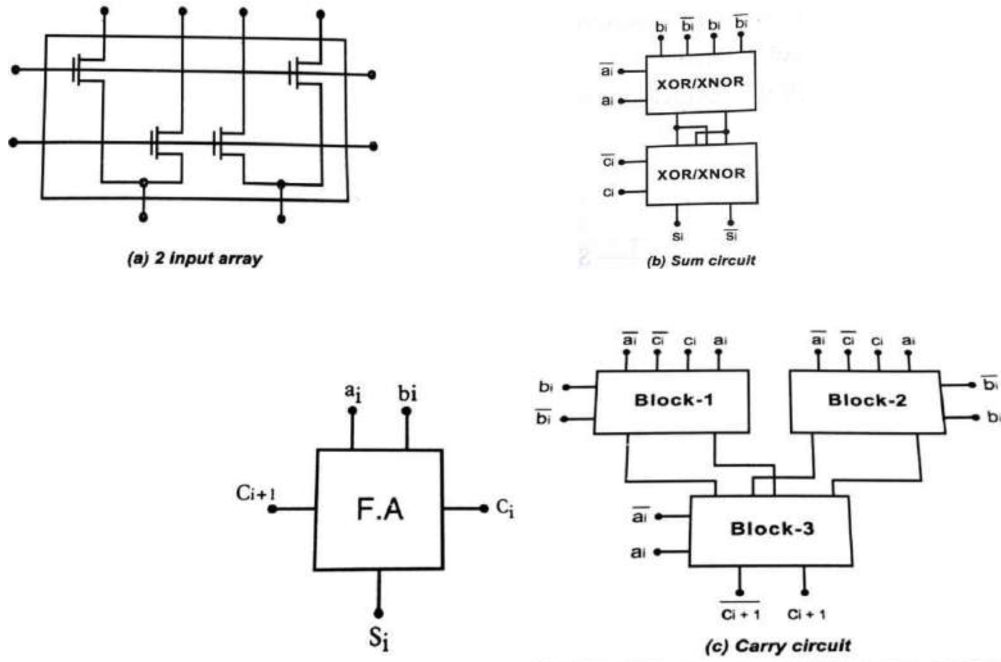


Figure : CPL Full adder design

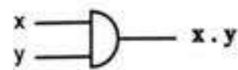
****

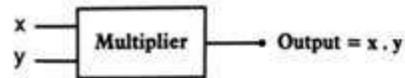
4.6. MULTIPLIERS:

- **Explain the design and operation of 4 x 4 multiplier circuit. (Apr. 2016, 2017, Nov 2016, 2018)**
- **Design a multiplier for 5 bit by 3 bit. Explain its operation and summarize the numbers of adders. Discuss it over Wallace multiplier. (Nov 2017, April 2018)**
- **Design a 4 bit unsigned array multiplier and analyze its hardware complexity. (April 2019-13M) (Nov 2019)**
- **Describe the hardware architecture of a 4-bit signed array multiplier. [Nov/Dec 2022]**

15

- A study of computer arithmetic processes will reveal that the most common requirements are for addition and subtraction.
- There is also a significant need for a multiplication capability.
- Basic operations in multiplication are given below.
  0x0=0, 0x1=0, 1x0=0, 1x1=1

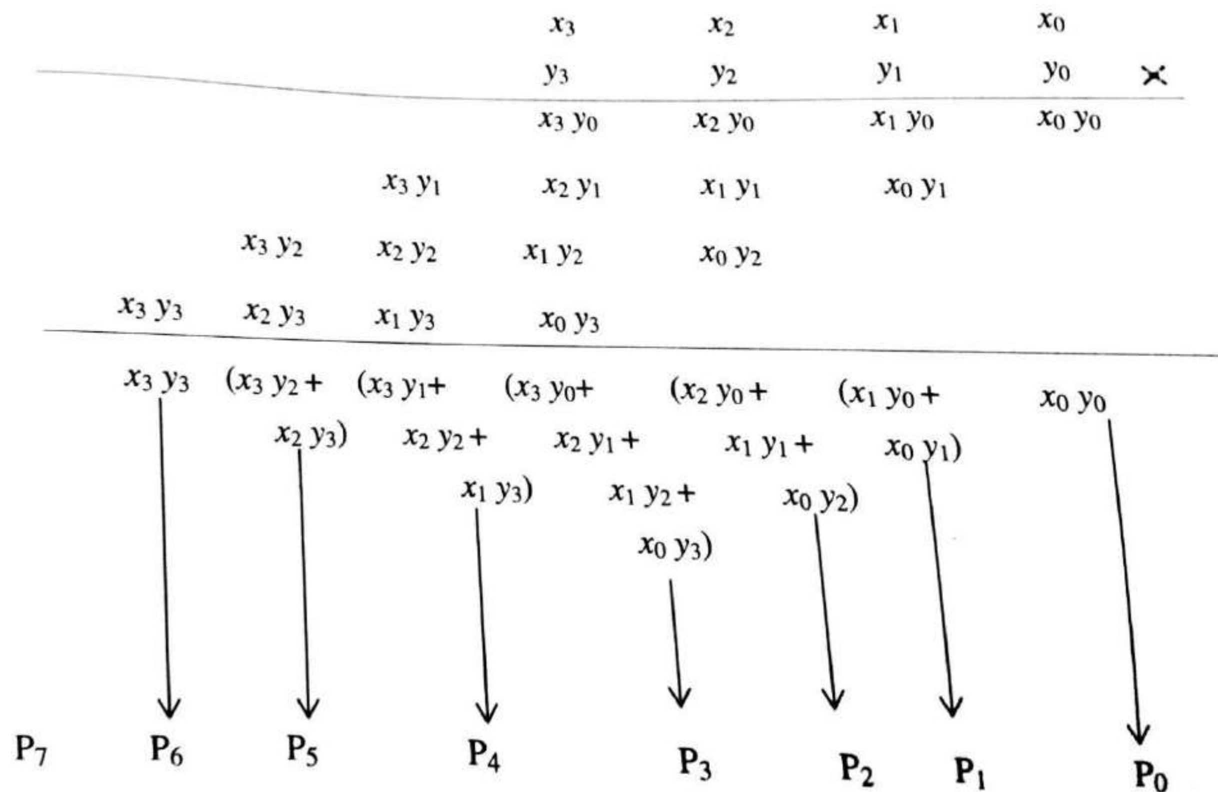




x	y	x . y
0	0	0
0	1	0
1	0	0
1	1	1

```
          1 0 1 0 1 0      Multiplicand
x             1 0 1 1      Multiplier
-----------------------
          1 0 1 0 1 0  \
        1 0 1 0 1 0     |
      0 0 0 0 0 0       |  Partial products
+   1 0 1 0 1 0        /
-----------------------
    1 1 1 0 0 1 1 1 0      Result
```

- If two different 4-bit numbers (x_0, x_1, x_2, x_3 & y_0, y_1, y_2, y_3) are multiplied then



16

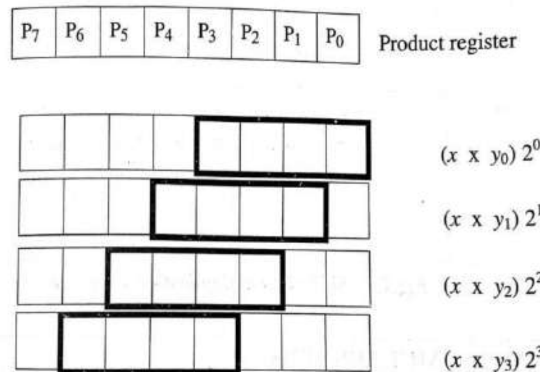
Product bits → (P_7 P_6 P_5 P_4 P_3 P_2 P_1 P_0)

$$P_i = \sum_{i=j+k} x_j y_k + c_{i-1}$$

$c_{i-1} = 0$ for $(i - 1) \leq 0$

Multiplication by shifting:

- If $x=(0010)_2 = (2)_{10}$
- If it is to be multiplied by 2, then we can shift x in left side. $x = (0100)_2 = (4)_{10}$
- If it is to be divided by 2, then we can shift in right side. $x = (0001)_2 = (1)_{10}$.
- So, shift register can be used for multiplication or division by 2.



- A practical implementation is based on the sequence. The product is obtained by successive addition and shift right operations

(i) Array multiplier:

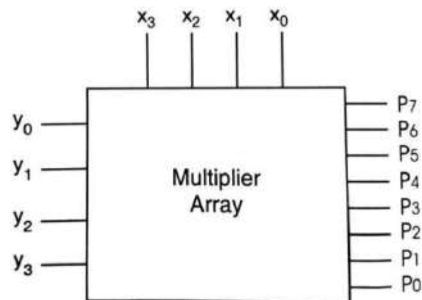


Figure: General block diagram of multiplier

- Array multiplier uses an array of cells for calculation.
- Multiplier circuit is based on repeated addition and shifting procedure. Each partial product is generated by the multiplication of the multiplicand with one multiplier digit.
- The partial products are shifted according to their bit sequences and then added.
- N-1 adders are required where N is the number of multiplier bits.
- The method is simple but the delay is high and consumes large area by using ripple carry adder for array multiplier.Product expression is given below

$$P_i = \sum_{i=j+k} x_j y_k + c_{i-1}$$

$$P = X.Y = \left(\sum_{j=0}^{n-1} x_j 2^j\right)\left(\sum_{k=0}^{n-1} y_k 2^k\right)$$

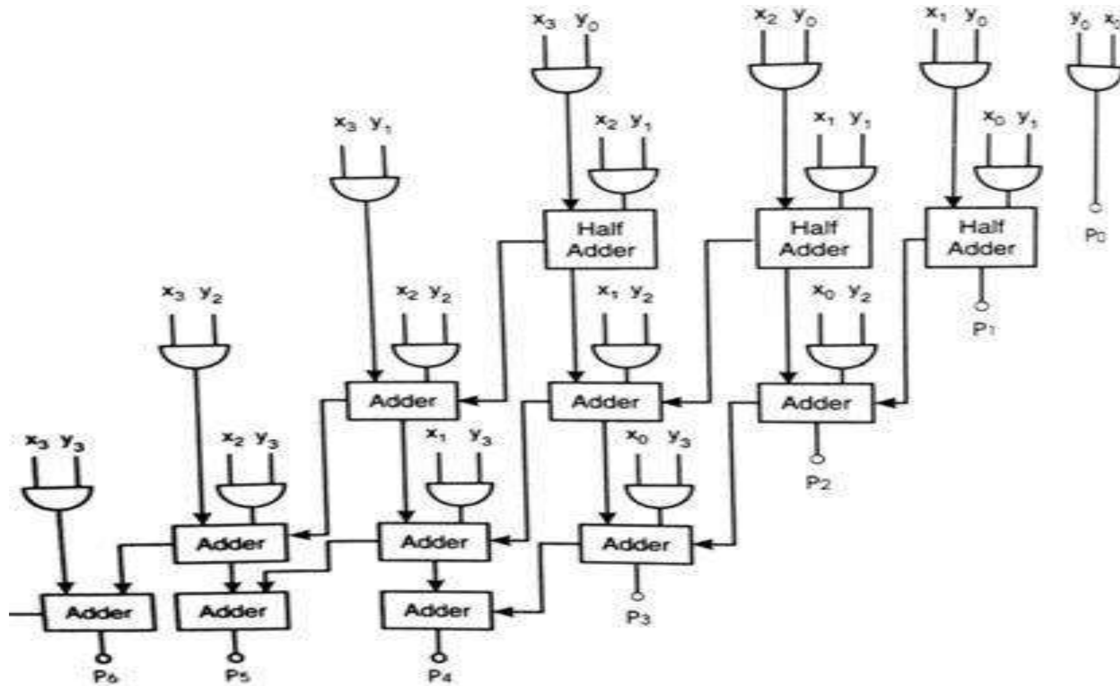


Figure: 4 x 4 array multiplier

- This multiplier can accept all the inputs at the same time. An array multiplier for n-bit word need n(n-2) full adders, n-half adder and n^2 AND gates.

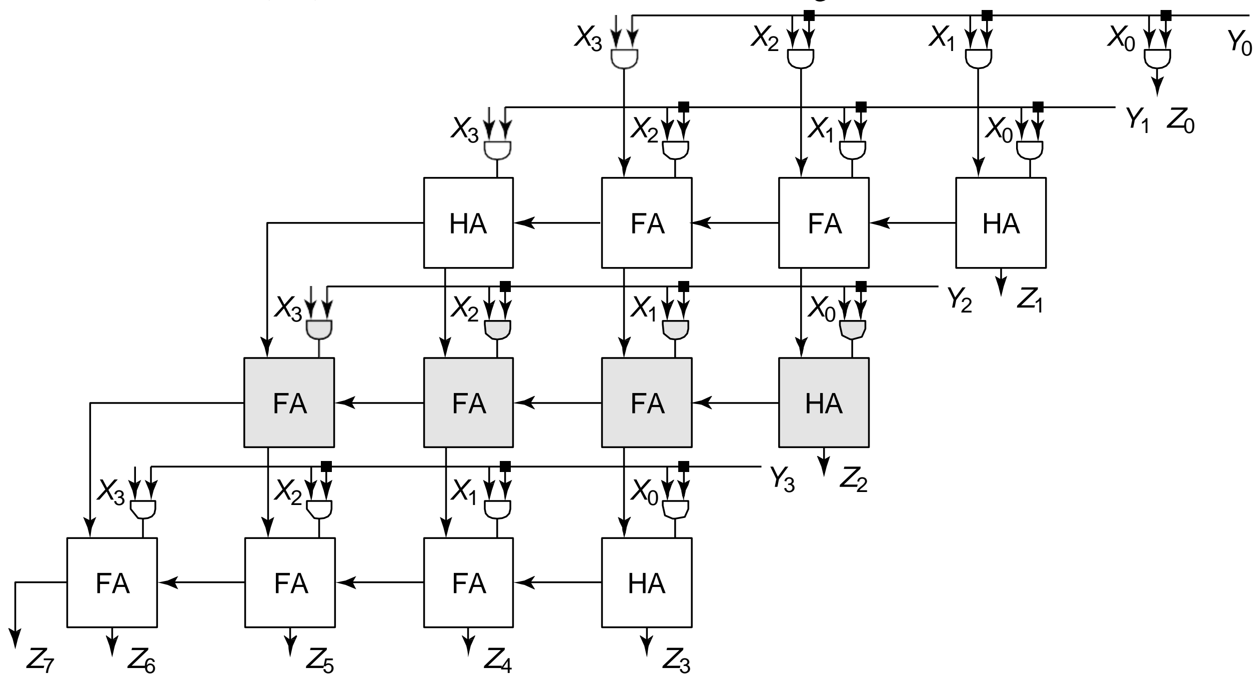


Figure: 4 x 4 array multiplier using Fulladder, Halfadder and AND gate.

(iv) Booth (encoding) multiplier:

- Booth's algorithm is an efficient hardware implementation of a digital circuit that multiplies two binary numbers in two's complement notation.
- Booth multiplication is a fastest technique that allows for smaller, faster multiplication circuits, by recoding the numbers that are multiplied.
- The Booths multipliers widely used in ASIC oriented products due to the higher computing speed and smaller area.
- In the binary number system, the digits called bits are to the set of {0,1}.
- The result of multiplying any binary number by binary bit is either 0 or original number.
- This makes the formation of partial products are more efficient and simple.
- Then adding all these partial products is time consuming task for any binary multipliers.
- The entire process consists of three steps partial product generation, partial product reduction and addition of partial products as shown in figure.

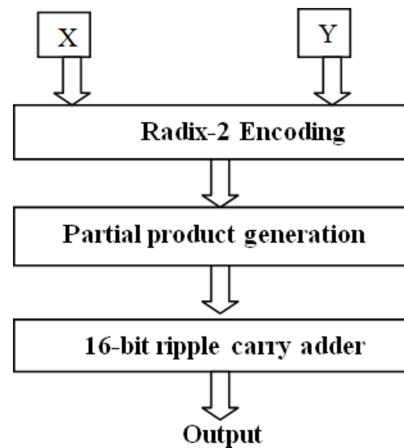


Figure: Block diagram of Booth multiplier

- But in booth multiplication, partial product generation is done based on recoding scheme e.g. radix 2 encoding.
- Bits of multiplicand (Y) are grouped from left to right and corresponding operation on multiplier (X) is done in order to generate the partial product.
- In radix-2 booth multiplication partial product generation is done based on encoding which is as given by Table.

Q_n	Q_{n+1}	Recoded Bits	Operation
0	0	0	Shift
0	1	+1	Add X
1	0	-1	Subtract X
1	1	0	Shift

Table: Booth encoding table with RADIX-2

19

- RADIX-2 PROCEDURE:

1) Add 0 to the LSB of the multiplier and make the pairing of 2 from the right to the left which shown in the figure.

11 1 1 1 0 0 0 1 1 0

Figure: 2- Bit pairing as per Booth recoding using Radix- 2.

2) 00 and 11: do nothing according to the encoding table.
3) 01: mark shows the end of the string' of 1and add multiplicand to the partial product.
4) 10: mark shows beginnings of the string of 1 subtract multiplicand from partial product.

With suitable example and with detailed steps explain Radix-4 modified booth encoding for an 8-bit signed multiplier. (Nov 2019)

Modified Booth Multiplier using Radix -4:

- **The disadvantage of Booth Multiplier with Radix-2 is increasing partial products.**
- **Modified Booth Multiplier with Radix-4 is reducing the half of the partial products in multipliers.**
- Modified Booth multiplication is a technique that allows for smaller, faster circuits by recoding the numbers that are multiplied.
- In Radix-4, encoding the multiplicands based on multipliers bits. It will compare 3-bits at a time with overlapping technique.
- Grouping starts from the LSB and the first block contains only two bits of the multipliers and it assumes zero for the third bit.

1 1 1 0 0 0 1 1 0

Figure. Grouping of 3-bit as per booth recoding

- These group of binary digits are according to the Modified Booth Encoding Table and it is one of the numbers from the set of (-2,2,0,1,-1).

groups	Partial products
000	0
001	1*multiplicand
010	1*multiplicand
011	2*multiplicand
100	-2*multiplicand
101	-1*multiplicand
110	-1*multiplicand
111	0

Table: Booth encoding table with RADIX-4

Multiplier Bits Block			Recoded 1-bit pair		2 bit booth	
i+1	i	i-1	i+1	i	Multiplier Value	Partial Product
0	0	0	0	0	0	Mx0
0	0	1	0	1	1	Mx1
0	1	0	1	-1	1	Mx1
0	1	0	1	0	2	Mx2
1	0	0	-1	0	-2	Mx-2
1	0	1	-1	1	-1	Mx-1
1	1	0	0	-1	-1	Mx-1
1	1	0	0	0	0	Mx0

x_{i+1}	x_i	x_{i-1}	$z_{i/2}$	Explanation
0	0	0	0	No string of 1s in sight
0	0	1	1	End of string of 1s
0	1	0	1	Isolated 1
0	1	1	2	End of string of 1s
1	0	0	-2	Beginning of string of 1s
1	0	1	-1	End a string, begin new one
1	1	0	-1	Beginning of string of 1s
1	1	1	0	Continuation of string of 1s

Example

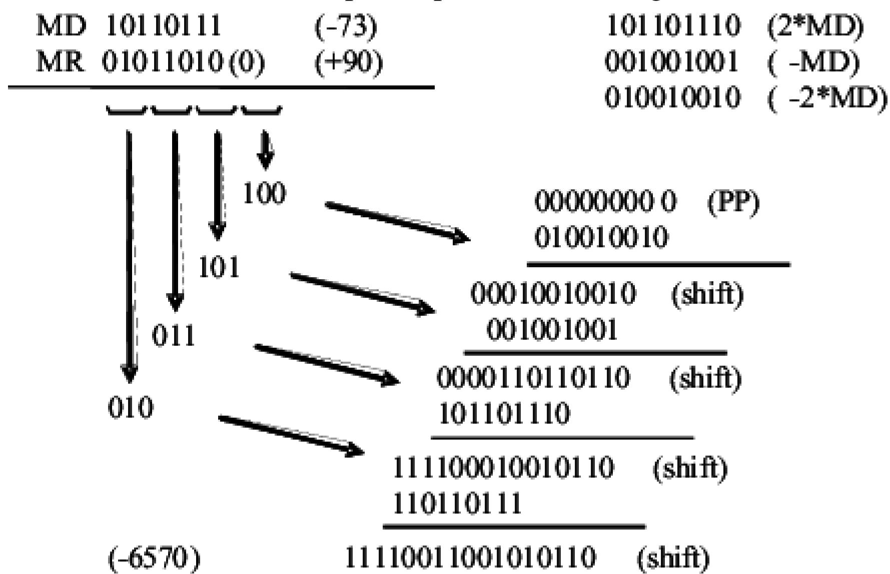
1 0 0 1 1 1 0 1 1 0 1 0 1 1 1 0 (0) Operand x

-2 2 -1 2 -1 -1 0 -2 Radix-4 version z

21

A		1	0	1	1				−5
X	×	1	1	0	1				−3
Y		0	$\bar{1}$	1	$\bar{1}$				recoded multiplier
Add $-A$		0	1	0	1				
Shift		0	0	1	0	1			
Add A	+	1	0	1	1				
		1	1	0	1	1			
Shift		1	1	1	0	1	1		
Add $-A$	+	0	1	0	1				
		0	0	1	1	1	1		
Shift		0	0	0	1	1	1	1	

- **RADIX-4 PROCEDURE: [May 2021 (Model)]**
  1) Add 0 to the right of the LSB of the multiplier.
  2) Extend the sign bit 1 position if it is necessary when n is even.
  3) Value of each vector,the partial product is coming from the set of (-2,2,0,1,-1).



(v) Wallace tree Multiplier:

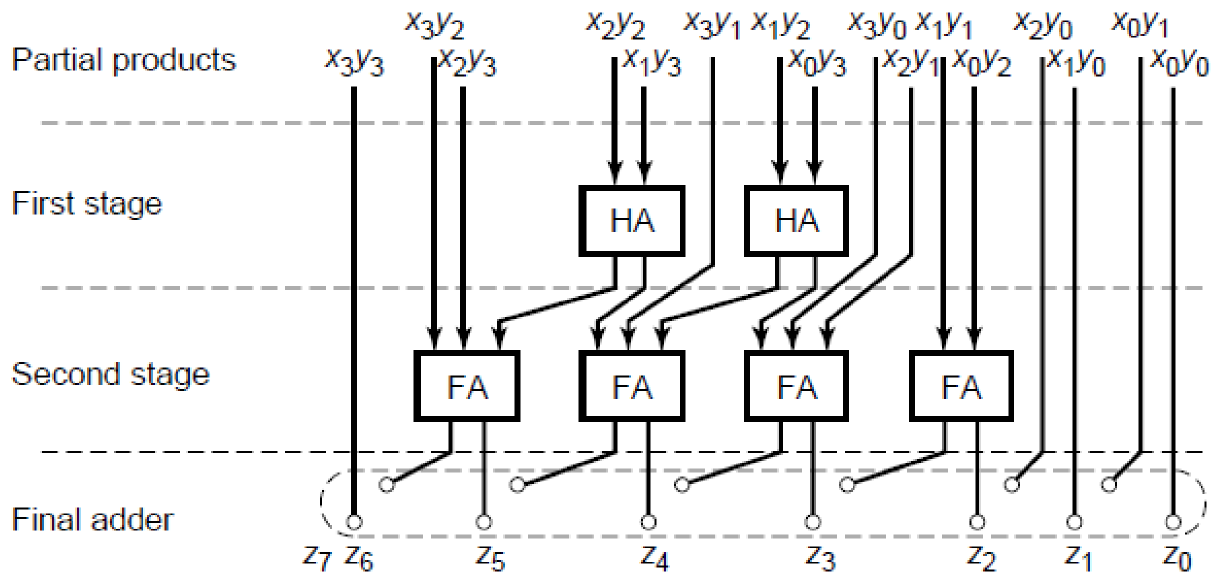
- A Wallace tree is an efficient hardware implementation of a digital circuit that multiplies two integer numbers.
- The Wallace tree multiplier has three steps to be followed,
  (a) Multiply each bit of one of the arguments, by each bit of the other, yielding n^2 results.
  (b) Reduce the number of partial products to two by layers of full and half adders.
  (c) Group the wires in two numbers and add them with a conventional adder.
- The second section works as follows,
  (a) Take any three wires with same weights and input them into a full adder. The result will be an output wire of the same weight and an output wire with a higher weight for each three input wires.

22

(b) If there are two wires of the same weight left, input them into a half adder.

(c) If there is just one wire left and connects it to next layer.

- The Wallace tree multiplier output structure is tree basis style. It reduces the number of components and reduces the area.
- The architecture of a 4 x4 Wallace tree multiplier is shown in figure.



❖ **Apply radix-2 booth encoding to realize a 4-bit signed multiplier for (-10)*(-11). (April 2019-15M) [Apr/May 2022][Nov/Dec 2022]**

Solution:

M= -10 =0110, Q= -11 =0101

	A	Q	Q_{-1}	
Step-I:	0000	0101	0	:last 2 bits are10; A=A-M
	1010	0101	0	: shift right
	1101	0010	1	
Step-II:	0011	0010	1	:last 2 bits are 01; A=A+M
	0001	1001	0	:shift right
Step-III:	1011	1001	0	:last 2 bits are10; A=A-M
	1101	1100	1	;shift right
Step-IV:	1101	1100	1	;last 2 bits are 01; A=A+M
	0110	1110	0	;shift right

****

4.7. DIVIDERS

Explain in detail about the design and procedure for dividers.

- There are two types of dividers, Serial divider and Parallel divider. Serial divider is slow and parallel divider is fast in performance.

23

- Generally division is done by repeated subtraction. If 10/3 is to be performed then,

  10 -3 =7, (divisor is 3, dividend is 10)

  7 – 3 = 4,

  4 – 3 = 1

- Here, repeated subtraction has been done, after 3 subtractions, the remainder is 1. It is less than divisor. So now the subtraction is stopped.
- Let see the example of binary division with use of 1's complement method

  1010 (10_d) / 0011 (3_d)

  Step1: find 1's complement of divisor

  Step2: add this with the dividend

  Step3: if carry is 1, then it is added with the output to get the difference output

  Step4: the same procedure is repeated until we are get carry 0.

  Step5: then the process is stopped.

```
                    1 0 1 0 (10)
(1's complement of 3) 1 1 0 0 +
              [1] 0 1 1 0
                        1
                  0 1 1 1 (7)
```

Carry is 1, so, it is added with the o/p.

10 – 3 = 7

```
      0 1 1 1 (7)
      1 1 0 0 +
  [1] 0 0 1 1
            1
      0 1 0 0 (3)
```

Carry is 1, so, it is added with the o/p.

7 – 3 = 4

```
      0  1  0  0  (4)
      1  1  0  0  +
 [1]  0  0  0  0
  └──────────► 1
      0  0  0  1  (1)
```

4 – 3 = 1

Carry is 1, so, it is added with the o/p.

```
      0  0  0  1  (1)
      1  1  0  0  +
      1  1  0  1
```

There is no carry, so, the process is stopped.

Quotient = 3

Remainder = 0 0 0 1 (Final Difference)

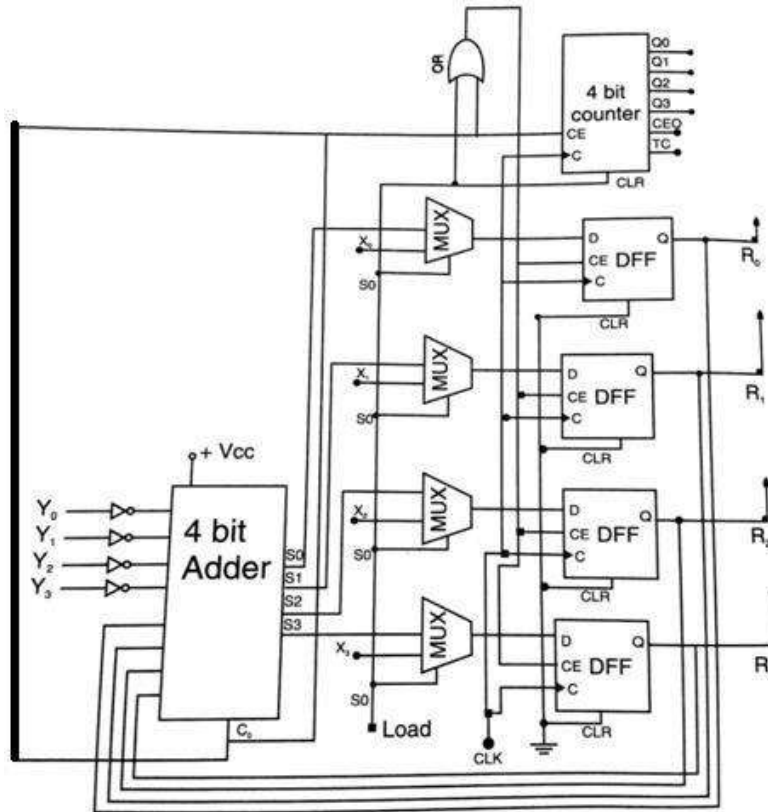
The implementation of this division is given below.

$$X \div Y$$

$$X_3\, X_2\, X_1\, X_0 \div Y_3\, Y_2\, Y_1\, Y_0$$

- Basic building blocks of serial adder are given below.
  1. 4 bit adder
  2. 4 bit binary up counter
  3. 2:1 MUX (4 MUXs are used)
  4. D flipflop

- Y_0 Y_1 Y_2 Y_3 are complemented and given to 4 bit adder block (figure shown below)
- X_0 X_1 X_2 X_3 are given to MUXs and MUX output is given to D flipflop. Select signal of MUX is high. It is connected to clear input of counter.
- Carry output of adder is connected with clock enable pin of counter. The same is given to OR gate. The output of this OR gate is given to clock enable signal of flipflops.
- The other input of OR gate is tied with select signal of MUX.
- If X > Y, C_0 of adder is high.
- After first subtraction, the counter output is incremented by 1.
- For each subtraction, the counter output is incremented.
- If C_0 of adder is low, then clock of counter and FF is disabled. Counting is stopped.
- Q_3 Q_2 Q_1 Q_0 is the counter output (Quotient)
- R_3 R_2 R_1 R_0 is the flipflop output (remainder)

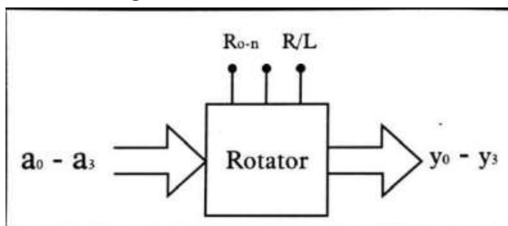


****

4.8. SHIFT REGISTERS:

Design 4 input and 4 output barrel shifter using NMOS logic. (NOV 2018, Nov 2019).
List the several commonly used shifters. Design the shifter that can perform all the commonly used shifters. [May 2021, NOV 2021]
Elaborate in detail the design of a 4-bit barrel shifter. [Nov/Dec 2022]

- An n-bit rotation is specified by using the control word R_{0-n} and L/R bit defines a left or right shifting.



- For example $y_3\ y_2\ y_1\ y_0 = a_3\ a_2\ a_1\ a_0$

  If it is rotated 1-bit in left side, we get $y_3\ y_2\ y_1\ y_0 = a_2\ a_1\ a_0\mathbf{a_3}$

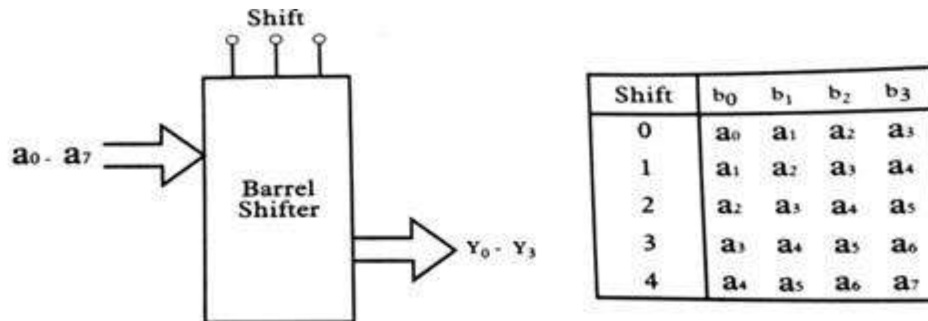
  If it is rotated 1-bit in right side, we get $y_3\ y_2\ y_1\ y_0 = \mathbf{a_0}\ a_3\ a_2\ a_1$

Barrel Shifter:

- A barrel shifter is a digital circuit that can shift a data word by a specified number of bits in oneclock cycle.

26

- It can be implemented as a sequence of multiplexers(MUX), and in such an implementation the output of one MUX is connected to the input of the next MUX in a way that depends on the shift distance.
- For example, take a four-bit barrel shifter, with inputs A, B, C and D. The shifter can cycle the order of the bits ABCD as DABC, CDAB, or BCDA; in this case, no bits are lost.
- That is, it can shift all of the outputs up to three positions to the right (thus make any cyclic combination of A, B, C and D).
- The barrel shifter has a variety of applications, including being a useful component in microprocessors (alongside the ALU).



Shift	b_0	b_1	b_2	b_3
0	a_0	a_1	a_2	a_3
1	a_1	a_2	a_3	a_4
2	a_2	a_3	a_4	a_5
3	a_3	a_4	a_5	a_6
4	a_4	a_5	a_6	a_7

Figure: 8 X 4 barrel shifter

- General symbol for barrel shifter is shown in figure. The outputs are given as y_3 y_2 y_1 y_0. S_0, S_1, S_2,S_3 are known as shift lines.
- A barrel shifter is often implemented as a cascade of parallel 2×1 multiplexers.
- For a 8-bit barrel shifter, two intermediate signals are used which shifts by four and two bits, or passes the same data, based on the value of S[2] and S[1].
- This signal is then shifted by another multiplexer, which is controlled by S[0].
- A common usage of a barrel shifter is in the hardware implementation of **floating-point arithmetic**.

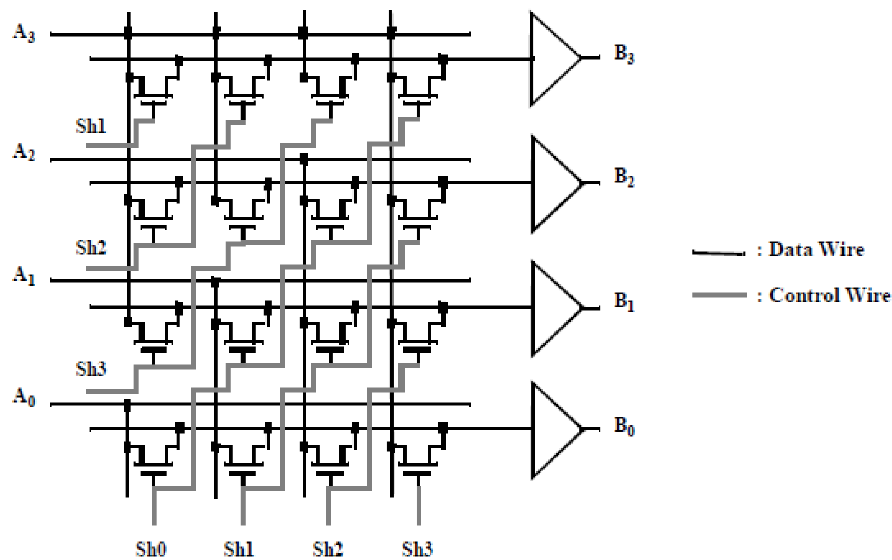


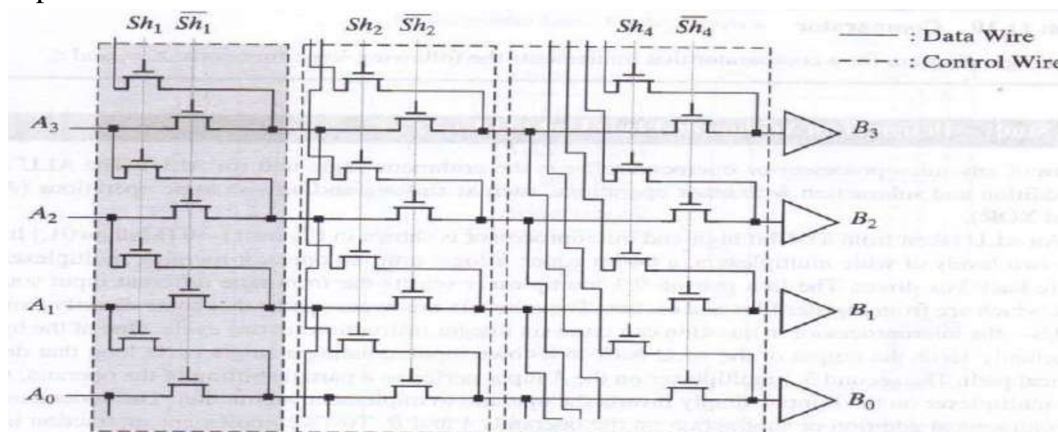
Figure: Barrel Shifter

27

- For a floating-point add or subtract operation, requires shifting the smaller number to the right, increasing its exponent, until it matches the exponent of the larger number.
- This is done by using the barrel shifter to shift the smaller number to the right by the difference, in one cycle.
- If a simple shifter were used, shifting by n bit positions would require n clock cycles.
- The disadvantages of FET array barrel shifter are the threshold voltage drop problem, parasitic limited switching time problem.
- The figure shown is known as a barrel shifter and a 8 x 4-bit barrel shifter circuit.

Logarithmic Shifter:

- A Shifter with a maximum shift width of M consists of a $\log_2 M$ stages, where the ith stage either shifts over 2^i or passes the data unchanged.
- Maximum shift value of seven bits is shown in figure, to shift over five bits, the first stage is set to shift mode, the second to pass mode and the last again to shift.
- The speed of the logarithmic shifter depends on the shift width in a logarithmic wa, M-bit shifter requires $\log_2 M$ stages.
- The series connection of pass transistors slows the shifter down for larger shift values.
- Advantage of logarithmic shifter is more effective for larger shift values in terms of both area and speed.



****

4.9. SPEED AND AREA TRADE OFF:

Discuss the details about speed and area trade off. (May 2017)
Discuss trade-off between speed Vs area. [Nov/Dec 2022]

Adder:

- The tradeoff in terms of power and performance is shown below.
- The performance is represented in terms of the delay(speed).
- The area estimations for each of the delays are given based on the fact that area is in relation to the power consumption.
- The area of a carry lookahead adder is larger than the area of a ripple carry for a particular delay.

- This is because the computations performed in a carry lookahead adder are parallel, which requires a larger number of gates and also results in a larger area.

CLA –Carry Lookahead Adder, RC, R – Ripple carry adder

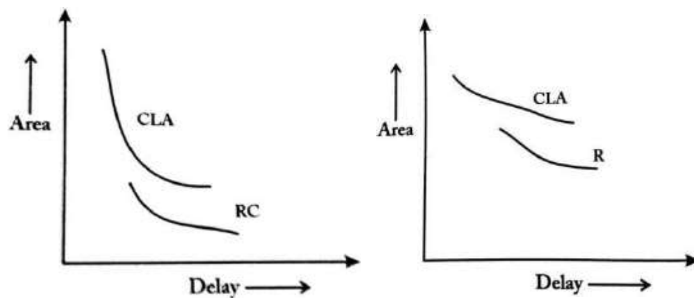


Figure: Area Vs Delay for 8 bit adder Figure: Area Vs Delay for 16 bit adder

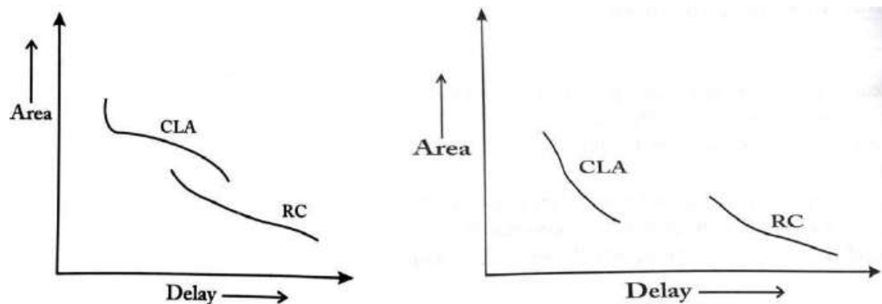


Figure: Area Vs delay for 32 bit adder Figure: Area Vs delay for 64 bit adder

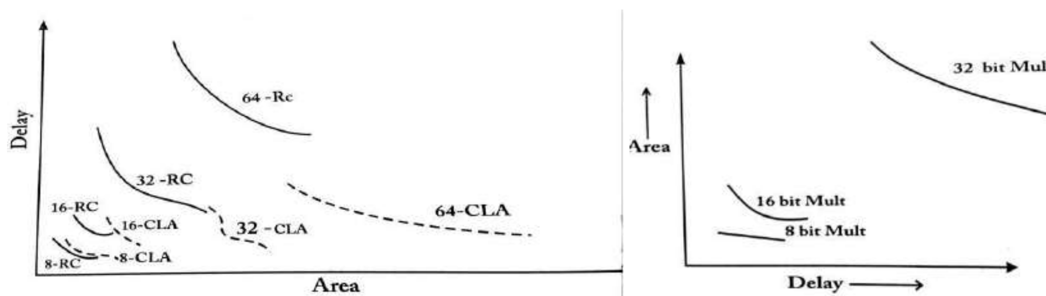


Figure: Delay Vs Area for all adders **Figure: Area Vs Delay for all multiplier**

- Above figures shows that the delay of the ripple carry adder increases much faster when compared to the carry lookahead adder as the number of bits is increased.
- In the carry lookahead adder, the cost is in terms of the area because computations are in parallel, and therefore more power is consumed for a specific delay.

A simple Data Path Summary

Architecture type	Voltage	Area	Power
Simple datapath (no pipelining or parallelism)	5V	1	1
Pipelined datapath	2.9V	1.3	0.37
Parallel datapath	2.9V	3.4	0.34

29

4.10 Memory Architecture and Memory Control Circuits:

Discuss Memory classification and its architecture and building blocks.

4.10.1 Memory Classification:

- Parameters used to characterize a memory device are area, power and speed.
- **Area:** area is important for its physical implementations by VLSI technology. Smaller the area per bit and more devices can be accommodated. So cost per bit is reduced.
- **Speed:** speed of operation plays a very important role. Memory can communicate at speed with processors.
- **Power:** Power is important, because MOS memories are used in many battery operated portable systems. Power dissipation of memory plays an important role. Memory devices will consume less power.

Classification based on operation mode:

1. ROM
2. RAM

Classification based on data storage mode:

It means on how it is stored and how long it remains there.

1. Volatile
2. Non-volatile

- Volatile memory devices will store information, as long as power is it.
- As soon as power is turned off, information is lost. Static RAM and dynamic RAM belong to the category of volatile memory.
- EPROM and mask programmable ROM are non-volatile memory devices.
- If the power is turned off information will not be lost.

Classification based on access method:

1. Random access
2. Non-random access

4.10.2 Memory Architecture and Building Blocks:

Explain the *memory architecture* and its control circuits in detail. (April 2018)

Illustrate the building blocks of Memory architectures and memory peripheral circuitry adapted to operate for non-volatile memory. [May 2021]

When n x m memory is implemented, then, n memory words are arranged in a linear fashion. One word will be selected at a time by using select line.

- If we want to implement the memory 8X8, n=8, m=8(number of bits).
- Then we need 8 select signals (one for each word).
- But by using decoder we can reduce the number of select signals.
- In case of 3 to 8 decoder, if 3 inputs are given to decoder, then we can get 8 select signals.
- If n=220, then we can give only 20 inputs to the decoder.

30

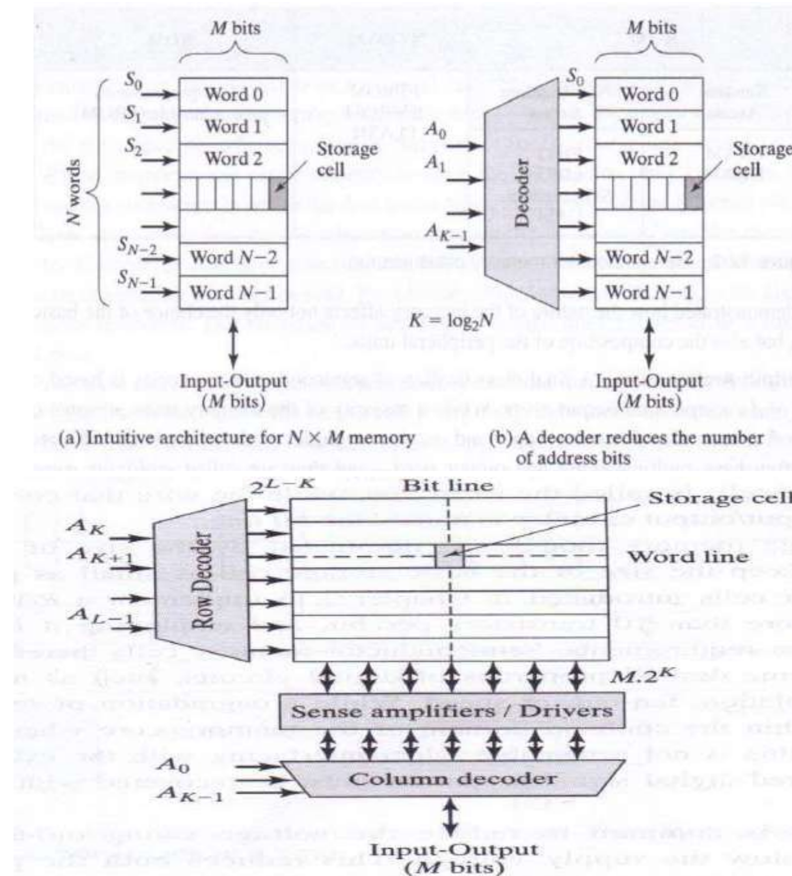


Figure: Array structured memory organization

- If basic storage cell size is approximately square, then the design is extremely slow. The vertical wire, which connects the storage cells to I/O will be excessively large.
- So, memory arrays are organized in such a way that vertical and horizontal dimensions are the same.
- The words are stored in a row. These words are selected simultaneously.
- The **column decoder** is used to route the correct word to the I/O terminals.
- The row address is used to select one row of memory and column address is used to select particular word from that selected row.
- **Word line**: The horizontal select line which is used to select the single row of cell is known as word line.
- **Bit line**: The wire which connects the cell in a single column to the input/output circuit is known as bit line.
- **Sense amplifier**: It requiresanamplificationoftheinternalswingtofullrail-to-rail amplitude.
- **Block address**: the memory is divided into various small blocks.
- The address which is used to select one of the small blocks to be read or written is known as block address.
- **Advantages:**
  1. Access time is fast

2. Power saving is good, because blocks not activated are in power saving mode.

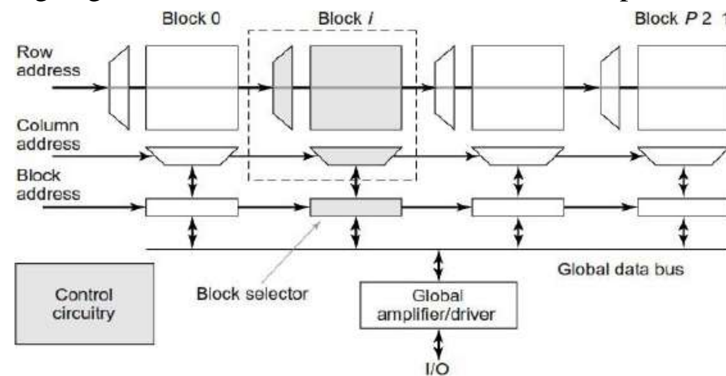


Figure: Hierarchical memory architecture

4.10.3 Memory Core

Discuss Memory core its types in detail.

4.10.3.1 Read Only Memory (ROM):

ROM is a memory where code is written only one time.

Diode ROM:

- It is simple where presence of diode in between bit line and word line is considered as logic 1 and absence of diode as logic 0.
- Disadvantage is used for small memories and no isolation between word line and bit line.

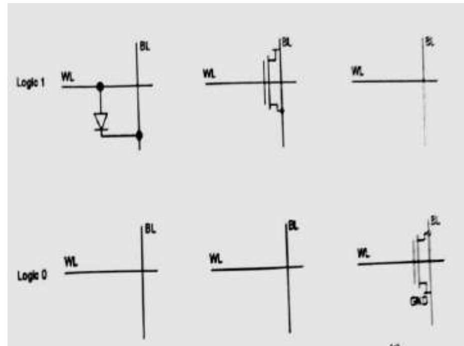


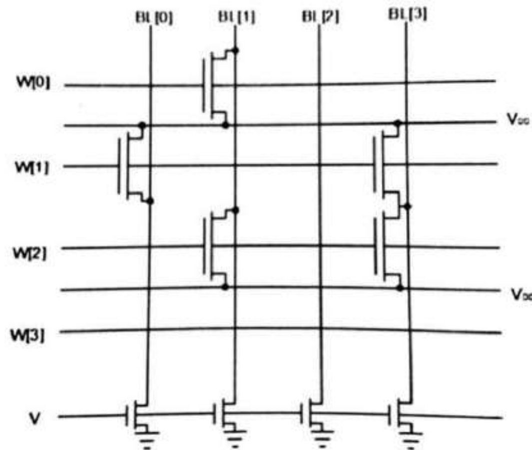
Figure: Diode ROM

MOS ROM:

Q: Draw the NOR and NAND implementation of 4-word, 4-bit ROM. (NOV 2021)

- Diode is replaced by gate source connection of nMOS. Drain is connected to V_{DD}.
- The charging and discharging of word line capacitance has been taken care by the word line driver.
- Absence of a transistor between word line and bit line means logic 1 is stored and if presence then logic 0 is stored.

32



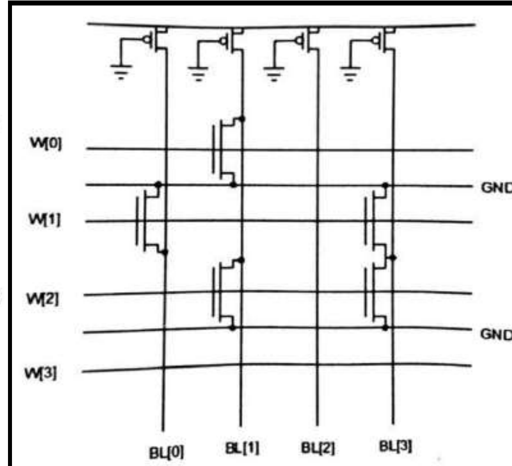


Figure: 4 x 4OR ROM cell array

Figure: 4 x 4 MOS NOR ROM

Programming ROM

- The transistor in the intersection of row and column is OFF when the associated word line is LOW. In this condition, we get logic 1 output.

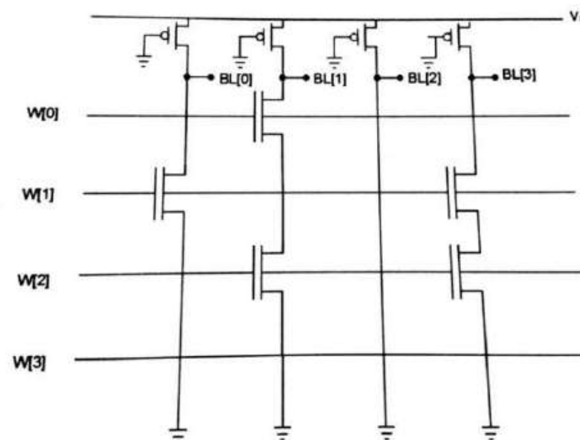


Figure: 4 x 4 MOS NAND ROM

Advantage: basic cell only consists of transistor. No need of connection to any of the supply voltage.

Disadvantage: As it has pseudo nMOS, it is ratioed logic and consumes static power.

To overcome this, precharged MOS NOR ROM logic circuit is used.

- This eliminate static dissipation ratioed logic requirement.

4.10.3.2 Non-Volatile READ-WRITE Memory:

- It consists of array of transistors. We can write the program by enabling or disabling these devices selectively.
- To reprogram, the programmed values to be erased, then the new programming is started.

Floating gate transistor:

- It is mostly used in all the reprogrammable memories.

33

- In floating gate transistor, extra polysilicon strip is used in between the gate and the channel known as floating gate.
- Floating gate doubles the gate oxides thickness and hence device transconductance is reduced and threshold voltage is increased.
- The threshold voltage is a programmable.
- If high voltage is (>10V) is applied between the source terminals and gate-drain terminals, then high electric field is generated. So, avalanche injection occurs.
- After acquiring energy, electron becomes hot and transverse through the first oxide insulator . They get trapped on the floated gate.
- The floating gate transistor is known as floating gate avalanche injection MOS or FAMOS.

Disadvantage: High programming voltage is need.

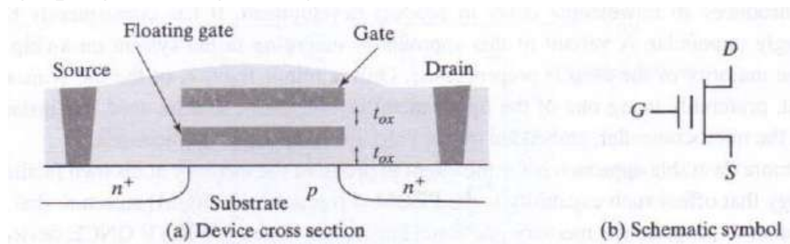


Figure (a) floating gate transistor (b) symbol

EPROM – Erasable Programmable Read Only Memory:

- Erasing is done by passing UV rays on the cell by using transparent window.
- This process will take some seconds to some minutes.
- It depends on intensity of UV source. The programming takes 5-10microseconds/word.
- During programming, chip is removed from the board and placed in EPROM programmer.

Advantages: simple and large families are fabricated with low cost.

Disadvantages:

- Number of erase/program cycle is limited upto 1000.
- Reliability is not good.
- Threshold voltage of the device may be varied with repeated program.

EEPROM – E²PROM:

- Electrically Erasable Programmable ROM. Here Floating gate tunneling oxide (FLOTOX) is used.
- It is similar to floating gate except that the portion of the floating gate is separated from the channel at the thickness of 10nm or <10nm.
- If 10V is applied, electron travels to and from the floating gate through Fowler-Nordheim tunneling.
- Erasing can be done by revering applied voltage which is used for writing.

34

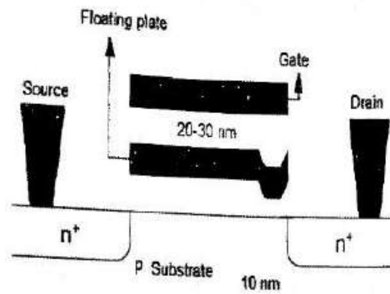


Figure: FLOTOX transistor

Advantage: High versatility and possible for 105 erase/write cycle.

Disadvantages: Larger than FAMOS transistor, Costly, Repeated programming causes a drift in threshold voltage.

Flash Memory – Flash Electrically Erasable Programmable ROM

- It is a combination of density of EPROM and versatility of EEPROM.
- Avalanche hot electron injection mechanism is used.
- Erasing can be done by Fowler-Nordheim tunneling concept. Here erasing is done in bulk.

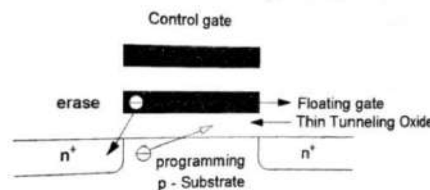


Figure: ETOX device

- It is similar to FAMOS gate.
- A very thin tunneling oxide layer (10nm thickness) is there.
- ***Erasing operation:*** Erasing can be performed when gate is connected to the ground and the source is connected to 12V.
- ***Write operation:*** High voltage pulse is applied to the gate of the selected device. Logic 1 is applied to the drain and hot electrons are injected into the floating gate.
- ***Read operation:***To select a cell, its word line is connected to 5V. It causes conditional discharge of the bit line.

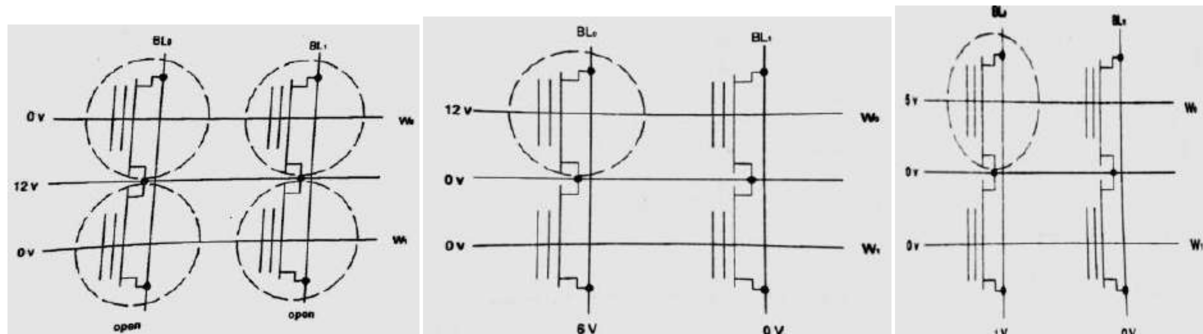


Figure: (a) Erase (b) Write (c) Read operation of NOR flash memory

35

4.10.3.3 RAM – Random Access Memory

Explain about static and dynamic RAM.
Construct 6T based SRAM cell. Explain its read and write operations. (NOV 2018) [Nov/Dec 2022]

4.10.3.3.1 Static RAM:

- SRAM cell needs 6 transistors per bit.
- M5 and M6 transistors are shared between read and write operations.
- Bit line(BL) and inverse Bit Line signals are used to improve the noise margin during read and write operations.

***Read operation**:*

- Let us assume logic 1 is stored at Q and BL and inverse BL are precharge to 2.5V before starting read operation.
- The read cycle is started by asserting word line then M_5 and M_6 transistors are enabled.
- After the small initial word line delay then the values stored at Q and inverse Q are transferred to the bit lines by leaving BL at 2.5V and the value at inverse Q is discharge through M_1, M_5.

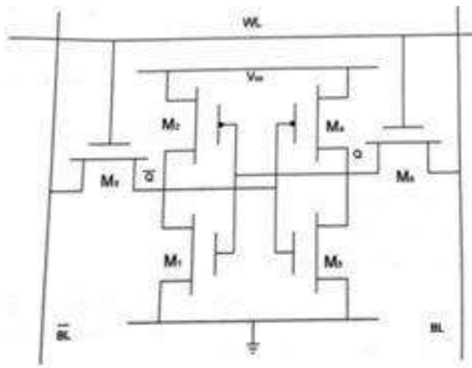


Figure: CMOS SRAM cell

***Write operation**:*

- Assume that Q=1, now logical 0 is to be written in the cell.
- Then inverse BL is set to 1 and BL is set to 0.
- The gate of M_1 is at V_{DD} and gate of M_4 is at ground as long as the switching is not commenced.
- Inverse Q is not pulled high enough to ensure the writing of logic 1.
- Cell voltage is kept below 0.4V. The new value of the cell is written through M_6.

4.10.3.3.2 Dynamic RAM:

Three transistors DRAM

- Content in the cell can be periodically rewritten through a resistive load, called as refresh operation.
- This refresh occurs for every 1-4ms. Dynamic memory has refresh operation.
- For example, logic 1 is to be written, and then BL1 is asserted high and write wordline(WWL) is asserted.
- This data is retained as charge on the capacitor once WWL is low.

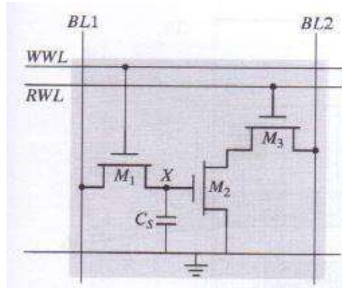


Figure: Three transistor dynamic memory cell

- To read the cell, the read word line (RWL) is raised. M_2 transistor is either ON or OFF depends upon the stored value.
- BL2 bit line is connected to V_{DD} or it is precharged to V_{DD} or V_{DD}-V_t.
- When logic 1 is stored, the series combination of $M_2$and M_3 pulls BL2 line low.
- If logic 0 is stored, then BL2 line is high.
- To refresh the cell, first the stored data is read, and its inverse is placed on BL1 and WWL line is asserted.

One transistor DRAM:

- In this cell, to write logic 1 then it is placed on bit line and word line is asserted high.
- The capacitor is charged or discharged depending upon the data. Before performing read operation, bit line is precharged.

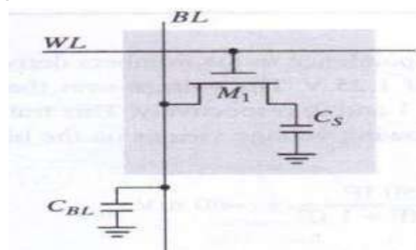


Figure: One transistor DRAM

4.10.3.3.3 CAM – Content Addressable or Associate Memory

Explain about CAM.

- It supports 3 operating modes,
  - Read
  - Write
  - Match
- In this memory, it is possible to compare all the stored data in parallel with the incoming data. It is not power efficient.
- Figure shows apossible implementation ofaCAM array.
- Thecellcombines atraditional 6TRAM storage cell *(M_4-M_9)*withadditional circuitry toperform al-bit digital comparison *(M_1-M_3)*.
- When thecellistobe written,complementarydataisforced ontothebitlines, whiletheword lineisenabled as inastandardSRAM cell.

37

- In the compare mode, stored data are compared using bit line. The match line is connected to all CAM blocks in a row. And it is initially precharged to V_{DD}.
- If there is some match occurs, then internal row is discharged. If even one bit in a row is mismatched, then the match line is low.

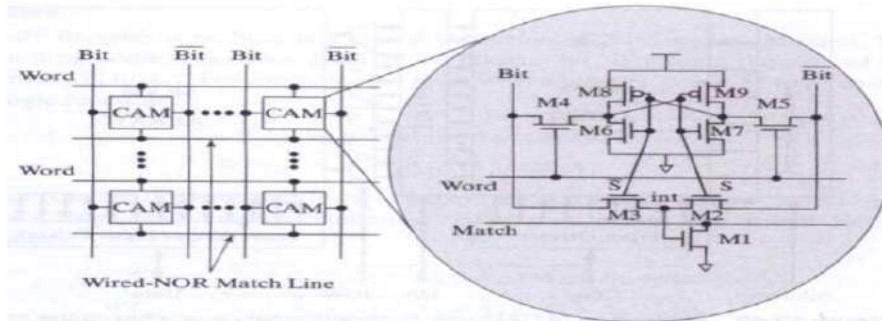


Figure: CAM cell

****

4.11 Memory peripheral (control) Circuits:

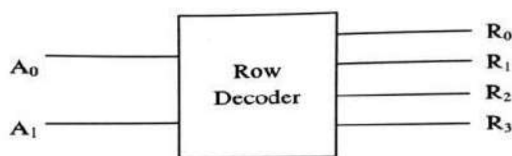
Explain the memory architecture and its *control circuits* in detail. (April 2018)

Illustrate the building blocks of Memory architectures and memory peripheral circuitry adapted to operate for non-volatile memory. [May 2021]

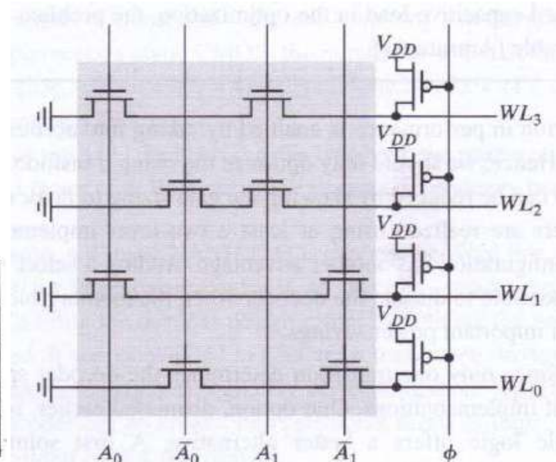
(i) Address & Block Decoders:

Row Decoder:

- Row and column address decoder are used to select the particular memory location in an array.
- Row decoder is used to drive NOR ROM array. It selects one of 2^n word lines.
- Dynamic 2 to 4 decoder reduces the number of transistors and propagation delay.



A_0	A_1	R_0	R_1	R_2	R_3
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0
1	1	0	0	0	0



Symbol and Truth table

Dynamic2-to-4NORdecoder

Column Decoder

- It should matchthebitlinepitchofthememory array.

- In column decoder, decoder outputs are connected to nMOS pass transistors.
- By using this circuit, we can selectively drive one out of m pass transistors.
- Only one nMOS pass transistor is ON at the time.

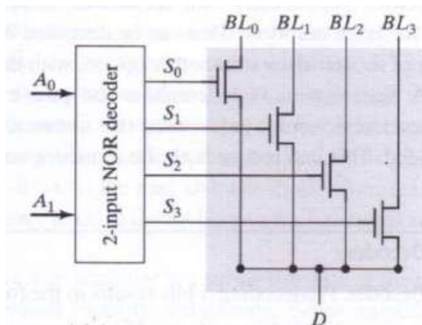
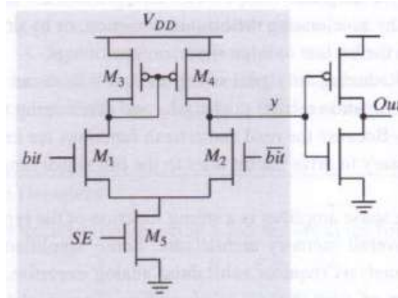


Figure: Four-input pass-transistor-based column decoder using a NOR predecoder

(ii) Sense Amplifier

- Sense amplifiers play a major role in the functionality, performance and reliability of memory circuits.
- Basic differential sense amplifier circuit shown in below figure.
- It performs the following performances



Amplification:

- In memory structures such as the 1TDRAM, amplification is required for proper functionality.

Delay Reduction:

- The amplifier compensates for the fan-out driving capability of the memory cell by detecting and amplifying small transitions on the bit line to large signal output swings.

Power reduction:

- Reducing the signal swing on the bit lines can eliminate large part of the power dissipation related to charging and discharging the bit lines.

(iii) Drivers/ Buffers

- The length of word and bit lines increases with increasing memory sizes.
- Large portion of the read and write access time can be attributed to the wire delays.
- A major part of the memory-periphery area is allocated to the drivers (address buffers and I/O drivers).

****

39

4.12: Low Power Memory design:

- ❖ **Discuss about Low power memory design. [Apr/May 2022]**
- ❖ **Elucidate in detail low power SRAM circuit. (April 2019-13M) (Nov 2019)**

(i) Active Power Reduction:

- Voltage reduction done by either an increase in the size of the storage capacitor and/or a noise reduction.

Techniques for power reductions:

• Half- V_{DD} precharge:

- ▪ Precharging a bit line to $V_{DD}/2$. It helps to reduce the active power dissipation in DRAM memories by a factor of 2.

• Boosted word line:

- ▪ Raising the value of the word line above V_{DD} during a write operation, eliminates the threshold drop over the access transistor, yielding a substantial increase in stored charge.

• Increased capacitor area or value:

- ▪ Keeping the "ground" plate of the storage capacitor at $V_{DD}/2$ reduces the maximum voltage over Cs, making it possible to use thinner oxides.

• Increasing the cell size:

- ▪ Ultra-low-voltage DRAM memory operation might requirea sacrifice in area efficiency.

(ii) Retention current Reduction:

- SRAM array should not have any static power dissipation. But the leakage current of the transistor will be the major problem and this is the main source of the retention current.
- This retention current can be reduced by the following factors.
  1. Turn OFF unused memory blocks
  2. Negative biasing voltage of the cells which are not active, thus reduce the leakage current.
  3. If low threshold voltage transistor is inserted between V_{DD} and SRAM array, leakage reduces.
  4. Leakage is a function of V_{DD}, thus if supply rail is lowered, then leakage current is reduced.

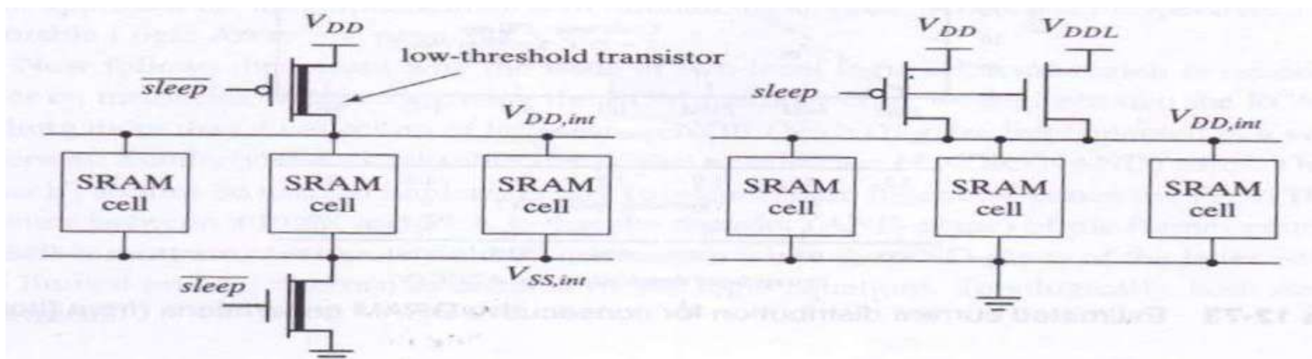


Figure: (a) Insertion of low threshold device (b) Reducing supply Voltage

****

40

Programmable devices (Programmable ASIC):

- Programmable devices can be divided into three areas
    1. Programmable logic structure
    2. Programmable interconnect
    3. Reprogrammable Gate array
- A programmable logic device (PLD) is an electronic component used to build reconfigurable digital circuits.
- Unlike a logic gate, which has a fixed function, a PLD has an undefined function at the time of manufacture.

1. Programmable Logic Structure:

- ❖ **Describe in detail the chip with programmable logic structures. (Nov 2009)**

(a) Programmable Logic Array:

- Programmable logic arrays (PLAs) is a type of fixed architecture logic devices with *programmable AND gates followed by programmable OR array.*
- Logic array is the structure unit which can be programmed to perform various functions.
- Programmable Logic Array (PLA) can be implemented as AND-OR plane devices.
- Structure of AND-OR PLA is shown below.

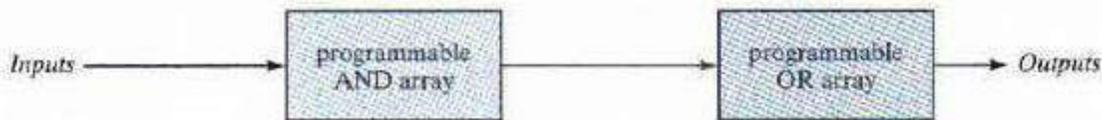


Figure: Programmable logic array

- PLA is used to implement a complex combinational circuit.
- The AND and OR gates inside the PLA are initially fabricated with fuses among them.
- The specific Boolean functions are implemented in sum of products (SOP) form by blowing appropriate fuses and leaving the desired connections.
- For an example, the Boolean expressions are,

$$F_1 = A\bar{B} + AC + \bar{A}B\bar{C}$$

$$F_2 = \overline{(AC + BC)}$$

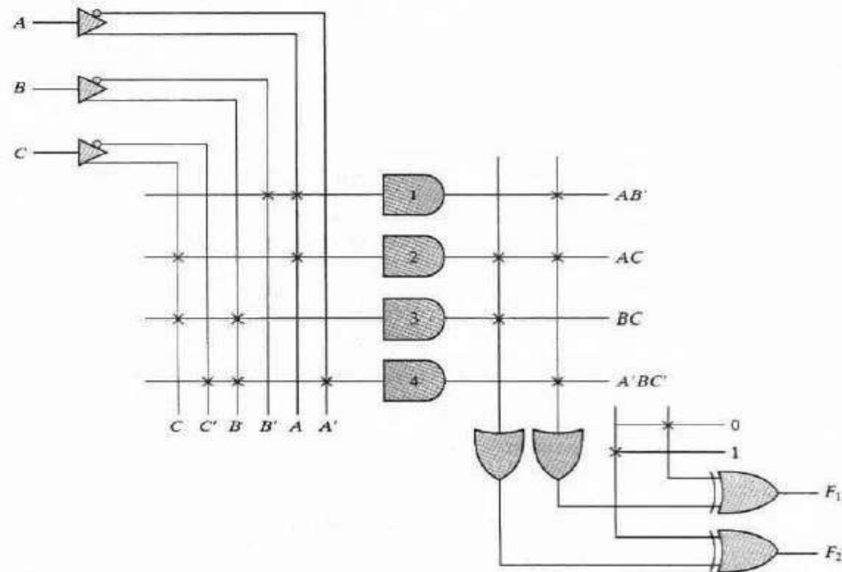


Figure: PLA with three inputs, four product terms and two outputs

(b) PAL (Programmable Array Logic) Architecture:

- The PAL is a programmable logic device with a *fixed OR array and a programmable AND* array.

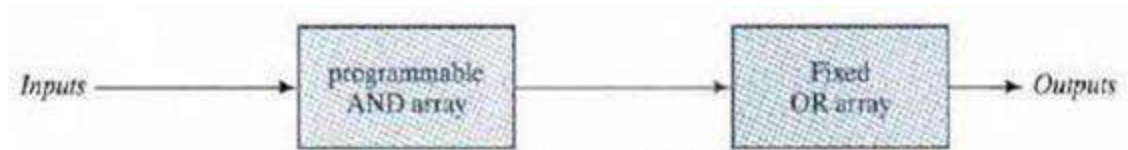


Figure: Programmable Array Logic

- Because only the AND gates are programmable, the PAL is easier to program than but is not as flexible as the PLA.
- The PAL is a programmable logic device with a fixed OR array and a programmable AND array.

42

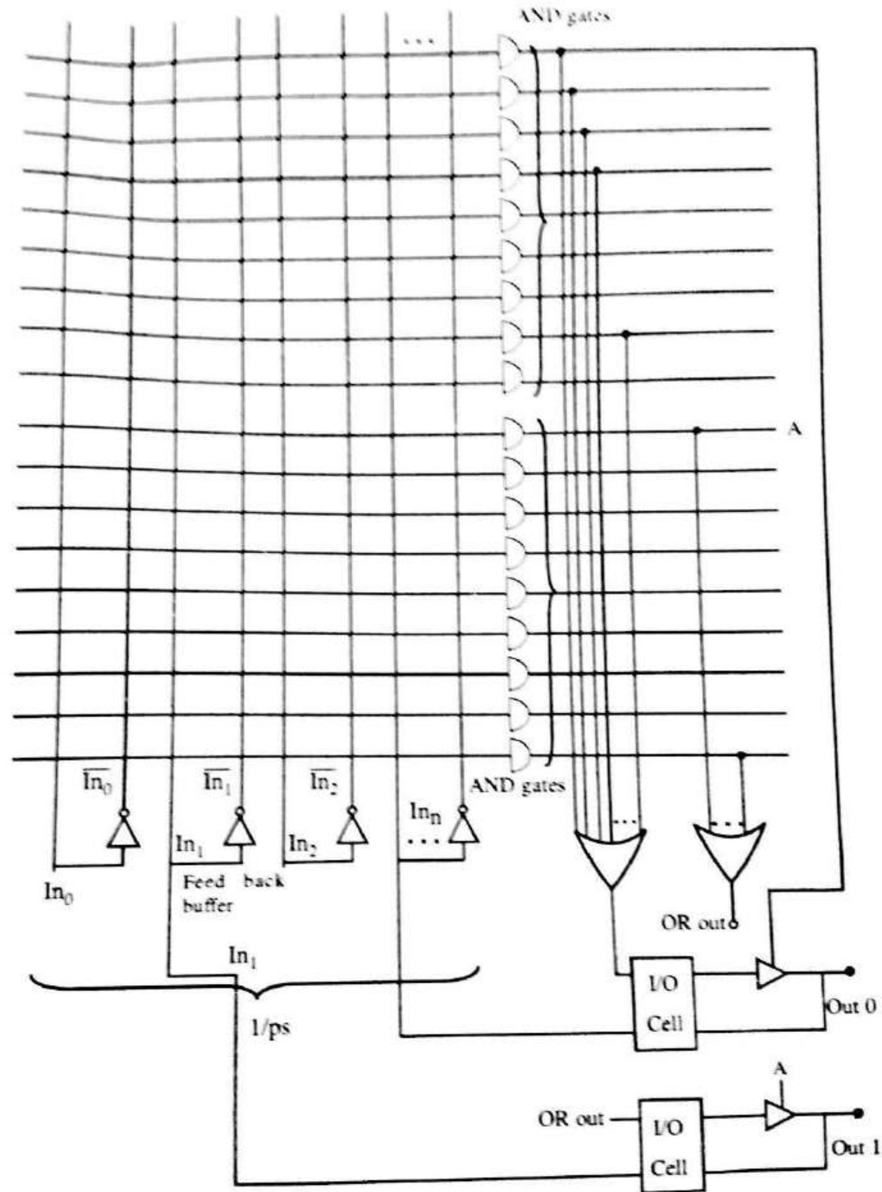


Figure: Example of PAL circuit

Reprogrammable Gate array:

- A field programmable gate array (FPGA) is a VLSI circuit that can be programmed at the user's location.
- A typical FPGA consists of an array of millions of logic blocks, surrounded by programmable input and output blocks and connected together via programmable interconnections.
- There is a wide variety of internal configurations within this group of devices.
- The performance of each type of device depends on the circuit contained in its logic blocks and the efficiency of its programmed interconnections.

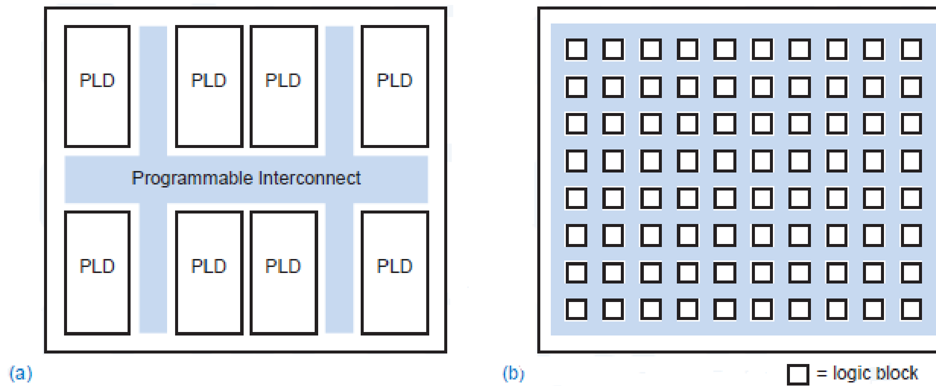


Figure: Programmable-logic-device approaches: (a) CPLD (b) FPGA.

- A typical FPGA logic block consists of lookup tables, multiplexers, gates, and flip-flops.
- A lookup table is a truth table stored in an SRAM and provides the combinational circuit functions for the logic block.
- The combinational logic section, along with a number of programmable multiplexers, is used to configure the input equations for the flip-flop and the output of the logic block.
- The *advantage* of using RAM instead of ROM to store the truth table is that the table can be programmed by writing into memory.
- The *disadvantage* is that the memory is volatile and presents the need for the lookup table's content to be reloaded in the event that power is disrupted.
- The program can be downloaded either from a host computer or from an onboard PROM.
- The program remains in SRAM until the FPGA is reprogrammed or the power is turned off. The device must be reprogrammed every time power is turned on.

****

<u>Programming technology used in FPGA :</u>

Discuss the different types of programming technology used in FPGA design. (NOV 2016)

- There are three types of programming technology.
  - ✓ Fusible link programming (Anti fuse)
  - ✓ SRAM Programming
  - ✓ EPROM and EEPROM programming

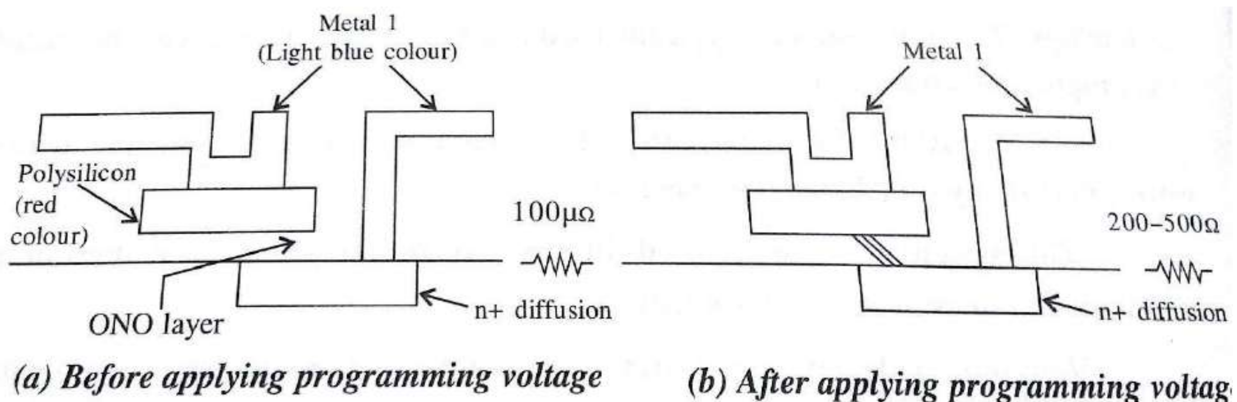
5.6.1: Fusible link programming:

- In this type, Platinum, Titanium tungsten is used to form link.
- It is blown when certain current is exceeded in the fuse. Higher voltage is applied to device to blow the fuses.

Draw and explain the operation of metal-metal antifuse and EPROM transistor. (June 2012)

ANTIFUSE:

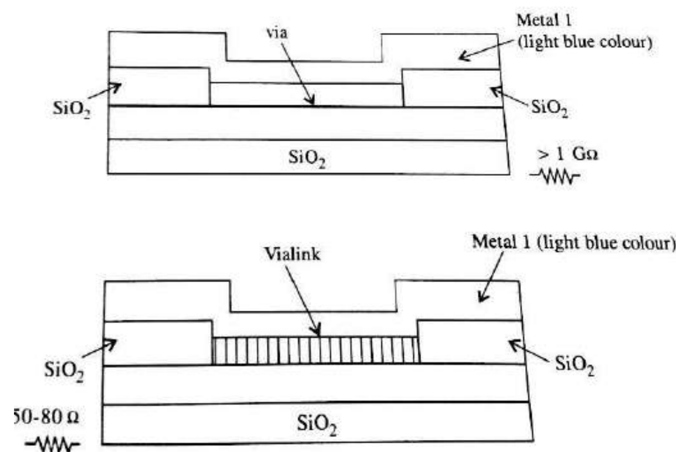
- In FPGA, the device is programmed by changing the characteristic of switching element (or) we can write the program for routing.
- Programming routing can be explained by using the product of ACTEL, Quick Logic Companies etc.
- In ACTEL, interconnect is done by PLICE (or) Antifuse.
- PLICE means Programmable Low Impedance Circuit Element.
- Antifuse is high resistance (>100MΩ) is changed into low resistance (200-500Ω) by applying programming voltage.
- It consists of ONO (Oxide-Nitride-Oxide) layer which is sandwiched between polysilicon layer and n+ diffusion.
- Antifuses separate interconnect wires on the FPGA chip and the programmer blows an antifuse to make a permanent connection.
- Once an antifuse is programmed, the process can't be reversed. This is an **OTP Technology.**



(a) Before applying programming voltage *(b) After applying programming voltag*

In-system programming (ISP):

- Possibility to program the chip after it has been assembled on the PCB.
- In Quick logic company, programmable interconnect is provided with Vialink (metal-metal anti-fuse).



45

Figure: Metal-metal anti-fuse

Advantages of metal-metal antifuse:

- Advantages of metal-metal antifuse over poly diffusion antifuse are:
    1. The connections are direct to metal wiring layers.
    2. It is easier to use larger programming currents to reduce the antifuse resistance.

UV-Erasable programming:

Find the reason for referring EPROM technology as floating gate avalanche MOS. (Dec. 2013)

EPROM programming:

- In this type floating gate transistor is used.
- We can reprogram by using UV-light.
- High electric field causes electrons flowing towards drain to move across the insulating gate oxide, where they trapped on the bottom, floating gate.
- These energetic electrons are HOT and this effect is known as Hot-electron injection (or) avalanche injection.
- EPROM technology is sometimes called floating –gate avalanche MOS (FAMOS).

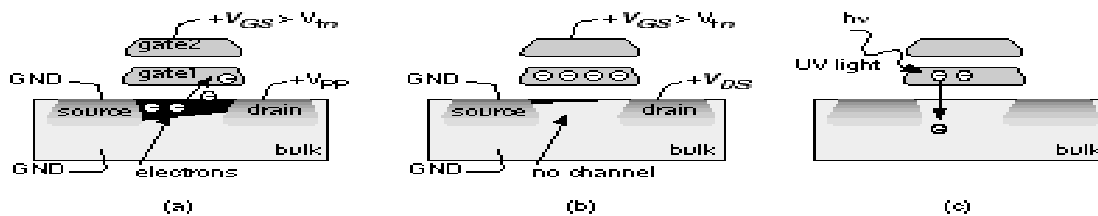


Figure: EPROM transistor

- (a) With a high (>12V) programming voltage, V_{PP} applied to the drain. Electrons gain enough energy to jump onto the floating gate (gate1).
- (b) Electrons stuck on gate1 raise the threshold voltage so that the transistor is always off for normal operating voltages.
- (c) Ultraviolet light provides enough energy for electrons stuck on gate1 to jump back to the bulk, allowing the transistor to operate normally.

EEPROM programming:

- Electrically Erasable programming is most popular CMOS technology.
- A very thin oxide between floating gate and the drain allow the electrons to tunnel to or from the floating gate (gate is charged or discharged).
- Thus enabling writing and erasing operation.

Advantages:

- The advantages of EEPROM technology are:
    - ✓ faster than using a UV lamp
    - ✓ chips do not have to be removed from the system
    - ✓ if the system contains circuits to generate both program and erase voltages, it may use ISP

46

SRAM Programming

- SRAM programming is shown in figure.
- SRAM configuration cell is constructed from two cross-coupled inverters and uses a standard CMOS process.
- The configuration cell drives the gates of other transistors on the chip (using pass transistors or transmission gates) to make a connection or off to break a connection.
- The cell is programmed using the WRITE and DATA lines.

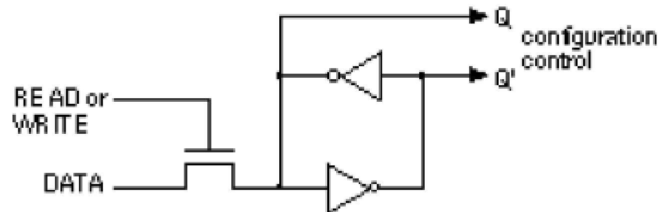


Figure: SRAM programming

Advantages:

- Designers can reuse chips during prototyping.
- Designers can update or change a system on the fly in reconfigurable hardware.

Disadvantage:

- Need to keep power supply for retaining the connection information.

****

- ❖ **Explain the reprogrammable device architecture with neat diagrams.**
- ❖ **With neat diagram explain the functional blocks in PDA (Programmable Device Architecture). (AU:June 2015, June 2016)**
- ❖ **With neat sketch explain the CLB, IOB and Programmable interconnects of an FPGA device. (May 2016)**
- ❖ **Explain about building block architecture of FPGA. (April 2017, 2018, NOV 2018)**
- ❖ **Elucidate in detail the basic FPGA architecture. (April 2019-13M)**
- ❖ **Describe in detail FPGA architecture and explain the main building blocks of FPGA. (Nov 2019)[Nov/Dec 2022]**
- ❖ **Illustrate the basic building block architectures of FPGA. [May 2021]**

<u>Re-Programmable Devices Architecture (FPGA)</u>

- FPGA provide the next generation in the programmable logic devices.
- It refers to the ability of the gate arrays to be programmed for a specific function by the user.
- The word Array is used to indicate a series of columns and rows of gates that can be programmed by the end user.
- As compared to standard gate arrays, the field programmable gate arrays are larger devices.

47

- The basic cell structure for FPGA is complicated than the basic cell structure of standard gate array.
- The programmable logic blocks of FPGA are called Configurable Logic Block (CLB).
- The FPGA architecture consists of **three types of configurable elements**-
    - (i) IOBs –Input/output blocks
    - (ii) CLBs- Configurable logic blocks
    - (iii) Resources for interconnection
- The IOBs provide a programmable interface between the internal, array of logic blocks (CLBs) and the device's external package pins.
- CLBs perform user-specified logic functions.
- The interconnect resources carry signals among the blocks.
- A configurable program stored in internal static memory cells.
- Configurable program determines the logic functions and the interconnections.
- The configurable data is loaded into the device during power-up reprogramming function.
- FPGA devices are customized by loading configuration data into internal memory cells.

The structure of FPGA:

The basic elements of the FPGA structure:

1. Logic blocks
    - Based on memories *(Flip-flop & LUT – Lookup Table)* Xilinx
    - Based on multiplexers *(Multiplexers)*-Actel
    - Based on PAL/PLA - Altera
    - Transistor Pairs
2. Interconnection Resources
    - Symmetrical FPGA-s
    - Row-based FPGA-s
    - *Sea-of-gates* type of FPGA-s
    - Hierarchical FPGA-s (*CPLD)*
3. Input-output cells (*I/O Cell)*
    - Possibilities for programming :
        - a. Input
        - b. Output
        - c. Bidirectional

48

RE-PROGRAMMABLE DEVICE ARCHITECTURE:

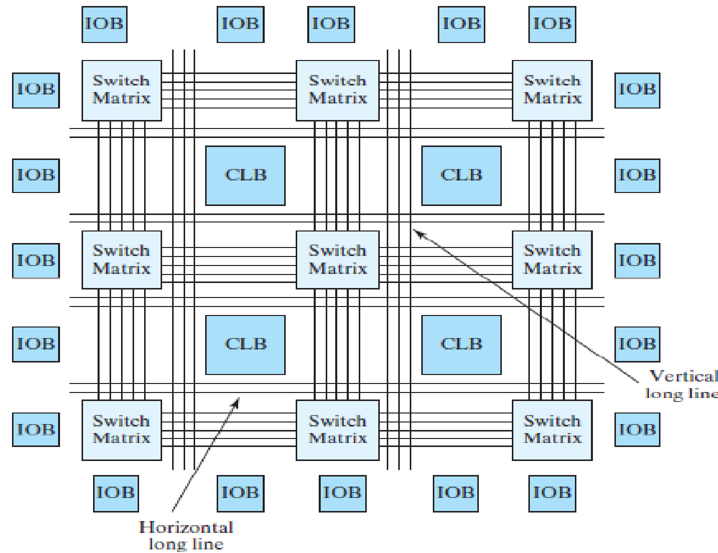
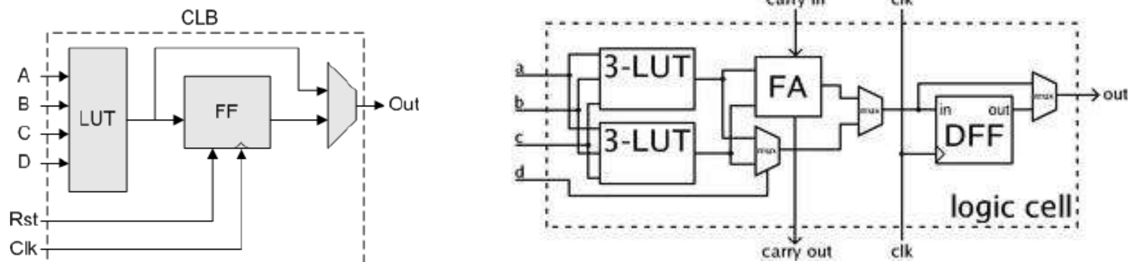


Figure: FPGA building blocks structure

- The figure shows the general structure of FPGA chip.
- It consists of a large number of programmable logic blocks surrounded by programmable I/O block.

Configurable Logic Block:



- **Figure: Various configurable Logic Block**

- The programmable logic blocks of FPGA are smaller and less capable than a PLD, but an FPGA chip contains a lot more logic blocks to make it more capable.
- As shown in figure the logic blocks are distributed across the entire chip.
- These logic blocks can be interconnected with programmable inter connections.
- The programmable logic blocks of FPGAs are called Configurable Logic Blocks (CLBs).
- CLBs contain LUT, FF, logic gates and Multiplexer to perform logic functions.
- The CLB contains RAM memory cells and can be programmed to realize any function of five variables or any two functions of four variables.
- The functions are stored in the truth table form, so the number of gates required to realize the functions is not important.

Interconnection resources:

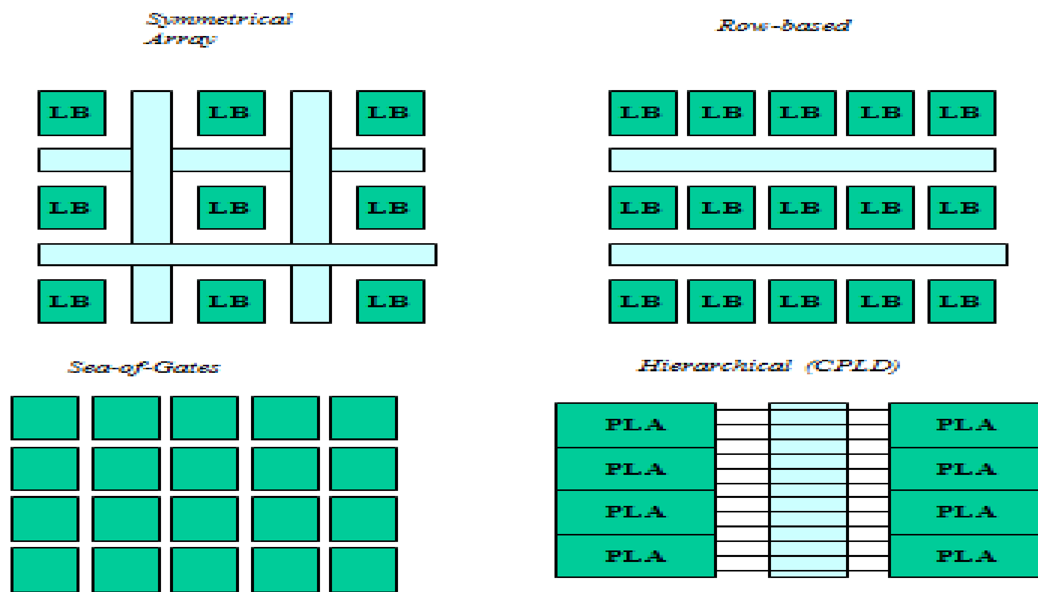


Figure: Types of interconnection resources

(a) Symmetrical Arrays

- It consists of logic elements (CLBs) arranged in rows and columns of a matrix and interconnect laid out between them.
- This symmrtrical martrix is surrounded by I/O blocks which connect it to outside world.

(b) Row based architecture:

- It consists of alternating rows of logic modules and programmable interconnect tracks.
- Input output blocks is located in the periphery of the rows.
- One row may be connected to adjacent rows via vertical interconnect.

(c) Hierarchical CPLD:

- This architecture is designed in hierarchical manner with top level containing only logic blocks and interconnects.
    1. Connections within macrocells
    2. Local connection resource within the logical block.
    3. Global connection resource *(Switch Matrix)*

(d) Sea of gates structure:

- It consists of logic elements (CLBs) arranged in rows and columns of a matrix inthe channel less gate arrays module.

I/O cells(Blocks):

- User-configurable input/output blocks (IOBs) provide the interface between external package pins and the internal logic.
- Each IOB controls one package pin and can be configured for input, output, or bidirectional signals.
- Figure shows a three-state bidirectional output buffer.

- When the output enable, OE is ‘1’ the output section is enabled and drives the I/O pad.
- When OE is ‘0’ the output buffer is placed in a high-impedance state.

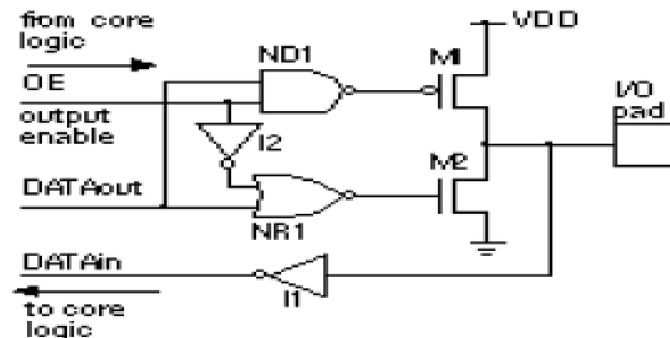


Figure: A three-state bidirectional output buffer

- We can limit the number of I/O drivers that can be attached to any one V_{DD} and GND pad.
- It allows employ the same pad for input and output bidirectional I/O.
- When we want to use the pad as an input, set OE low and take the data from DATAin.
- We can build output-only or input-only pads.

****

FPGA(*PROGRAMMABLE ASIC *)interconnect routing procedures (Architectures):

- ❖ **Give short notes on FPGA interconnect routing procedures. (May 2016, May 2021)**
- ❖ **Describe FPGA interconnect routing resources with neat diagram. (April 2019-13M)**
- ❖ **Give a note on standard cell design and FPGA interconnecting resources. (Nov 2019) [Apr/May 2022]**

- Routing architecture comprises of programmable switches and many wires.
- Routing provides connection between logic blocks, I/O blocks, and between one logic block and another logic block.
- The type of routing architecture decides area consumed by routing as well as density of logic blocks.
- Routing techniques decide the amount of area used by wire segments and programmable switches as compared to area consumed by logic blocks.

Types of FPGA interconnect routing procedures:

- ✓ Hierarchical Routing Architecture
- ✓ Island-Style Routing Architecture
- ✓ Xilinx Routing Architecture
- ✓ Altera Routing Architecture
- ✓ Actel Routing Architecture

(a) Hierarchical Routing Architecture:

- Hierarchical routing architectures separates FPGA logic blocks into distinct groups.

- Connections between the logic blocks within a group can be made using wire segments at the lowest level of the routing hierarchy.
- Connections between the logic blocks in distant groups require the traversal of one or more levels of routing segments.
- As shown in Figure, only one level of routing directly connects to the logic blocks.
- Programmable connections are represented with the crosses and circles.

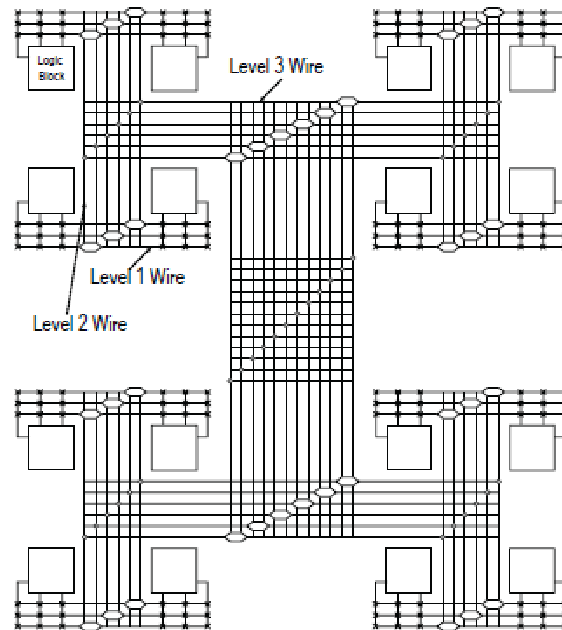


Figure: Example of Hierarchical FPGA

(b) Xilinx Routing Architecture:

- In Xilinx routing, connections are made from logic block into the channel through a connection block.
- As SRAM technology is used to implement Lookup Tables, connection sites are large.
- A logic block is surrounded by connection blocks on all four sides.

52

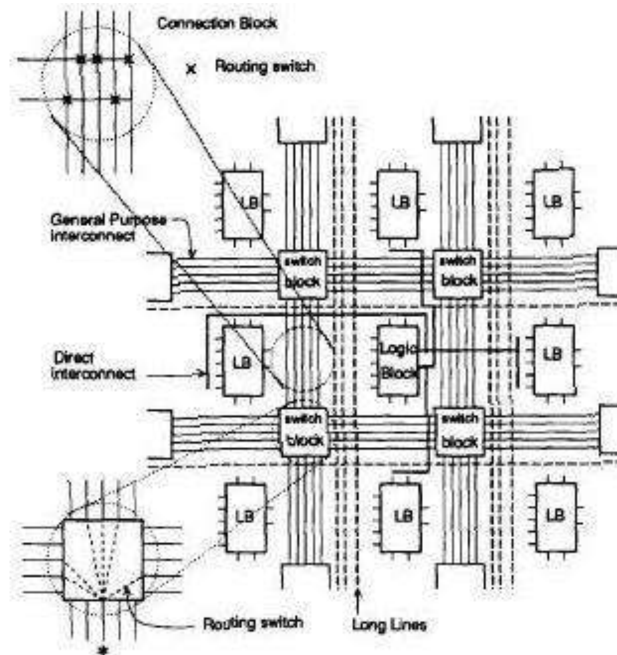


Figure: Xilinix Routing Architecture

- They connect logic block pins to wire segments.
- Pass transistors are used to implement connection for output pins, while use of multiplexers for input pins saves the number of SRAM cells required per pin.
- The logic block pins connecting to connection blocks can then be connected to any number of wire segments through switching blocks.
- Figure shows the Xilinx routing architecture.
- There are four types of wire segments available:
    - ✓ General purpose segments that pass through switches in the switch block.
    - ✓ Direct interconnect connects logic block pins to four surrounding connecting blocks
    - ✓ Long line: high fan out uniform delay connections
    - ✓ Clock lines: clock signal provider which runs all over the chip.

(c) Altera Routing Architecture :

- Altera routing architecture has two level hierarchies.
- At the first level of the hierarchy, 16 or 32 of the logic blocks are grouped into a Logic Array Block (LAB).
- The channel here is set of wires that run vertically along the length of the FPGA.
- Figure shows Alter Max 5000 routing architecture.
- Tracks are used for four types of connections:
    - ✓ Connections from output of all logic blocks in LAB.
    - ✓ Connection from logic expanders.
    - ✓ Connections from output of logic blocks in other LABs
    - ✓ Connections to and from Input output pads

53

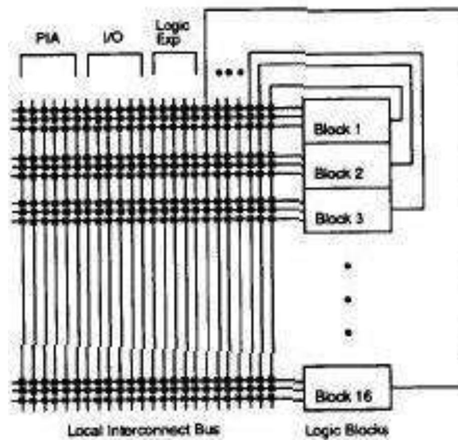


Figure: Altera Max 5000 Routing Architecture

- All four types of tracks connect to every logic block in the array block.
- Any track can connect to into any input which makes this routing simple.
- Advantage: It allows to be packed tightly and efficiently.
- Disadvantage: Large number of switches required, which adds to capacitive load.

(d) Island-Style Routing Architecture:

- As shown in Figure, island-style FPGAs logic blocks are arranged in a two dimensional mesh with the routing resources evenly distributed throughout the mesh.
- An island-style global routing architecture typically has the routing channels on all four sides of the logic blocks.
- The number of wires contained in the channel, W, is pre-set during fabrication, and is one of the key choices made by the architect.
- It employs wire segments of different lengths in each channel to provide the most appropriate length for each given connection.

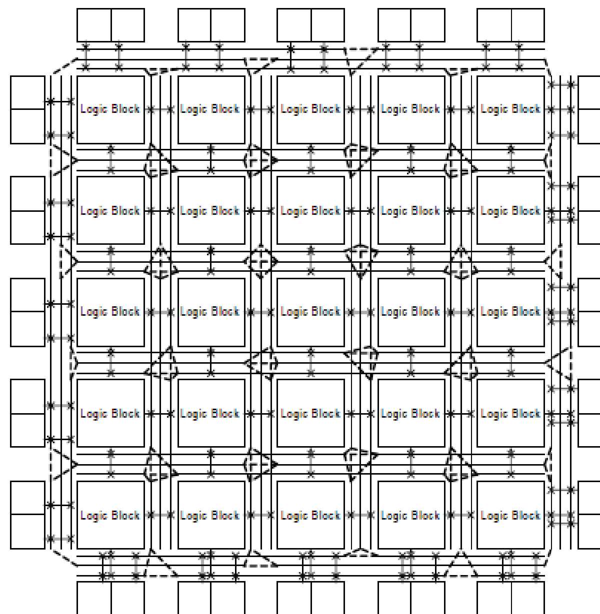


Figure: Island-Style Routing Architecture

(e) Actel Routing Architecture:

- Actel's design has more wire segments in horizontal direction than in vertical direction.
- The input pins connect to all tracks of the channel that is on the same side as the pin.
- The output pins extend across two channels above the logic block and two channels below it.
- Output pin can be connected to all 4 channels that it crosses.
- The switch blocks are distributed throughout the horizontal channels.
- All vertical tracks can make a connection with every incidental horizontal track.
- This allows for the flexibility that a horizontal track can switch into a vertical track, thus allowing for horizontal and vertical routing of same wire.
- The drawback is more switches are required which add up to more capacitive load.

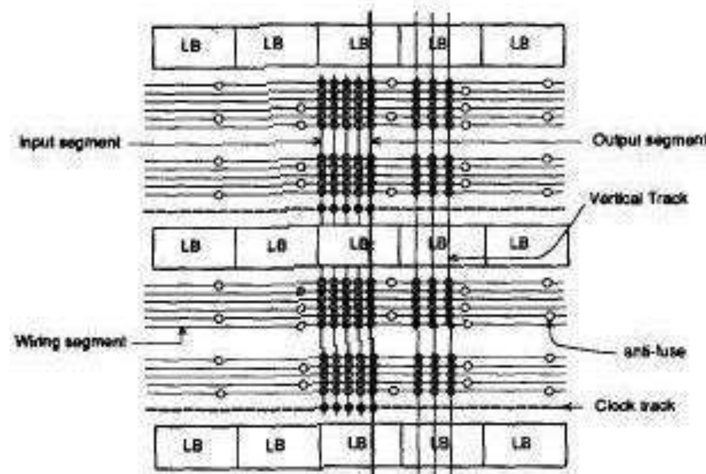


Figure: Actel Routing Architecture

***INTERCONNECT*:**

Explain in detail about the interconnect.

Interconnect in an integrated circuit are physical connections between two transistors and/ or the externalsurroundings.

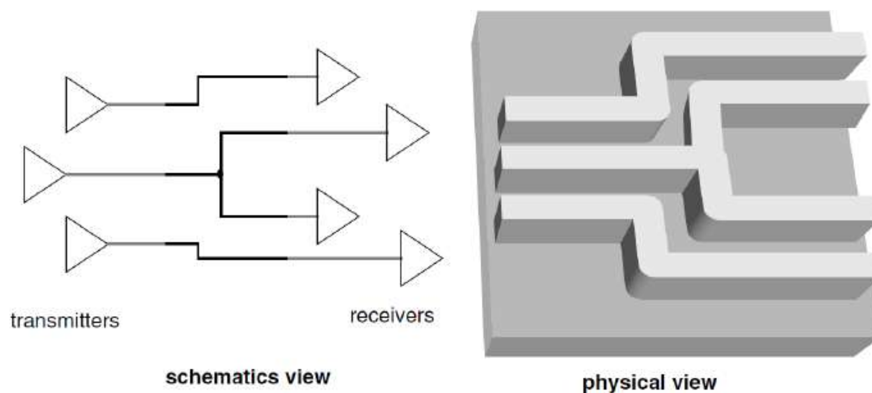
- An electronic circuit designer has multiple choices in realizing the interconnections between the various devices that make up the circuit.
- Here the start of the art processes offers multiple layers of aluminium or copper, and at least one layer of polysilicon. Even the heavily doped n^+ and p^+ diffusion layers are typically used for the realization of source and drain regions can be employed for wiring purposes. These wires appear in the schematic diagrams of electronic circuit as simple lines with no apparent impact on the circuit performance.

These wiring of integrated circuits forms a complex geometry that introduces the following parasitics:

1. Capacitive Parasitics
2. Resistive Parasitics and
3. Inductive Parasitics

The capacitive, resistive and inductive parasitics have multiple effects of the circuit's behaviour i.e.

- They all cause an increase in propagation delay, or equivalent, a drop in performance.
- They all have an impact on the energy dissipation and the power distribution.
- They all cause the introduction of extra noise sources, which affect the reliability of the circuit.



(FOR UNDERSTANDING- NO NEED TO DRAW)

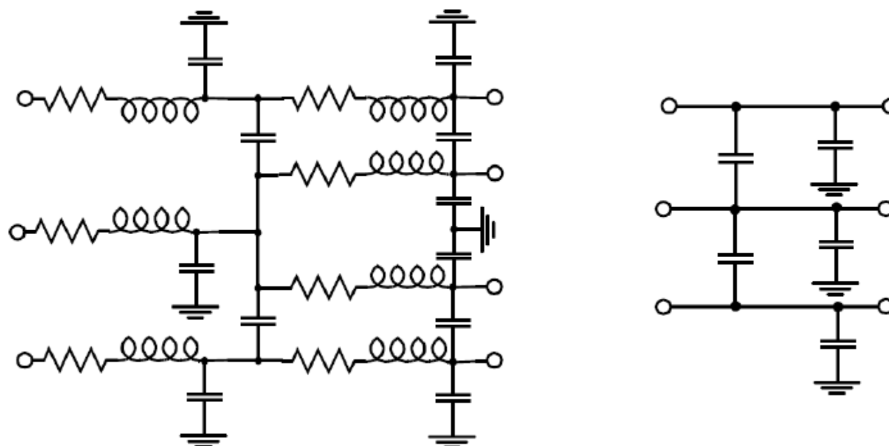
SCHEMATIC AND PHYSICAL VIEWS OF WIRING OF BUS-NETWORK

It is important that the designer has a clear insight in the parasitic wiring effects, their relative importance, and their models. This is best illustrated with the simple example as shown above. Each wire in a bus network connects a transmitter (or transmitters) to a set of receivers and is implemented as a link of wire segments of various lengths and geometries. Assume that all segments are implemented on a single interconnect layer, isolated from the silicon substrate and from each other by a layer of dielectric material.Be aware that the reality may be far more complex.

56

Analyzing the behavior of this schematic, which only models a small part of the circuit, is slow and cumbersome. Fortunately, substantial simplifications can often be made, some of which are enumerated below,

- Inductive effects can be ignored if the resistance of the wire is substantial- this is for instance the case for long Aluminum wires with a small cross-section- or if the rise and fall times of the applied signals are slow.
- When the wires are short, the cross-section of the wire is large, or the interconnect material used has a low resistivity, a capacitance- only model can be used (figure FOR WIRE PARASITICS WITH CAPACITANCE ONLY shown below)
- The separation between neighbouring wires is large, or when the wires only run together for a short distance, inter-wire capacitance can be ignored, and all the parasitic capacitance can be modeled as capacitance to ground. Obviously, the latter problems are the easiest to model, analyze, and optimize.



WIRE PARASITICS (WITH THE EXCEPTION OF INTER-WIRE RESISTANCE AND MUTUAL INDUCTANCE) WIRE PARASITICS WITH CAPACITANCE ONLY

WIRE MODELS FOR PARASITICS

The various interconnect parameters whose values can be estimated, simple models to evaluate their impact, and a set of rules- of- thumb to decide i.e. when and where a particular model or effect should be consideredare:

1. Capacitance Parameter
2. Resistance Parameter
3. Inductance Parameter

CAPACITANCE INTERCONNECT PARAMETER:

The capacitance of such a wire is a function of its shape, its environment, its distance to the substrate, and the distance to surrounding wires. An accurate modeling of the wire capacitance(s) in a state-of-the-art integrated circuit is a non-trivial task and is even today the subject of advanced research.

In capacitance parameter there are two types of capacitance occurring i.e.

1. Parallel plate Capacitance and
2. Fringe Capacitance

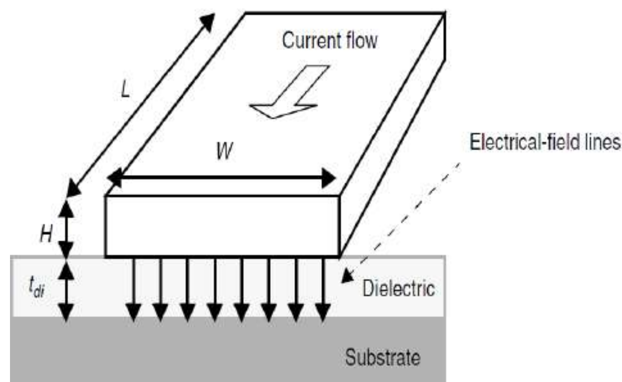
57

Parallel plate Capacitance:

Consider first a simple rectangular wire placed above the semiconductor substrate, as shown in figure below.If the width of the wire is substantially larger than the thickness of the insulating material, it may be assumed that the electrical-field lines are orthogonal to the capacitor plates, and that its capacitance can be modeled by the *parallel-plate capacitance model.* Under those circumstances, the total capacitance of the wire can be approximated as,

$$c_{int} = \frac{\varepsilon_{di}}{t_{di}} WL$$

Where W and L are respectively the width and length of the wire, and t_{di} and ε_{di} represent the thickness of the dielectric layer and its permittivity. SiO2 is the dielectric material of choice in integrated circuits, although some materials with lower permittivity, and hence lower capacitance, are coming in use.

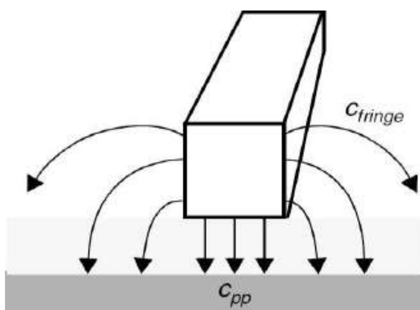


PARALLEL-PLATE CAPACITANCE MODEL OF INTERCONNECT WIRE

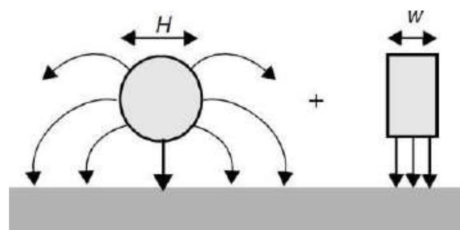
In actuality, this model is too simplistic. To minimize the resistance of the wires while scaling technology, itis desirable to keep the cross-section of the wire ($W \times H$) as large as possible. On the other hand, small values of W lead to denser wiring and less area overhead. As a result, we have over the years witnessed a steady reduction in the *W/H*- ratio, such that it has even dropped below unity in advanced processes.

Fringe/ Fringing Capacitance:

The capacitance between the side-walls of the wires and the substrate, called the *fringing capacitance*, as shown below.



Fringing fields/ the fringing-field capacitance



model of fringing-field capacitance- decomposes the Capacitance into two contributions: a parallel-plate capacitance,and a fringing capacitance, modeled by a cylindrical wire with a diameter equal to the thickness of the wire

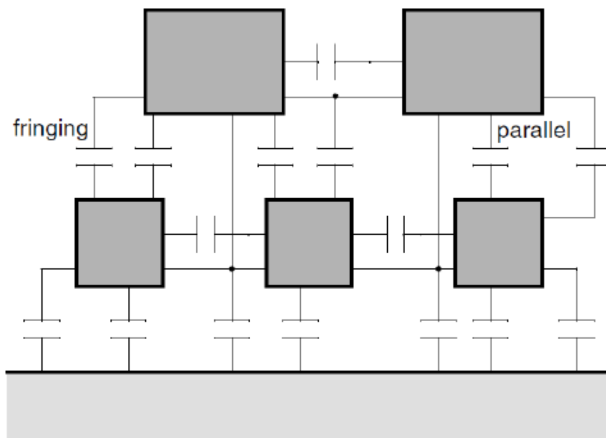
Therefore, the parallel plate capacitance and fringing capacitance constitutes the overall capacitance. Which is given as,

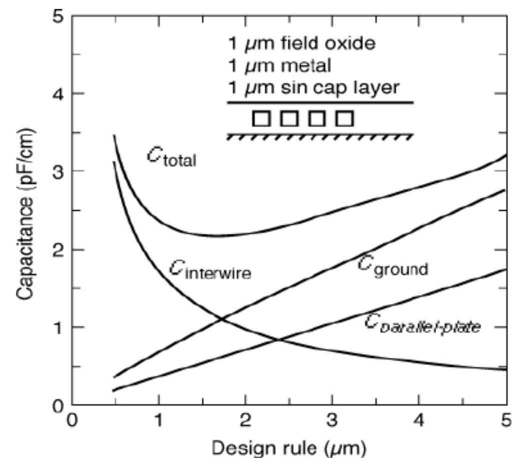
$$c_{wire} = c_{pp} + c_{fringe} = \frac{w\varepsilon_{di}}{t_{di}} + \frac{2\pi\varepsilon_{di}}{\log(t_{di}/H)}$$

With $w = W - H/2$ a good approximation for the width of the parallel-plate capacitor.

CAPACITANCE COUPLING/ CAPACITANCE COUPLING EFFECT:

Assuming that a wire is completely isolated from its surrounding structures and is only capacitively coupled to ground, becomes untenable. This is illustrated in figure, where the capacitance components of a wire embedded in an interconnect hierarchy are identified. Each wire is not only coupled to the grounded substrate, but also to the neighbouring wires on the same layer and on adjacent layers. The main differenceis that not all its capacitive components do terminate at the grounded substrate, but that a large number of them connect to other wires, which have dynamically varying voltage levels, these floating capacitors causescrosstalk and a negative effect to the circuit also.





CAPACITIVE COUPLING BETWEEN WIRES IN INTERCONNECT HIERARCHY ***INTERCONNECT CAPACITANCE AS A FUNCTION OF DESIGN RULES***

Inter- wire capacitances become a dominant factor in multi- layer interconnect structures. This effect is more important for wires in the higher interconnect layers, as these wires are farther away from the substrate. The increasing contribution of the inter- wire capacitance to the total capacitance with decreasing feature sizes is illustrated by graphical figure as shown, which plots the capacitive components of a set of parallel wires routed above a ground plane, it is assumed that dielectric and wire thickness are held constant while scaling all other dimensions. When W becomes smaller than 1.75 H, the inter-wire capacitance starts to dominate.

Wiring Capacitances for 0.25 µm CMOS Technology:

The table rows represent the top plate of the capacitor, the columns the bottom plate. The area capacitancesare expressed in aF[1]/µm², while the fringe capacitances (given in the shaded rows) are in aF/µm.

	Field	Active	Poly	Al1	Al2	Al3	Al4
Poly	88						
	54						
Al1	30	41	57				
	40	47	54				
Al2	13	15	17	36			
	25	27	29	45			
Al3	8.9	9.4	10	15	41		
	18	19	20	27	49		
Al4	6.5	6.8	7	8.9	15	35	
	14	15	15	18	27	45	
Al5	5.2	5.4	5.4	6.6	9.1	14	38
	12	12	12	14	19	27	52

Inter- Wire Capacitance per unit wire length for different interconnect layers of 0.25 µm CMOSTechnology Process:

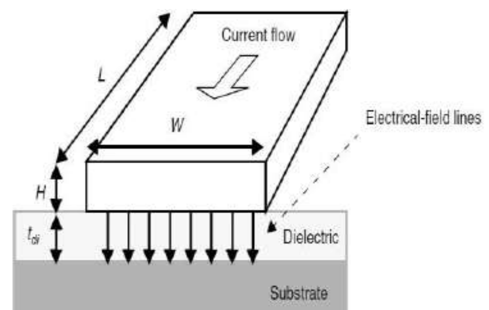
The capacitances are expressed in aF/mm, and are for minimally-spaced wires

Layer	Poly	Al1	Al2	Al3	Al4	Al5
Capacitance	40	95	85	85	85	115

RESISTANCE INTERCONNECT PARAMETER:

The resistance of a wire is proportional to its length *L* and inversely proportional to its cross- section *A*.The resistance of a rectangular conductor as shown in figure below can be expressed as,

$$R = \frac{\rho L}{A}\ ohm = \frac{\rho L}{H\,W}\ ohm$$



Where:

ρ = resistivity

A = HW = area of cross section of the rectangular wireIf L = W, i.e. square of resistive material, then

$$R = \frac{\rho}{H} = R_s \text{ in } {}^{ohm}/_{square} \text{ (Sheet Resistance)}$$

the *sheet resistance* of the material, having units of Ω/ sq. This expresses that the resistance of a square conductor is independent of its absolute size, as is apparent from $R = R_S \frac{L}{W}$.

To obtain the resistance of a wire, simply multiply the sheet resistance by its ratio (L/ W).

Resistivity of Commonly used Conductors/ Interconnect Resistance:

Material	ρ (Ω-m)
Silver (Ag)	1.6×10^{-8}
Copper (Cu)	1.7×10^{-8}
Gold (Au)	2.2×10^{-8}
Aluminum (Al)	2.7×10^{-8}
Tungsten (W)	5.5×10^{-8}

Aluminum is the interconnect material most often used in integrated circuits because of its low cost and its compatibility with the standard integrated- circuit fabrication process. Unfortunately, it has a large resistivity compared to materials such as Copper. With ever-increasing performance targets, this is rapidly becoming a liability and top- of- the- line processes are now increasingly using Copper as the conductor of choice.

Typical values of the Sheet Resistance of various Interconnect Materials using 0.25 µm CMOS Technology:

Material	Sheet Resistance (Ω/❑)
n- or p-well diffusion	1000 – 1500
n^+, p^+ diffusion	50 – 150
n^+, p^+ diffusion with silicide	3 – 5
n^+, p^+ polysilicon	150 – 200
n^+, p^+ polysilicon with silicide	4 – 5
Aluminum	0.05 – 0.1

From the table, we conclude that Aluminum is the preferred material for the wiring of long interconnections. Polysilicon should only be used for local interconnect. Although the sheet resistance of the diffusion layer (*n*+, *p*+) is comparable to that of polysilicon, the use of diffusion wires should be avoided due to its large capacitance and the associated *RC* delay.

INDUCTANCE INTERCONNECT PARAMETER:

The inductance of a section of a circuit states that a changing current passing through an inductor generates avoltage drop ΔV.

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

On-chip inductance include ringing and overshoot effects, reflections of signals due to impedance mismatch, inductive coupling between lines, and switching noise due to *Ldi*/*dt* voltage drops.

It is possible to compute the inductance a wire directly from its geometry and its environment. A simpler approach relies on the fact that the capacitance c and the inductance l (per unit length) of a wire are relatedby the following expression,

$$cl = \varepsilon\mu$$

With ε and μ respectively the permittivity and permeability of the surrounding dielectric.

Other interesting relations, obtained from Maxwell's laws, can be pointed out. The constant product of permeability and permittivity also defines the speed□ at which an electromagnetic wave can propagate through the medium,

$$\nu = \frac{1}{\sqrt{lc}} = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{c_0}{\sqrt{\varepsilon_r\mu_r}}$$

c_0 equals the speed of light (30 cm/ nsec) in a vacuum.

Considering a lumped RLC model we get $Z_{RL} = R + J\omega L, where\ \omega = 2\pi f.$

$$If\ R \gg J\omega L, then\ inductance\ effect\ is\ not\ important.$$

$$If\ R \ll J\omega L, then\ inductance\ effect\ will\ be\ observed\ in\ signal\ reduction\ in\ resistance$$

Dielectric constants and wave-propagation speeds for various materials used in electronic circuits;(The relative permeability μ_r of most dielectrics is approximately equal to 1)

Dielectric	ε_r	Propagation speed (cm/nsec)
Vacuum	1	30
SiO_2	3.9	15
PC board (epoxy glass)	5.0	13
Alumina (ceramic package)	9.5	10

INTERCONNECT MODELING:

Describe about interconnect modelling.

As we know the parasitic elements have an impact on the electrical behaviour of the circuit and influence its delay, power dissipation, and reliability. To study these effects requires the introduction of electrical models that estimate and approximate the real behaviour of the wire as a function of its parameters. These models vary from very simple to very complex depending upon the effects that are being studied and the required accuracy.

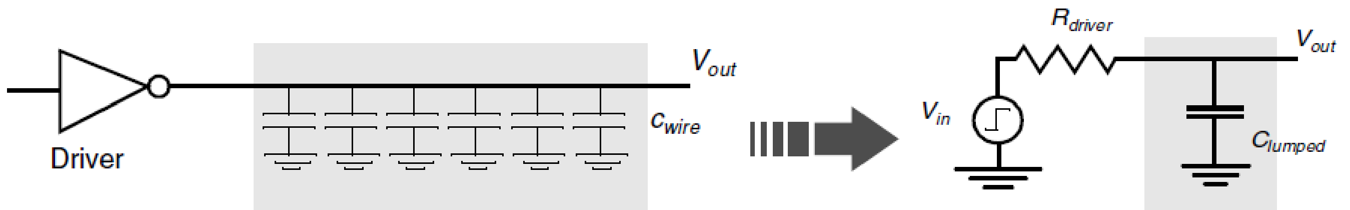
The types of interconnect modelling are:

1. Lumped Model
2. Lumped RC Model- The Elmore Delay
3. Distributed RC line Model/ Distributed *rc* line Model
4. Transmission Line Model

Lumped Model:

The circuit parasitics of a wire are distributed along its length and are not lumped into a single position.Yet, when only a single parasitic component is dominant, when the interaction between the components is small, or when looking at only one aspect of the circuit behaviour, it is often useful to lump the different fractions into a single circuit element. The advantage of this approach is that the effects of the parasitic then can be described by an ordinary differential equation.

As long as the resistive component of the wire is small and the switching frequencies are in the low to medium range, it is meaningful to consider only the capacitive component of the wire, and to lump the distributed capacitance into a single capacitor as shown in figure. It is observed that in this model the wire still represents an equipotential region, and that the wire itself does not introduce any delay. The only impacton performance is introduced by the loading effect of the capacitor on the driving gate. This capacitive lumped model is simple, yet effective, and is the model of choice for the analysis of most interconnect wires in digital integrated circuits.



DISTRIBUTED VERSUS LUMPED CAPACITANCE MODEL OF WIRE. C_{LUMPED} = L×C_{WIRE}, WITH L THE LENGTH OF THE WIRE AND C_{WIRE} THE CAPACITANCE PER UNIT LENGTH. THE DRIVER IS MODELED AS A VOLTAGE SOURCE AND A SOURCE RESISTANCE R_{DRIVER}

The operation of this simple RC network is described by the following ordinary differential equation,

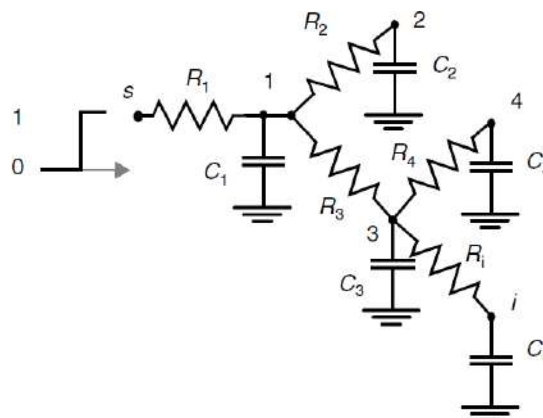
$$C_{lumped}\frac{dV_{out}}{dt} + \frac{V_{out} - V_{in}}{R_{driver}} = 0$$

Lumped RC Model/ The Elmore Delay:

On-chip metal wires of over a few mm length have a significant resistance. The equipotential assumption, presented in the lumped-capacitor model, is no longer adequate, and a resistive-capacitive model has to be adopted.

A first approach lumps the total wire resistance of each wire segment into one single *R* and similarly combines the global capacitance into a single capacitor *C*. This simple model, called the ***lumped RC model***, is pessimistic and inaccurate for long interconnect wires, which are more adequately represented by a *distributed rc-model*. Yet, before analyzing the distributed model, it is worthwhile to spend some time on theanalysis and the modeling of lumped *RC* networks for the following reasons:

1. The distributed *rc*-model is complex and no closed form solutions exist. The behaviour of the distributed *rc*- line can be adequately modeled by a simple *RC* network.
2. A common practice in the study of the transient behavior of complex transistor-wire networks is to reduce the circuit to an *RC* network. Having a means to analyze such a network effectively and to predict its first-order response would add a great asset to the designers tool box.



TREE- STRUCTURED RC NETWORK

An interesting result of this particular circuit topology is that there exists a unique resistive path between the source node *s* and any node *i* of the network. The total resistance along this path is called the *path resistance* R_{ii}. For example, the path resistance between the source node *s* and node 4 equals,

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

The definition of the path resistance can be extended to address the *shared path resistance* R_{ik}, which represents the resistance shared among the paths from the root node *s* to nodes *k* and *i*:

$$R_{ik} = \sum R_j \Rightarrow (R_j \in [path(s \rightarrow i) \cap path(s \rightarrow k)])$$

Here,

$\boldsymbol{R_{i4} = R_1 + R_3}$ **while** $\boldsymbol{R_{i2} = R_1}$

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

$$\tau_{Di} = \sum_{k=1}^{N} C_k R_{ik}$$

Therefore, the Elmore delay is equivalent to the first-order time constant of the network (or the first moment of the impulse response). The designer should be aware that this time- constant represents a simple approximation of the actual delay between source node and node *i*. Yet in most cases this approximation has proven to be quite reasonable and acceptable. It offers the designer a powerful mechanism for providing a quick estimate of the delay of a complex network.

The RC delay of a tree structured network is given as,

$$\tau_{Di} = R_1C_1 + R_1C_2 + (R_1 + R_3)C_3 + (R_1 + R_3)C_4 + (R_1 + R_3 + R_i)C_i$$

i.e. using

$$\tau_{Di} = \sum_{k=1}^{N} C_k R_{ik}$$

We can compute the Elmore delay for node *i*

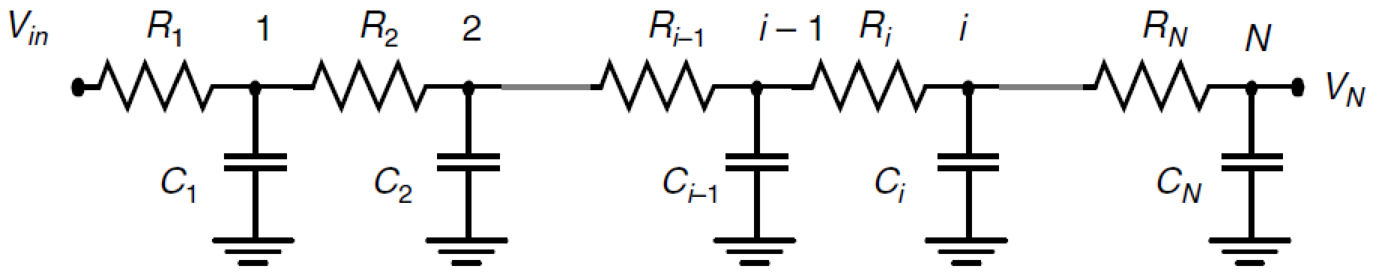
RC Chain/ The Elmore RC Chain Delay:

As a special case of the *RC* tree network, let us consider the simple, non- branched *RC* chain (or ladder) shown in figure. This network is worth analyzing because it is a structure that is often encountered in digital circuits, and also because it represents an approximative model of a resistive-capacitive wire. The Elmore delay of this chain network can be derived with the aid of

$$\tau_{Di} = \sum_{k=1}^{N} C_k R_{ik}$$

As

$$\tau_{DN} = \sum_{i=1}^{N} C_i \sum_{j=1}^{i} R_j = \sum_{i=1}^{N} C_i R_{ii}$$



RC CHAIN MODEL

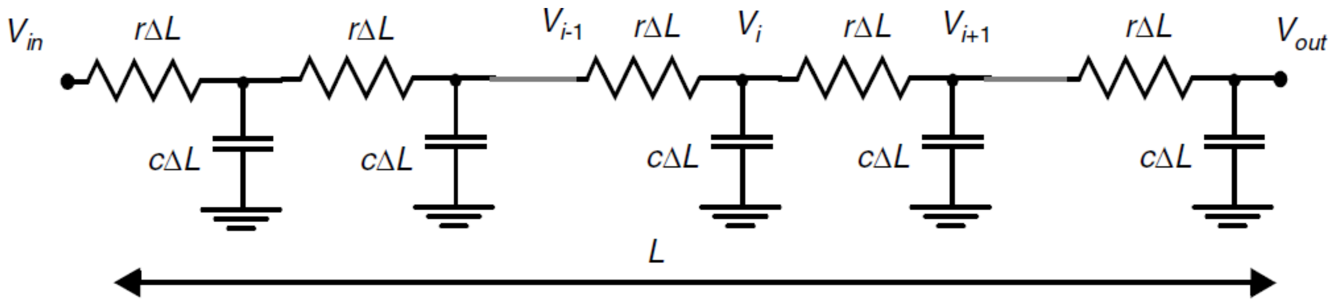
The component of node 1 consists of C_1R_1 with R_1 the total resistance between the node and the source, while the contribution of node 2 equals $C_2(R_1 + R_2)$. The equivalent time constant at node *2* equals $C_1R_1 + C_2(R_1 + R_2)$. τ_i of node *i* can be derived in a similar way.

$$\tau_{Di} = C_1R_1 + C_2(R_1 + R_2) + \ldots + C_i(R_1 + R_2 + \ldots + R_i)$$

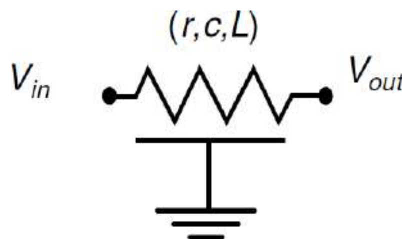
Thus, the Elmore delay formula has proven to be extremely useful. Besides making it possible to analyze wires, the formula can also be used to approximate the propagation delay of complex transistor networks. The evaluation of the propagation delay is then reduced to the analysis of the resulting *RC* network. More precise minimum and maximum bounds on the voltage waveforms in an *RC* tree have further been established.

Distributed RC line Model/ Distributed *rc* line Model:

A distributed *rc* line model is a more appropriate model as shown below which has, *r* and *c* stand for the resistance and capacitance per unit length.



DISTRIBUTED RC LINE MODEL



SCHEMATIC SYMBOL FOR DISTRIBUTED RC LINE

The voltage at node i of this network can be determined by solving the following set of partial differential equations:

$$c\Delta L\frac{\partial V_i}{\partial t} = \frac{(V_{i+1} - V_i) + (V_{i-1} - V_i)}{r\Delta L}$$

The correct behavior of the distributed rc line is then obtained by reducing ΔL asymptotically to 0. For $\Delta L \rightarrow 0$, the above equation becomes the well-known *diffusion*

$$rc\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2}$$

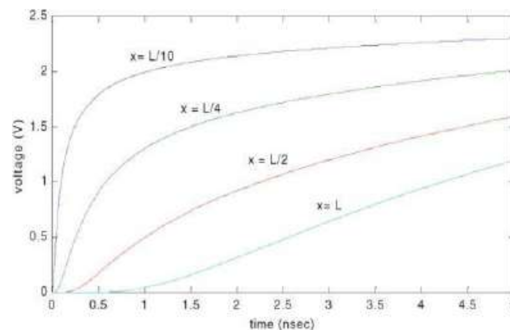
equation:

Where V is the voltage at a particular point in the wire, and x is the distance between this point and the signal source. No closed form solution exists for this equation, but approximative expressions such as the formula written below can be derived:

$$V_{out}(t) = 2erfc\left(\sqrt{\frac{RC}{4t}}\right) \qquad t \ll RC$$

$$= 1.0 - 1.366e^{-2.5359\frac{t}{RC}} + 0.366e^{-9.4641\frac{t}{RC}} \qquad t \gg RC$$

The graph below shows the response of a wire to a step input, plotting the waveforms at different points in the wire as a function of time. It is observable how the step waveform “diffuses” from the start to the end of the wire, and the waveform rapidly degrades, resulting in a considerable delay for long wires. Driving these rc lines and minimizing the delay and signal degradation is one of the trickiest problems in modern digital integrated circuit design.

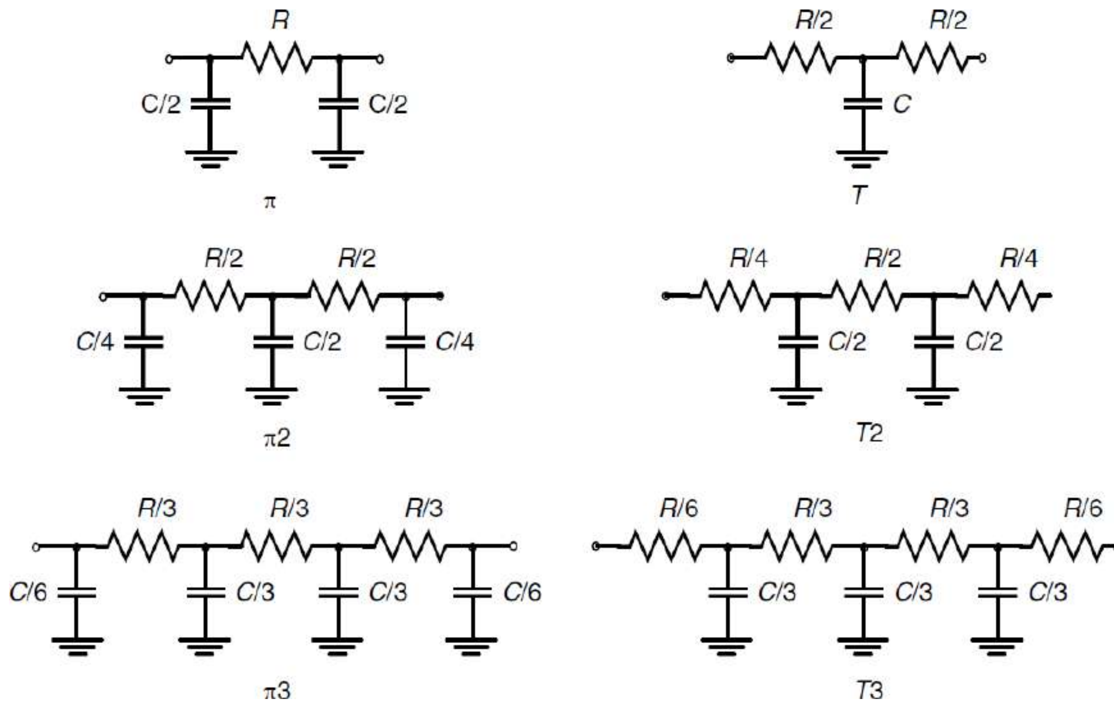


SIMULATED STEP RESPONSE OF RESISTIVE-CAPACITIVE WIRE AS A FUNCTION OF TIME AND PLACE

***rc* delays should only be considered when the rise (fall) time at the line input is smaller than *RC*, the rise (fall) time of the line.**

$$t_{rise} < RC$$

With R and C the total resistance and capacitance of the wire. When this condition is not met, the change in signal is slower than the propagation delay of the wire, and a lumped capacitive model suffices.



SIMULATION π AND T MODELS FOR DISTRIBUTED RC LINE

Step response of lumped and distributed *RC* networks- Points of Interest:

Voltage range	Lumped *RC* network	Distributed *RC* network
$0 \rightarrow 50\%$ (t_p)	0.69 *RC*	0.38 *RC*
$0 \rightarrow 63\%$ (τ)	*RC*	0.5 *RC*
$10\% \rightarrow 90\%$ (t_r)	2.2 *RC*	0.9 *RC*
$0\% \rightarrow 90\%$	2.3 *RC*	1.0 *RC*

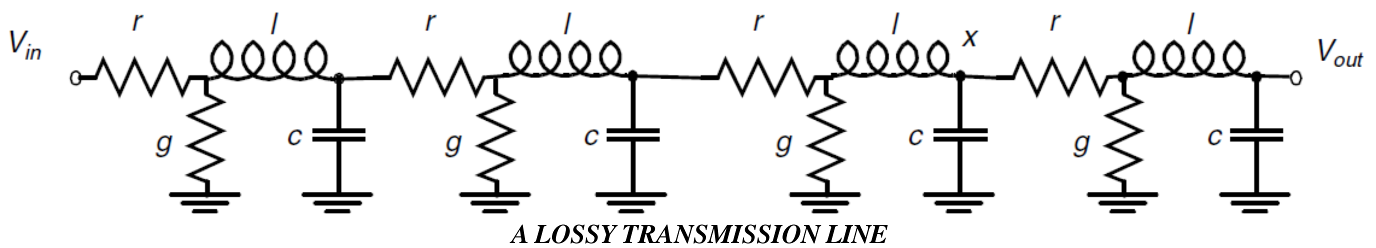
The Transmission Line Model:

Similar to the resistance and capacitance of an interconnect line, the inductance is distributed over the wire. A distributed *rlc* model of a wire, known as the transmission line model, becomes the most accurate approximation of the actual behaviour.

The transmission line has the prime property that a signal propagates over the interconnection medium as a *wave*. This is in contrast to the distributed *rc* model, where the signal *diffuses* from the source to the destination governed by the diffusion equation i.e.

$$rc\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2}$$

In the wave mode, a signal propagates by alternatively transferring energy from the electric to the magnetic fields, or equivalently from the capacitive to the inductive modes.



A LOSSY TRANSMISSION LINE

Consider the point *x* along the transmission line of figure as shown above at time *t*. The following set of equations holds:

$$\frac{\partial v}{\partial x} = -ri - l\frac{\partial i}{\partial t}$$

$$\frac{\partial i}{\partial x} = -gv - c\frac{\partial v}{\partial t}$$

Assuming that the leakage conductance *g* equals 0, which is true for most insulating materials, and eliminating the current *i* yields the *wave propagation equation*,

$$\frac{\partial^2 v}{\partial x^2} = rc\frac{\partial v}{\partial t} + lc\frac{\partial^2 v}{\partial t^2}$$

where *r*, *c,* and *l* are the resistance, capacitance, and inductance per unit length respectively.

COPING WITH INTERCONNECT:

As till now we have concentrated on the growing impact of interconnect parasitics on all design metrics of digital integrated circuits. As mentioned, interconnect introduces three types of parasitic effects i.e. capacitive, resistive, and inductive- all of which influence the signal integrity and degrade the performanceof the circuit. While so far we have concentrated on the modeling aspects of the wire, we now analyze how interconnect affects the circuit operation, and we present a collection of design techniques to cope with theseeffects with considering each parasitics- this is referred to as *coping with interconnect*.

CAPACITIVE PARASITICS:

- **Capacitance reliability and Cross talk:**

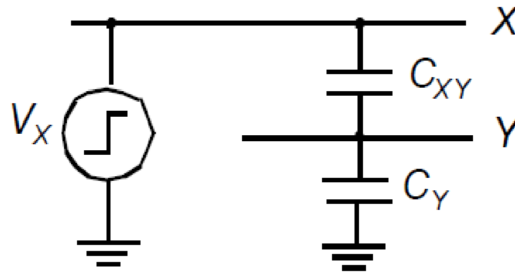
An unwanted coupling from a neighbouring signal wire to a network node introduces an interference that is generally called *cross talk*. The resulting disturbance acts as a noise source and can lead to hard-to-trace intermittent errors, since the injected noise depends upon the transient value of the other signals routed in theneighbourhood. In integrated circuits, this inter signal coupling can be both capacitive and inductive.

Capacitive cross talk is the dominant effect at current switching speeds, although inductive coupling forms a major concern in the design of the input-output circuitry of mixed-signal circuits. The potential impact of capacitive crosstalk is influenced by the impedance of the line under examination. If the line is floating, the

disturbance caused by the coupling persists and may be worsened by subsequent switching on adjacentwires. If the wire is driven, on the other hand, the signal returns to its original level.

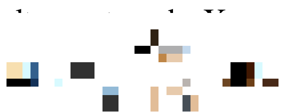
o **Floating Lines:**

Approaching capacitive parasitic with respect to capacitive coupling to a floating line.



CAPACITIVE COUPLING TO A FLOATING LINE

Considering the circuit shown as above, where line X is coupled to wire Y by a parasitic capacitance C_{XY}. Line Y sees a total capacitance to ground equal to C_Y. Assuming that the



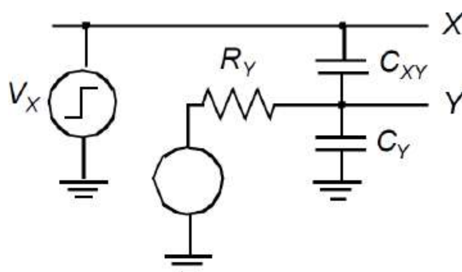
eriences a step change equal to ΔV_X. This step appears on node Y tive voltage divider.

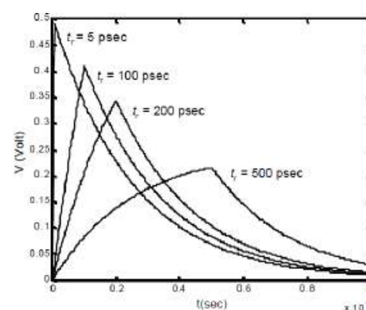
$$\Delta V_Y = \frac{C_{XY}}{C_Y + C_{XY}} \Delta V_X$$

Circuits that are particularly susceptive to capacitive cross talk are networks with low- swing pre chargednodes, located in adjacent to full- swing wires (with $\Delta V_X = V_{DD}$).

Examples of these are dynamic memories, low swing on chip busses and some dynamic families.To address the cross talk issue, level- restoring device or keepers are a must in dynamic logic.

o **Driven Lines:**





CAPACITIVE COUPLING TO A DRIVEN LINE AND ITS VOLTAGE RESPONSE

As seen from the figure, if the line Y is driven with a resistance R_Y, a step on line X results in a transient on line Y. The transient decays with a time constant $\tau_{XY} = R_Y (C_{XY} + C_Y)$. The actual impact on the victim line is a strong function of the rise- fall time of the interfering signal.

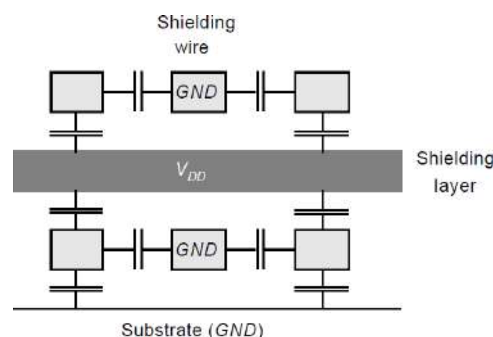
If the rise time is comparable or larger than the time constant, the peak value of disturbance is diminished. This can be observed in the response figure.

Obvious, keeping the driving impedance of a wire and hence τ_{XY} low goes a long way towards reducing the impact of capacitive cross talk. The keeper transistor added to a dynamic gate or pre charged wire is an excellent example of how impedance reduction helps to control noise.

Therefore, the impact of cross talk on the signal integrity of driven nodes is rather limited. The resulting glitches may cause malfunctioning of connecting sequential elements, and should therefore be carefully monitored. The most important effect is an increase in delay.

Design Techniques to Deal with Capacitive Cross talk:

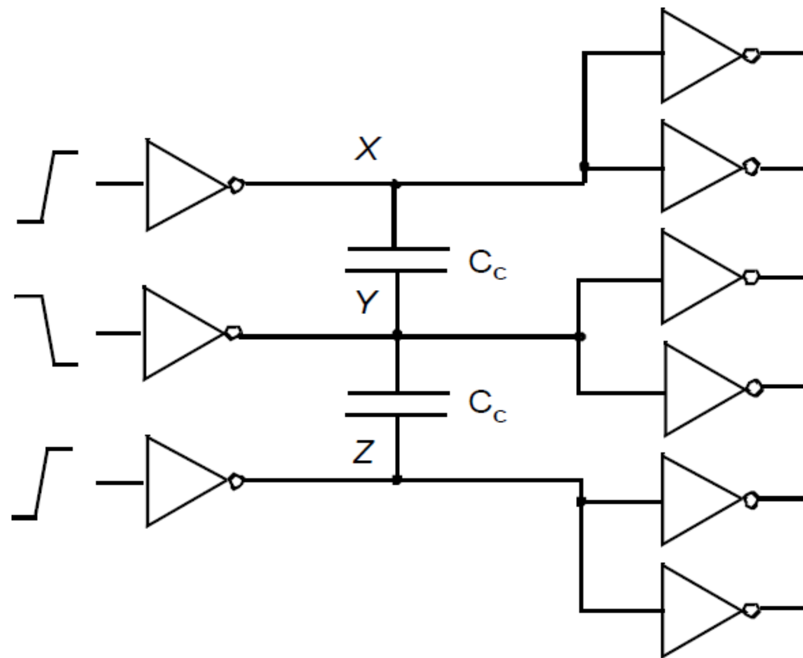
1. If possible avoid floating nodes, nodes sensitive to cross talk problems such as pre charged busses,should be equipped with keeper devices to reduce the impedance.
2. Sensitive nodes should be well separated from full swing signals.
3. Making the rise- fall time as large as possible subjection to timing constraints.
4. Use differential signalling in sensitive low swing wiring networks. This turns the cross talk signal into a common mode noise source that does not impact the operation of the circuit.
5. To keep the cross talk minimum, do not allow the capacitance between the two signal wires to grow too large.
6. If necessary provide shielding wire- GND or V_{DD} between the two signals as show below. This effectively turns the interwire capacitance into a capacitance to ground and eliminates interference. An adverse effect of shielding is the increased capacitive load.



CROSS SECTION OF ROUTING LAYERS ILLUSTRATING THE USE OF SHIELDING TO REDUCE CAPACITIVE CROSS TALK

7. The interwire capacitance between signals on different layers can be further reduced by addition of extra routing layers.

- **Impact of Cross talk on Propagation Delay (With respect to CMOS):**



IMPACT OF CROSS TALK ON PROPAGATION DELAY

The circuit schematic illustrates of how capacitive cross talk may result in a data-dependent variation of the propagation delay. Assume that the inputs to the three parallel wires X, Y, and Z experience simultaneous transitions. Wire *Y* (called the victim wire) switches in a direction that is opposite to the transitions of its neighbouring signals *X* and *Z*. The coupling capacitances experience a voltage swing that is double the signal swing, and hence represent an effective capacitive load that is twice as large as C_c- the by now well known *Miller effect*.
Since the coupling capacitance represents a large fraction of the overall capacitance in the deep-submicron dense wire structures, this increase in capacitance is substantial, and has a major impact on the propagation delay of the circuit. Observe that this is a worst-case scenario. If all inputs experience a simultaneous transition in the same direction, the voltage over the coupling capacitances remains constant, resulting in a zero contribution to the effective load capacitance.

The total load capacitance C_L of gate *Y*, hence depends upon the data activities on the neighbouring signals and varies between the following bounds:

$$C_{GND} \leq C_L \leq C_{GND} + 4C_c$$

with C_{GND} the capacitance of node *Y* to ground, including the diffusion and fan out capacitances.

Design Techniques for Circuit Fabrics with Predictable delay:

With cross talk making wire-delay more and more unpredictable, a designer can choose between a numberof different methodology options to address the issue, some of which are,

1. Evaluate and improve: After detailed extraction and simulation, the bottlenecks in delay are identified,and the circuit is appropriately modified.
2. Constructive layout generation: Wire routing programs take into account the effects of the adjacent wires,ensuring that the performance requirements are met.
3. Predictable structures: By using predefined, known, or conservative wiring structures, the designer is that the circuit will meet his specifications and that cross talk will not be a show stopper.

- **Capacitive Load (With respect to CMOS):**

The increasing values of the interconnect capacitances, especially those of the global wires, emphasize the need for effective driver circuits that can (dis)charge capacitances with sufficient speed. This need is further highlighted by the fact that in complex designs a single gate often has to drive a large fan-out and hence hasa large capacitive load.

Typical examples of large on-chip loads are busses, clock networks, and control wires. The latter include,for instance, reset and set signals. These signals control the operation of a large number of gates, so fan-outis normally high. Other examples of large fan-outs are encountered in memories where a large number of storage cells is connected to a small set of control and data wires.

The capacitance of these nodes is easily in the multi-pico farad range. The worst case occurs when signalsgo off-chip. In this case, the load consists of the package wiring, the printed circuit board wiring, and the input capacitance of the connected ICs or components.

Typical off-chip loads range from 20 to 50 pF, which is multiple thousand times larger than a standard on- chip load. Driving those nodes with sufficient speed becomes one of the most crucial design problems.

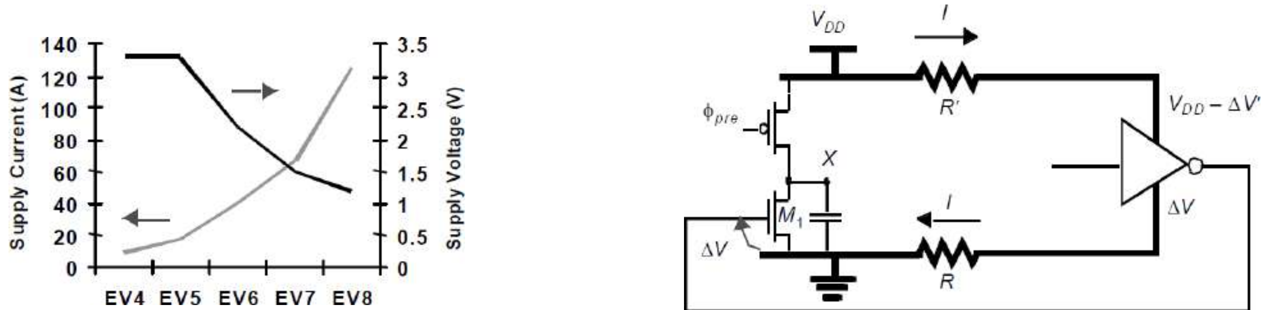
The main secrets to the efficient driving of large capacitive loads are:

1. Adequate transistor sizing is instrumental when dealing with large loads.
2. Partitioning drivers into chains of gradually-increasing factors.

RESISTIVE PARASITICS:

➢ Resistance and Reliability- Ohmic Voltage Drop:

Current flowing through a resistive wire results in an ohmic voltage drop that degrades the signal levels. This is especially important in the power distribution network, where current levels can easily reach amperesas shown below.



EVOLUTION OF POWER SUPPLY CURRENT AND SUPPLY VOLTAGE OHMIC VOLTAGE DROP ON THE SUPPLY REDUCES NOISE MARGIN

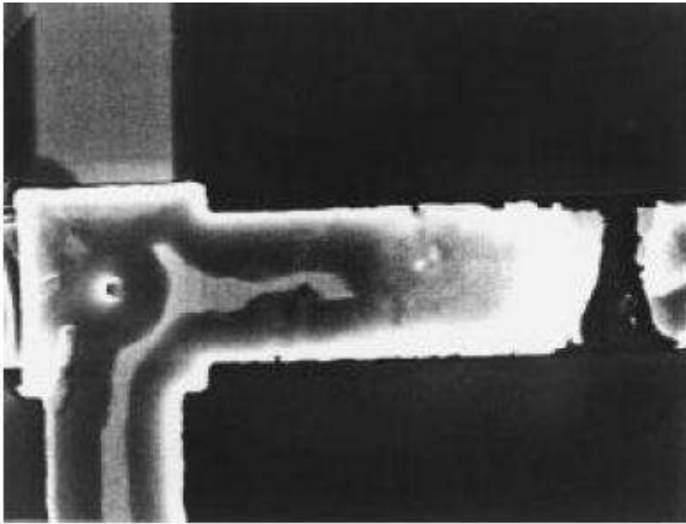
Consider a 2 cm long V_{DD} or *GND* wire with a current of 1mA per µm width. This current is about the maximum that can be sustained by an aluminum wire due to *electromigration* and assuming asheet resistance of 0.05 Ω/sq, the resistance of this wire (per µm width) equals 1 kΩ. A current of 1 mA/µm would result in a voltage drop of 1 V. The altered value of the voltage supply reduces noise margins and changes the logic levels as a function of the distance from the supply terminals. This is demonstrated by the circuit shown above, where an inverter placed far from the power and ground pins connects to a devicecloser to the supply.

The difference in logic levels caused by the *IR* voltage drop over the supply rails might partially turn on transistor M_1. This can result in an accidental discharging of the pre charged, dynamic node *X*, or causestatic power consumption if the connecting gate is static. In short, the current pulses from the on-chip logic, memories and I/O pins cause voltage drops over the power- distribution network and are the major sourcefor on- chip power supply noise. Beyond causing a reliability risk, IR drops on the supply network also impact the performance of the system. A small drop in the supply voltage may cause a significant increase indelay.

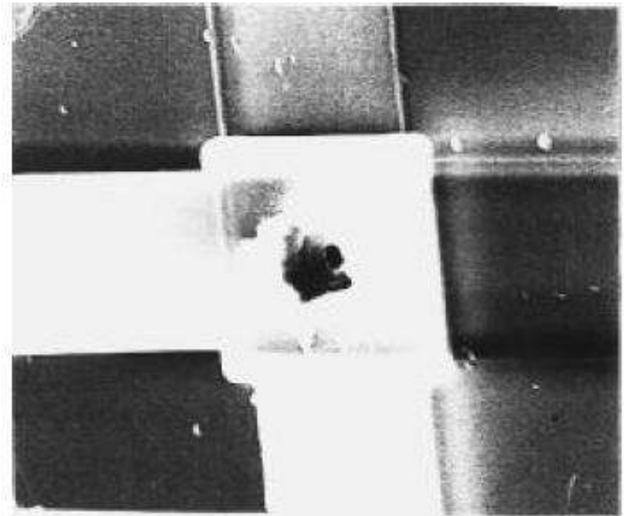
The most obvious problem is to reduce the maximum distance between the supply pins and the circuit supply connections which is most easily accomplished through a structured layout of the power distribution network. A number of on- chip power distribution networks with peripheral bonding.

➢ Electromigration:

The current density (current per unit area) in a metal wire is limited due to an effect called *electromigration*. A *direct* current in a metal wire running over a substantial time period, causes a transport of the metal ions. Eventually, this causes the wire to break or to short circuit to another wire. This type of failure will onlyoccur after the device has been in use for some time.



Line Open Failure



Open Failure in Contact Plug

ELECTROMIGRATION RELATED FAILURE MODES

The rate of the electromigration depends upon the temperature, the crystal structure, and the average current density. The latter is the only factor that can be effectively controlled by the circuit designer. Keeping the current below 0.5 to 1 mA/ µm normally prevents migration. This parameter can be used to determine the minimal wire width of the power and ground network. Signal wires normally carry an ac- current and are less susceptible to migration. The bidirectional flow of the electrons tends to anneal any damage done to the crystal structure. Most companies impose a number of strict wire-sizing guidelines on their designers, based on measurements and past experience.

Electromigration effects are proportional to the average current flow through the wire, while IR voltage drops are a function of the peak current.

From designing point of view, at the technology level, a number of precautions can be taken to reduce themigration risk i.e.

1. To add alloying elements (such as Cu or Tu) to the aluminum to prevent the movement of the Alions.
2. To control the granularity of the ions.
3. The introduction of new interconnect materials is a big help as well. For instance, the use of Copper interconnect increases the expected lifetime of a wire with a factor of 100 over Al.

➢ **Resistance and Performance- RC Delay:**

The delay of a wire grows quadratically with its length. Doubling the length of a wire increases its delay by a factor of four. The signal delay of long wires therefore tends to be dominated by the *RC* effect. This is becoming an ever larger problem in modern technologies, which feature an increasing average length of the global wires, at the same time that the average delay of the individual gates is going down. This leads to the rather bizar situation that it may take multiple clock cycles to get a signal from one side of a chip to its opposite end.

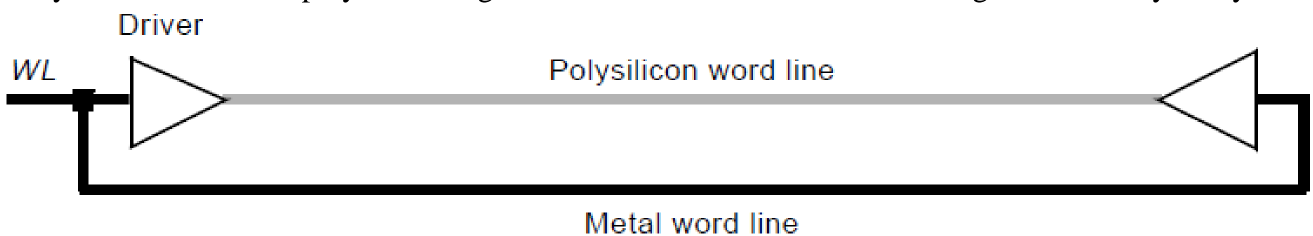
Providing accurate synchronization and correct operation becomes a major challenge under these circumstances. Therefore the different design techniques to cope with the delay imposed by the resistance ofthe wire are,

- **Better Interconnect Materials:**

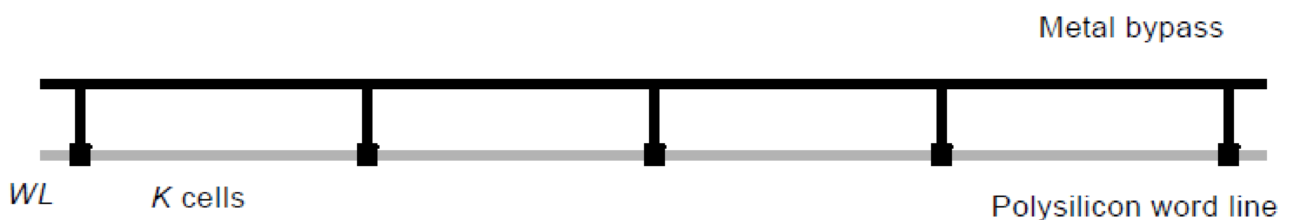
Use better interconnect materials when they are available and appropriate. The introduction of silicides and Copper have helped to reduce the resistance of polysilicon (and diffused) and metal wires, respectively, while the adoption of dielectric materials with a lower permittivity lowers the capacitance. Both Copper and low- permittivity dielectrics have become common in advanced CMOS technologies.

Here the designer should be aware that these new materials only provide a temporary respite of one or two generations, and do not solve the fundamental problem of the delay of long wires. Innovative design techniques are often the only way of coping with the latter.

Sometimes, it is hard to avoid the use of long polysilicon wires. A good example of such circumstance are the address lines in memories, which must connect to a large number of transistor gates. Keeping the wiresin polysilicon increases the memory density substantially by avoiding the overhead of the extra metal contacts. The polysilicon- only option unfortunately leads to an excessive propagation delay. One possible solution is to drive the word line from both ends, as shown in Figure. This effectively reduces the worst-casedelay by a factor of four. Another option is to provide an extra metal wire, called a *bypass*, which runs parallel to the polysilicon one, and connects to it every *k* cells as shown in figure. The delay is now dominated by the much shorter polysilicon segments between the contacts. Providing contacts only every *k*



(a) Driving the word line from both sides



(b) Using a metal bypass

cells helps to preserve the implementation density.

APPROACHES TO REDUCE THE WORD LINE DELAY

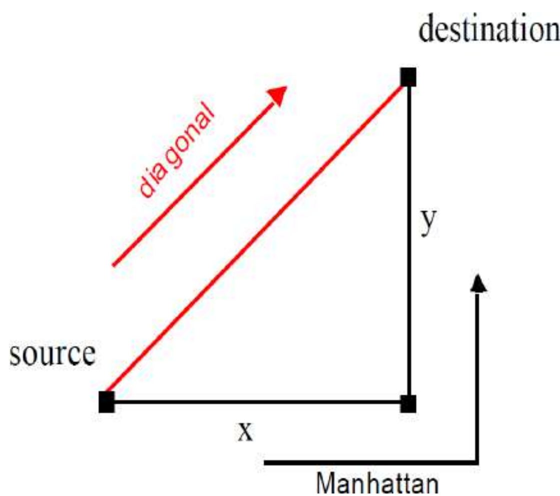
o **Better Interconnect Strategies:**

The length of the wire being a prime factor in both the delay and the energy consumption of an interconnectwire, any approach that helps to reduce the wire length is bound to have an essential impact.

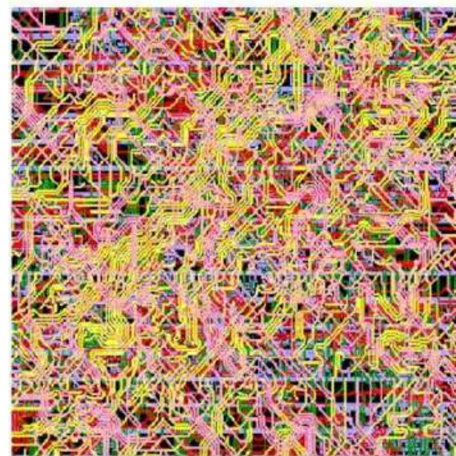
There are two wiring strategies i.e. the Manhattan- Style Routing and Diagonal- Style Routing.

- In Manhattan style routing, interconnections are first routed along the one of the preferred directions,followed by a connection in the other direction as shown.

- In Diagonal style routing less size of the wire length is required, on comparison to Manhattan 29% inbest case. And the use of 45°lines is ironical in integrated circuits. The main issues of diagonal routing are its complexity, impact on tools and masking concerns.



MANHATTAN VS DIAGONAL ROUTING

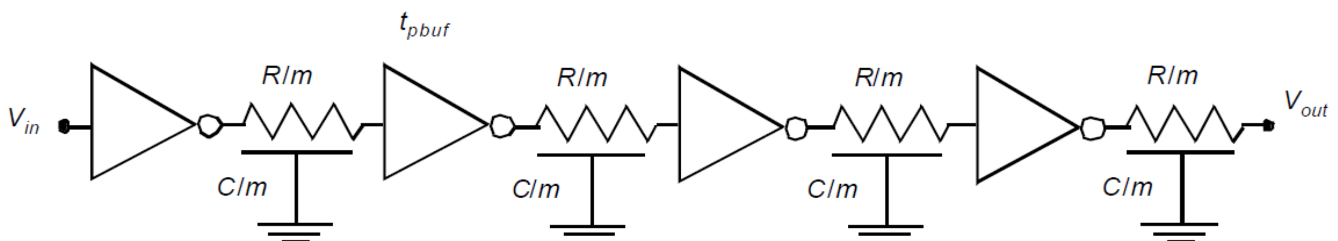


LAYOUT EXAMPLE OF 45• LINES (FOR UNDERSTANDING- NO NEED TO DRAW)

Earlier Manhattan routing was preferred because of the issues of diagonal routing inspite of its features. Now diagonal routing is preferred due to its features i.e. less wire length and 45° lines, its issues of complexity, impact on tools and masking concerns are easily overcomed nowadays by using CAD tools (Computer Aided Design Tools) like Cadence. Therefore the impact on wiring is quite tangible, a reduction of 20% in wire length, resulting in higher performance, lower power dissipation and smaller chip area.

o **Introducing Repeaters/ Buffer Insertion for very long wires:**

The most popular design approach to reducing the propagation delay of long wires is to introduce intermediate buffers, also called *repeaters*, in the interconnect line as shown below.



REDUCING RC INTERCONNECT DELAY BY USING REPEATERS

Making an interconnect line m times shorter reduces its propagation delay quadratically, and is sufficient to offset the extra delay of the repeaters when the wire is sufficiently long. Assuming that the repeaters have a fixed delay t_{pbuf}, we can derive the delay of the partitioned wire.

$$t_p = 0.38rc\left(\frac{L}{m}\right)^2 m + mt_{pbuf}$$

The optimal number of buffers that minimizes the overall delay can be found by setting $\frac{\partial t_p}{\partial m} = 0$,

$$m_{opt} = L\sqrt{\frac{0.38rc}{t_{pbuf}}} = \sqrt{\frac{t_{pwire(unbuffered)}}{t_{pbuf}}}$$

yielding a minimum delay of,

$$t_{p,opt} = 2\sqrt{t_{pwire(unbuffered)}t_{pbuf}}$$

and is obtained when the delay of the individual wire segments is made equal to that of a repeater.

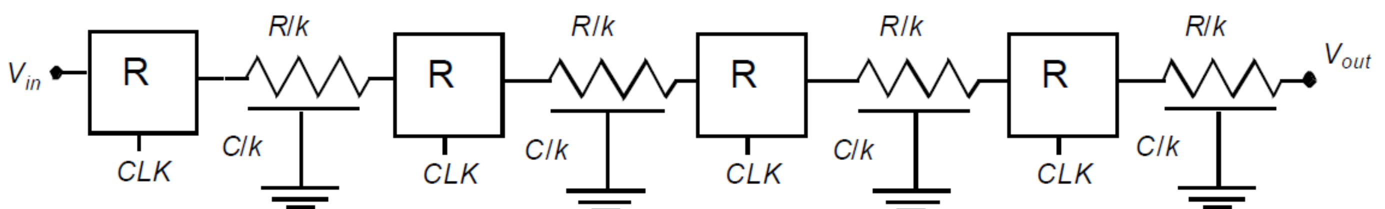
- **Optimizing the Interconnect Architecture:**

Even with buffer insertion, the delay of a resistive wire cannot be reduced below the minimum dictated by Equation,

$$t_{p,opt} = 2\sqrt{t_{pwire(unbuffered)}t_{pbuf}}$$

Long wires hence often exhibit a delay that is longer than the clock period of the design. For instance, the 10cm long Al wire of comes with a minimum delay of 4.7 nsec, even after optimal buffer insertion and sizing, while the 0.25 m CMOS process featured in this text can sustain clock speeds in excess of 1 GHz (this is, clock periods below 1 nsec). The wire delay all-by-itself hence becomes the limiting factor on the performance achievable by the integrated circuit. The only way to address this bottleneck is to tackle it at thesystem architecture-level.

Wire pipelining is a popular performance- improvement technique in this category which improves the throughput performance of logic modules with long critical paths. Similar approach can be used to increase the throughput of a wire, as is illustrated in figure below.



WIRE PIPELINING IMPROVES THE THROUGHPUT OF A WIRE

The wire is partitioned in k segments by inserting registers or latches. While this does not reduce the delay through the wire segment, it takes k clock cycles for a signal to proceed through the wire, it helps to increase its throughput, as the wire is handling k signals simultaneously at any point in time. The delay of the individual wire segments can further be optimized by repeater insertion, and should be below a single clock period.

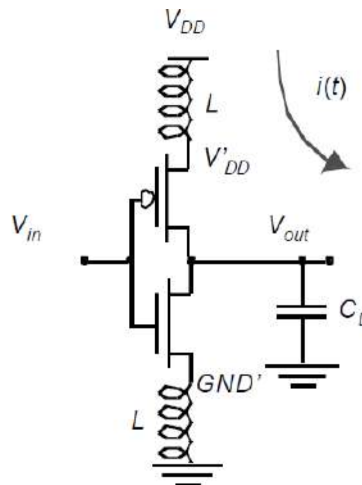
This is only one example of the many techniques that the chip architect has at her disposal to deal with the wire delay problem. The most important concern from this is that the wires have to be considered early on inthe design process, and can no longer be treated as an afterthought as was most often the case in the past.

INDUCTIVE PARASITICS:

Interconnect wires also exhibit an inductive parasitic. An important source of parasitic inductance is introduced by the bonding wires and chip packages. Even for intermediate- speed CMOS designs, the current through the input- output connections can experience fast transitions that cause voltage drops as well as ringing and overshooting, phenomena not found in RC circuits. At higher switching speeds, wavepropagation and transmission line effects can come into the picture.

- **Inductance and Reliability-** $\boldsymbol{L\frac{di}{dt} Voltage\ Drop}$**:**

During each switching action, a transient current is sourced from (or sunk into) the supply rails to charge (or discharge) the circuit capacitances as shown. Both V_{DD} and V_{SS} connections are routed to the external supplies through bonding wires and package pins and possess a non ignorable series inductance. Hence, a change in the transient current creates a voltage difference between the external and internal (*V'DD*, GND') supply voltages. This situation is especially severe at the output pads, where the driving of the large external capacitances generates large current surges. The deviations on the internal supply voltages affect the logic levels and result in reduced noise margins.

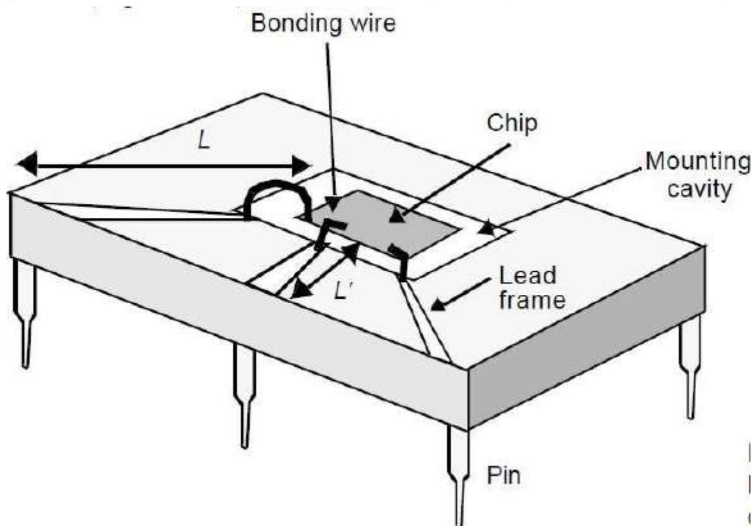


INDUCTIVE COUPLING BETWEEN EXTERNAL AND INTERNAL SUPPLY VOLTAGES

In an actual circuit, a single supply pin serves a large number of gates or output drivers. A simultaneous switching of those drivers causes even worse current transients and voltage drops. As a result, the internal supply voltages deviate in a substantial way from the external ones. For instance, the simultaneous switching of the 16 output drivers of an output bus would cause a voltage drop of at least 1.1 V if the supply connections of the buffers were connected to the same pin on the package. Improvements in packaging technologies are leading to ever-increasing numbers of pins per package. Packages with up to 1000 pins are currently available. Simultaneous switching of a substantial number of those pins results in huge spikes on the supply rails that are bound to disturb the operation of the internal circuits as well as other external components connected to the same supplies.

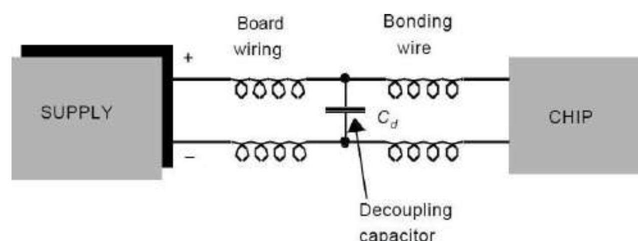
Design techniques to address $L\frac{di}{dt}$

1. Separate pins for I/O pads and chip core. Since the I/O drivers require the largest switching currents, they also cause the largest current changes. Therefore, it is wise to isolate the core of the chip where most of the logic action occurs, from the drivers by providing different power and ground pins.
2. Multiple power and ground pins in order to reduce the per supply pin, we can restrict the number of I/O drivers connected to a single supply pin.
3. Careful selection of positions of the power and ground pins on the package. The inductance of pins located at the corners of the package is substantially higher as shown below.



THE INDUCTANCE OF A BONDING WIRE/ PIN COMBINATION DEPENDS UPON THE PIN POSITIONS

4. Schedule current consuming transitions so that they do not occur simultaneously.
5. Increase the rise and fall times of the off-chip signals to the maximum extent allowable, and distributed all over the chip, especially under the data busses.
6. Use advanced packaging technologies such as surface-mount or hybrids that come with a substantially reduced capacitance and inductance per pin.
7. Adding decoupling capacitances on the board. These capacitances, which should be added for every supply pin, act as local supplies and stabilize the supply voltage seen by the chip. They separate the bonding- wire inductance from the inductance of the board interconnect as shown below. The bypasscapacitor, combined with the inductance, actually acts as a low- pass network that filters away the high-frequency components of the transient voltage spikes on the supply lines.



DECOUPLING CAPACITORS ISOLATE THE BOARD INDUCTANCE FROM THE BONDING WIRE AND IN INDUCTANCE

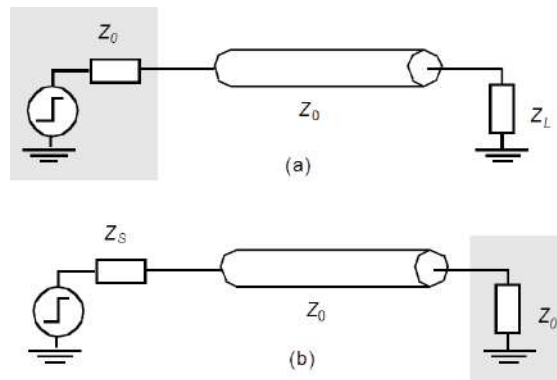
- **Inductance and Performance- Transmission Line Effects:**

When an interconnection wire becomes sufficiently long or when the circuits become sufficiently fast, the inductance of the wire starts to dominate the delay behaviour, and transmission line effects must beconsidered. This is more precisely the case when the rise and fall times of the signal become comparable to the time of flight of the signal waveform across the line as determined by the speed of light. As advancing technology increases line lenghts and switching speeds, this situation is gradually becoming common in fastest CMOS circuits as well, and transmission- line effects are bound to become a concern of the CMOS designer as well.

Some of the techniques to minimize the impact of the transmission line behaviour are:

- **Termination:**

To avoid the negative effects of transmission-line behaviour such as ringing or slow propagation delays, the line should be terminated, either at the source (series termination), or at the destination (parallel termination) with a resistance matched to its characteristic impedance Z_0.



MATCHED TERMINATION SCENARIOS FOR WIRES BEHAVING AS TRANSMISSION LINES: (a) SERIES TERMINATION AT THESOURCE, (b) PARALLEL TERMINATION AT THE DESTINATION

The two scenarios- series and parallel termination as shown are depicted in figure. Series termination requires that the impedance of the signal source is matched to the connecting wire. This approach is appropriate for many CMOS designs, where the destination load is purely capacitive. The impedance of the driver inverter can be matched to the line by careful transistor sizing.

- **Shielding**

If we want to control the behaviour of a wire behaving as a transmission line, we should carefully plan and manage how the return current flows. A good example of a well-defined transmission line is the coaxial cable, where the signal wire is surrounded by a cylindrical ground plane. To accomplish similar effects on a board or on a chip, designers often surround the signal wire with ground (supply) planes and shielding wires. Being shielding, adding shielding makes the behaviour and the delay of an interconnection a lot more predictable. Yet even with these precautions, powerful extraction and simulation tools will be needed in the future for the high-performance circuit designer.

UNIT – IV

INTERCONNECT, MEMORY ARCHITECTURE AND ARITHMETICCIRCUITS
TWO MARK QUESTIONS AND ANSWERS

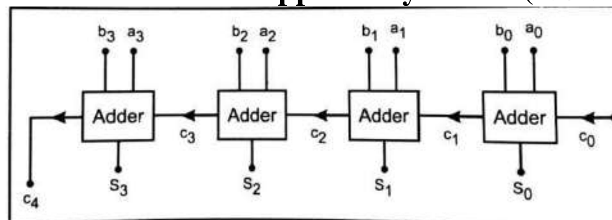
1. What is meant by data path circuits? (APR 2016)

- Data path circuits are meant for passing the data from one segment to other segment for processing or storing.
- The data path is the core of processors, where all computations are performed.

2. What is ripple carry adder?

- If n bits are added, then we can get n-bit sum and carry of C_n. C_i= Carry in bit from the previous column. N bit ripples carry adder needs n full adders with C_{i+1} carry out bit.

3. Draw the circuit for 4 bit ripple carry adder. (NOV 2018)



4. Write the equation for total delay in 4 bit ripple carry adder.

The total delay using the following equation,

$t_{4b} = t_d (c_{in} \rightarrow S_3) + 2t_d (c_{in} \rightarrow c_{out}) + t_d (a_0, b_0 \rightarrow c_1)$

5. Write the equation for worst case delay in 4 bit ripple carry adder.

If it is extend to n-bit, then the worst case delay is

$t_{n\text{-}bit} = t_d(c_{in} \rightarrow S_{n-1}) + (n-2)t_d(c_{in} \rightarrow c_{out}) + t_d(a_0,b_0 \rightarrow c_1)$

6. What is meant by Carry Lookahead Adder (CLA)?

- A carry-lookahead adder (CLA) or fast adder is a type of adder used in digital logic.
- A carry-lookahead adder improves speed by reducing the amount of time required to determine carry bits.
- The carry-lookahead adder calculates one or more carry bits before the sum, which reduces the wait time to calculate the result of the larger value bits.

7. Write the general expression for carry signal in CLA.
Write the full adder output interms of propagate and generate. (April 2018)

- We can write carry look-ahead expressions in terms of the generate gi and propagate pi signals. The general form of carry signal ci thus becomes

$c_{i+1} = a_i.b_i + c_i.(a_i \oplus b_i) = g_i + c_i.p_i$, If $a_i.b$ =1, then $c_{i+1} =$ 1,

8. Write the equation for generate term in CLA.

- In the case of binary addition, A + B generates if and only if both A and B are 1. If we write G(A,B) to represent the binary predicate that is true if and only if A + B generates, write generate term as, $g_i = a_i.b_i$

82

9. Write the equation for propagates term in CLA.

- In the case of binary addition, A + B propagates if and only if at least one of A or B is 1. If we write P(A,B) to represent the binary predicate that is true if and only if A + B propagates, we have:Write the propagate term as, $p_i = a_i \oplus b_i$

10. What are the two factors that Carry lookahead adder depends on?

- Carry lookahead depends on two things:
  - Calculating, for each digit position, whether that position is going to propagate a carry if one comes in from the right.
  - Combining these calculated values to be able to deduce quickly whether, for each group of digits, that group is going to propagate a carry that comes in from the right.

11. Write the generalized equation for CLA.

$$S_i = P_i \oplus c_i$$
$$c_1 = g_0 + P_0.c_0$$
$$c_2 = g_1 + P_1.c_1 = g_1 + P_1.(g_0 + P_0c_0)$$
$$c_3 = g_2 + P_3.c_3$$
$$c_3 = g_3 + P_3.c_3$$
$$= g_3 + P_3.g_2 + P_3.P_2.g_1 + P_3.P_2.P_1.g_0 + P_3.P_2.P_1.P_0.C_0$$

12. Name the limitations of MODL.

- MODL has following limitations as
  i. clocking in mandatory
  ii. The output is subject to charge leakage and charge sharing.
  iii. Series connected nFET chains can give long discharge times.

13. What is called Manchester Carry Chain Adder?

- The Manchester carry chain is a variation of the carry-lookahead adder that uses shared logic to lower the transistor count.
- As seen in CLA implementation section, the logic for generating each carry contains all of the logic used to generate the previous carries.
- A Manchester carry chain generates the intermediate carries by tapping off nodes in the gate that calculates the most significant carry value.

14. Write the basic equation for Manchester Carry Chain Adder?
Define kill term, propagate and generate term in a carry look ahead adder. (April 2019)

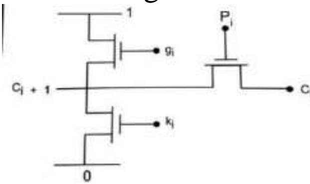
- In this adder, the basic equation is $c_{i+1} = g_i + c_i.p_i$

  Where $p_i = a_i \oplus b_i$ and $g_i = a_i.b_i$
- Carry kill bit $k_i = \overline{a_i + b_i} = \overline{a_i}.\overline{b_i}$
- If K_i=1, then p_i=0 and g_i=0. Hence, k_i is known as carry kill bit.

15. Draw the switch level circuit for Manchester carry chain adder.

83

The switch level circuit is given as



16. What are high (wide) adders?

- The adders with more than 4 bits can be designed. This is known as wide or high speed adders. Brute-force approach can be used to design 8 bit adder.

17. What are the types of high speed adders?

Types of high speed adders are

1. Carry Skip adder
2. Carry Select adder
3. Carry Save adder.

18. What is Carry skip adder?

- Carry skip adder is one of the high speed adders.
- When BP= $P_0P_1P_2P_3$ =1, the incoming carry is forwarded immediately to the next block.
- Hence the name carry bypass adder or carry skip adder.
- Idea: if (P_0 and P_1 and P_2 and P_3 =1) the C_{03} = C_0, else "kill" or "generate".

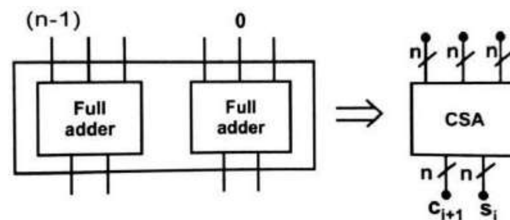
19. What is Carry Select adder?

Write the principle of any one fast multiplier. (NOV 2016)

- Adding two n-bit numbers with a carry-select adder is done with two adders in order to perform the calculation twice.
- After the two results are calculated, the correct sum, as well as the correct carry-out, is then selected with the multiplexer once the correct carry-in is known.

20. What is Carry save adder?

- In carry save adder, the carry does not propagate. So, it is faster than carry propagate adder.
- It has three inputs and produces 2 outputs, carry-out is saved. It is not immediately used to find the final sum value.



n-bit Carry Save Adder

21. What are accumulators?

- Accumulator acts as a part of ALU and it is identified as register A. The result of an operation performed in the ALU is stored in the accumulator.
- It is used to hold the data for manipulation (arithmetic and logical)

84

22. What are multipliers?

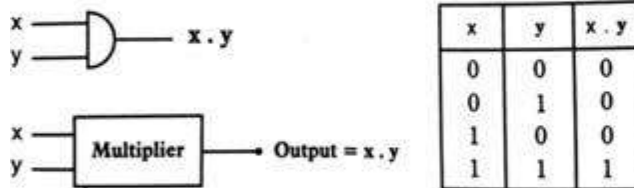
- Multiplier is used in computation process, which multiplies two binary numbers.
- Basic operations in multiplication are given below.

  0 x 0 = 0, 0 x 1 = 0, 1 x 0 = 0, 1 x 1 = 1

.

23. Draw the truth table of multiplier.

The truth table of multiplier is



x	y	x . y
0	0	0
0	1	0
1	0	0
1	1	1

24. Mention the steps involved in multiplying by shifting.

- If x=(0010)2 = (2)10
- If it is to be multiplied by 2, then we can shift x in left side. x = (0100)2 = (4)10
- If it is to be divided by 2, then we can shift in right side. x = (0001)2 = (1)10.
- So, shift register can be used for multiplication or division by 2.

25. Write the delay equation for array multiplier.

The equation for array multiplier is

$$P_i = \sum_{i=j+k} x_j y_k + c_{i-1}$$

$$P = X.Y = \left(\sum_{j=0}^{n-1} x_j 2^j\right)\left(\sum_{k=0}^{n-1} y_k 2^k\right) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\left(x_j y_k 2^{j+k}\right)$$

26. State radix-2 booth encoding table. (April 2019)

In radix-2 booth multiplication partial product generation is done based on encoding which is as given by Table.

Q_n	Q_{n+1}	Recoded Bits	Operation
0	0	0	Shift
0	1	+1	Add X
1	0	-1	Subtract X
1	1	0	Shift

Table: Booth encoding table with RADIX-2

27. What is meant by divider circuit?

- Divider circuit is used in arithmetic operation in digital circuits. Dividing is carry out by repeated subtraction and addition.

28. What are the types of dividers available in VLSI?

There are two types of dividers. They are serial divider and parallel divider.

85

29. Compare serial divider and parallel divider.

- Serial divider is slow and parallel divider is fast in performance. Array divider is fast compared with the serial divider. But hardware requirement is increased.

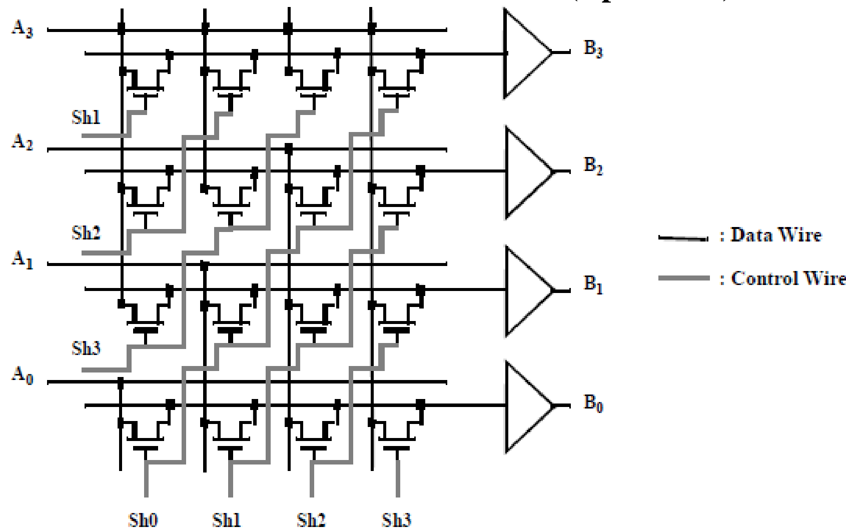
30. What is shift register?

- An n-bit rotation is specified by using the control word $R_{0\text{-}n}$ and L/R bit defines a left or right shifting.
- For example y3 y 2 y 1 y 0 = a3 a2 a1 a0
  If it is rotated 1-bit in left side, we get y3 y 2 y 1 y 0 = a2 a1 a0 a3
  If it is rotated 1-bit in right side, we get y3 y 2 y 1 y 0 = a0 a3 a2 a1

31. What is meant by Barrel shifter?

- A barrel shifter is a digital circuit that can shift a data word by a specified number of bits in one clock cycle.
- It can be implemented as a sequence of multiplexers (MUX). The output of one MUX is connected to the input of the next MUX in a way that depends on the shift distance.

32. Draw the structure of 4 X 4 barrel shifter. (April 2018)



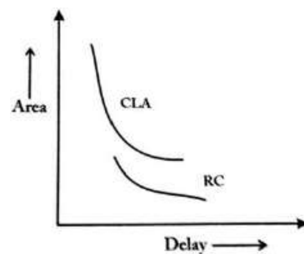
33. What is the area constraint between carry lookahead adder and ripple carry adder?

- The area of a carry lookahead adder is larger than the area of a ripple carry adder.
- Carry lookahead adder are parallel, which requires a larger number of gates and also results in a larger area.

34. What is the drawback of carry lookahead adder?

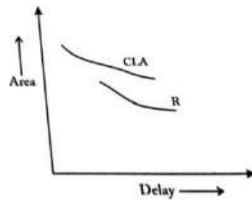
- In the carry lookahead adder, need large area because computations are in parallel and more power is consumed.

35. Draw the graph between area Vs delay of carry lookahead and ripple carry adder for 8 bit.

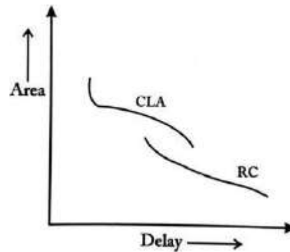


86

36. Draw the graph between area Vs delay of carry lookahead and ripple carry adder for 16 bit.



37. Draw the graph between area Vs delay of carry lookahead and ripple carry adder for32 bit.



38. What is meant by bit – sliced data path organization?(May 2016)

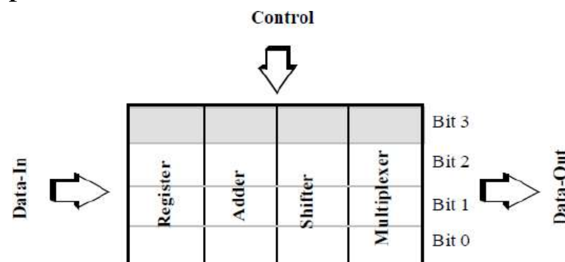
- Datapaths are arranged in a bit sliced organization, instead of operating on single bit digital signals.
- The data in a processor are arranged in a word based fashion. Bit slices are either identical or resemble a similar structure for all bits.

39. Determine propagation delay of n-bit carry select adder. (May 2016)

✓ Propagation delay, P of n-bit carry select adder is equal to $\sqrt{2N}$ where N = N- bit adder

40. Draw and list out the components of data path. (May 2017)

- Data path block consists of arithmetic operation, logical operation, shift operation and temporary storage of operands.



41. Mention the application of Barrel shift register.
Why is barrel shifter very useful in the designing of arithmetic circuits? (NOV 2016)

- A common usage of a barrel shifter is in the hardware implementation of floating-point arithmetic.
- For a floating-point add or subtract operation, requires shifting the smaller number to the right.
- This is done by using the barrel shifter to shift the smaller number to the right by the difference, in one cycle.

42. What is latency? (Nov 2017)

- Clock latency (or clock insertion delay) is defined as the amount of time taken by the clock signal in traveling from its source to the sinks.

43. Give the applications of high speed adder. (May 2017)

- CMOS high speed adder adders used in processor, data processing application and data path application with low power consumption.

44. What is meant by booth multiplier?

- Booth's algorithm is an efficient hardware implementation of a digital circuit that multiplies two binary numbers in two's complement notation.
- Booth multiplication is a fastest technique that allows for smaller, faster multiplication circuits, by recoding the numbers that are multiplied.

45. What is meant by array multiplier?

- Array multiplier uses an array of cells for calculation.
- Multiplier circuit is based on repeated addition and shifting procedure. Each partial product is generated by the multiplication of the multiplicand with one multiplier digit.
- N-1 adders are required where N is the number of multiplier bits.

46. What is Wallace tree multiplier? And give its advantages.

The Wallace tree multiplier output structure is tree basis style. It reduces the number of components and reduces the area.

47. What are parameters used to characterize the memory?

Parameters used to characterize a memory device are area, power and speed.

48. How can you classify memory based on operation mode?

Classifications of memory based on operation mode are 1. ROM , 2. RAM

49. How can you classify memory based on data storage mode?

Classifications of memory based on data storage mode are 1. Volatile, 2. Non-volatile

50. Define ROM. Give some examples.

ROM is a memory where code is written only one time. Examples are washing machine, calculator, games etc.

51. What are advantage and disadvantages of programming ROM?

Advantage: basic cell only consists of transistor. No need of connection to any of the supply voltage.

Disadvantage: As it has pseudo nMOS, it is ratioed logic and consumes static power.

52. What is meant by non-volatile memory?

Non-volatile consists of array of transistors. These are placed on a word line – bit line grid. We can write the program by enabling or disabling these devices selectively.

53. What is floating gate transistor?

Floating gate transistor is mostly used in all the reprogrammable memories. In floating gate transistor, extra polysilicon strip is used in between the gate and the channel known as floating gate.

54. What is RAM? And give types of RAM.

RAM is read and write memory. Types are static and dynamic RAM.

55. Distinguish Static and dynamic RAM.[Nov/Dec 2022]

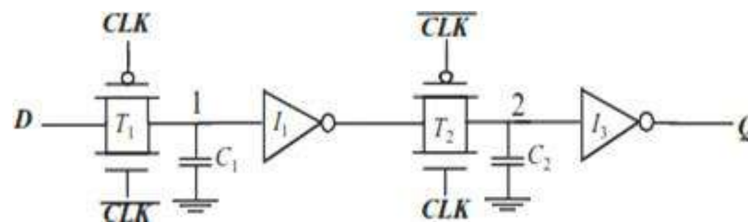
Static RAM:

SRAM cell needs 6 transistors per bit. Bit line (BL) and inverse Bit Line signals are used to improve the noise margin during read and write operations.

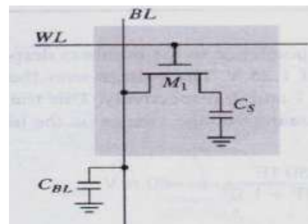
Dynamic RAM:

Content in the cell can be periodically rewritten through a resistive load, called as refresh operation. Here cell content is read follow by write operation.

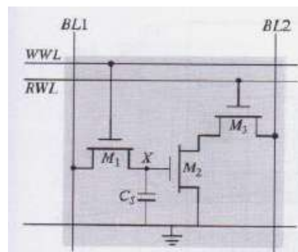
56. Draw the schematic of dynamic edge –triggered register. (Dec. 2016)



57. Design a one transistor DRAM cell. (Nov 2013, April 2015)
Draw a 1-transistor Dynamic RAM cell. (April 2019) [Nov/Dec 2022]



58. Design a three transistors DRAM cell.



59. State the merits of barrel shifter. (Nov 2019)

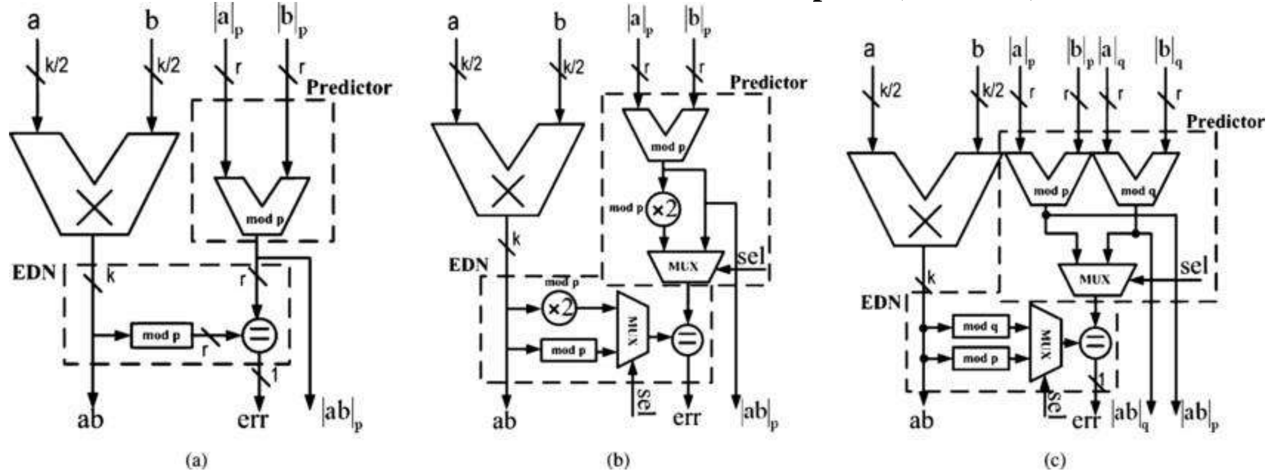
- It has small area and does not require a decoder
- Logarithmic shifter is more effective for larger shift values in terms of both area and speed.

60. How to design a high speed adder? (Nov 2017)

- Design of high speed adder using CMOS and transmission gates in submicron.
- Design of high speed adder using parallel adder manner.

89

61. Mention the different hardware architecture used for multiplier. (Nov 2019)



Hardware architectures for multipliers protected by (a) linear arithmetic codes, (b) [jxj ; j2xj] multilinear codes, and (c) multi-modulus multilinear codes.

62. Draw the dot diagram for Wallace tree multiplier. [May 2021]

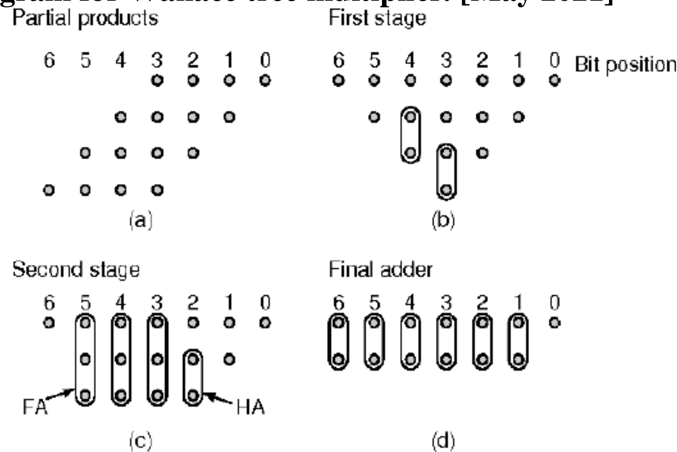


Figure 2: Dot diagram of 4-bit Wallace tree multiplier

63. List the categories of memory arrays. [May 2021]

Classification based on operation mode:

1. ROM 2. RAM

Classification based on data storage mode:

It means on how it is stored and how long it remains there.

1. Volatile 2. Non-volatile

Classification based on access method:

1. Random access 2. Non-random access

64. State the need of a sense amplifier in a memory cell. (NOV 2021)[Apr/May 2022]

It senses the low power signals from a bitline that represents a data bit (1 or 0) stored in a memory cell, and amplify the small voltage swing to recognizable logic levels so the data can be interpreted properly.

90

65. What are the building blocks of digital architecture?

These blocks include arithmetic circuits, counters, shift registers, memory arrays, and logic arrays.

66. Mention the steps for single bit addition.

1-Bit Adder (Half Adder)

- The simplest case arises when two one bit numbers are to be added.
- With one bit, only the numbers 0 and 1 can be represented.
- All possible scenarios can be summarized by the following table:

91

UNIT – IV

INTERCONNECT, MEMORY ARCHITECTURE AND ARITHMETICCIRCUITS

Question Bank

1. **Design a multiplier for 5 bit by 3 bit. Explain its operation and summarize the numbers of adders. Discuss it over Wallace multiplier.**
2. Describe the different approaches of improving the speed of the adder.
3. **How the drawback in ripple carry adder overcome by carry look ahead adder and discuss.**
4. Draw the ripple carry adder & explain its operation.
5. **Design a carry bypass adder and discuss its features.**
6. **Design 4 input and 4 output barrel shifter using NMOS logic.**
7. Explain in detail about the interconnect.
8. Describe about interconnect modelling.

92

UNIT V ASIC DESIGN AND TESTING

Introduction to wafer to chip fabrication process flow. Microchip design process & issues in test and verification of complex chips, embedded cores and SOCs, Fault models, Test coding. ASIC Design Flow, Introduction to ASICs, Introduction to test benches, Writing test benches in Verilog HDL, Automatic test pattern generation, Design for testability, Scan design: Test interface and boundary scan.

Introduction:

- **ASIC** - Application Specific Integrated Circuit is an Integrated Circuit (IC) designed to perform a specific function for a specific application.
- **Levels of integration:**
  The levels of integration are:
  - ✓ SSI - Small scale integration
  - ✓ MSI - Medium scale integration
  - ✓ LSI - Large scale integration
  - ✓ VLSI - Very large scale integration
  - ✓ USLI - Ultra large scale integration
- **Implementation technology**
- The implementation technologies used in ASIC are:
  - ✓ TTL – Transistor Transistor Logic
  - ✓ ECL – Emitter Coupled Logic
  - ✓ MOS – Metal Oxide Semiconductor (NMOS, CMOS)

5.1: Types of ASICs

- ❖ **Explain about different types of ASICs with neat diagram. (April 2016, 2017, 2018)**
- ❖ **Write brief notes on: (a) Full custom ASIC (b) Semi custom ASIC (May 2010, May 2016)**
- ❖ **Compare the different types of ASICs. (Nov 2007, Nov 2008)**

- The ASICs are classified as follows:

I. Full-Custom ASICs

II. Semi-custom ASICs
   a. Standard-Cell–Based ASICs (CBIC)
   b. Gate-Array–Based ASICs (MPGA)
      i. Channeled Gate Array
      ii. Channelless Gate Array
      iii. Structured Gate Array

III. Programmable ASICs
   a. Complex Programmable Logic Devices (CPLD)
   b. Field-Programmable Gate Arrays (FPGA)

5.1.1 :Full-Custom ASICs

Explain the full custom ASICs.

- In full custom ASIC, engineer can design full logic cells in IC. So, this technique is known as Full custom ASIC technique.

- Engineer uses mixed analog and digital technique to manufacture IC. All the logic cells are specifically designed for one ASIC.

Uses of bipolar technology:

- The characteristics of bipolar components in the same IC are matched very well.
- But the characteristic of components in different IC are not matched well

Uses of CMOS:

- This is widely used technology to manufacture IC.
- Mixing of analog and digital function are integrated in the same IC for which CMOS technology suits well.
- Designers give importance to performance.
- When large volume is manufactured, overall cost will be reduced.
- In super computer, quality is important so this design is implemented.
- All mask layers are customized in a full-custom ASIC
- Generally, the designer lays out all cells by hand
- Some automatic placement and routing may be done
- Critical (timing) paths are usually laid out completely by hand
- Full-custom design offers the ***highest performance and lowest part cost*** (smallest die size) for a given design.
- The ***disadvantages*** of full-custom design include ***increased design time, complexity, design expense***, and ***highest risk.***
- Microprocessors (strategic silicon) were exclusively full-custom, but designers are increasingly turning to semicustom ASIC techniques in this area as well.

5.1.2 : Semi-custom ASICs – Design

❖ **Briefly explain the semi-custom Asics with its classification. (May 2016, NOV 2016)**

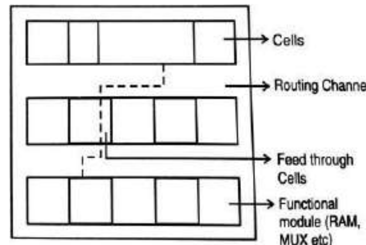
I. Standard cell based design:

- Standard cells are referred to AND gate, OR gate, multiplexer, flip flop, NOR gate etc.
- Standard cells can be used with larger predefined cells.
- This approach standardizes design entry level at logic gate.
- A design is generated automatically from HDL language.
- Then layout is created. In standard cell design, cells are placed in rows, and rows are separated by routing channel.
- All cells in library are in identical heights, widths of the cells can be varied to accommodate for variations in complexity between cells.
- A substantial fraction of area is allotted for signal routing.
- The minimization of interconnect overhead is most important goal of standard-cell placement routing tools. It is done by **feed through cells**.
- By using **feed through cell**, cells in different rows can be connected through vertical routing. So length of wire is reduced by feed through cells.

Semi-custom ASICs – CBIC

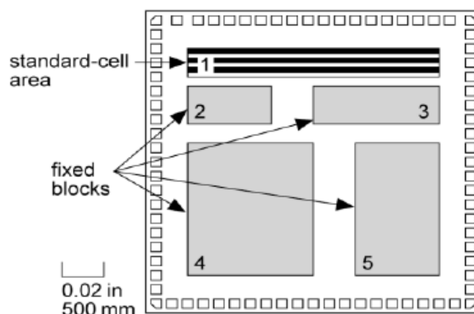
- CBIC means Cell Based ASICs.
- All the mask layers of CBIC are customized.

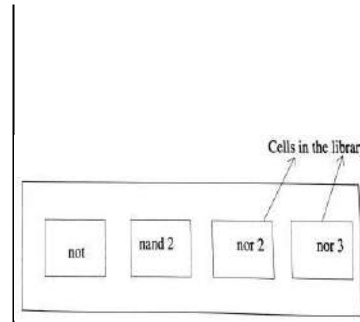
- It allows mega cells (SRAM, MPEG, decoder etc) to be placed in the same IC with standard cells (adder, gates etc).
- Mega cells are supplied by ASIC Company.
- Data path logic means the logic that operates on multiple signals across a data bus.
- Some of the ASIC library companies provide data path compiler which automatically generate data path logic.
- Data path library contains cells like adders, multiplexer, simple ALUs.
- ASIC Library Company provide data book which has functional description.



Features:

- It is a cell-based ASIC (CBIC —“sea-bick”)
- It has Standard cells. Standard cell is logic elements used CMOS technology.
- Possibly megacells , megafunctions , full-custom blocks , system-level macros (SLMs), fixed blocks , cores , or Functional Standard Blocks (FSBs)
- All mask layers are customized - transistors and interconnect
- Automated buffer sizing, placement and routing. And custom blocks can be embedded.





- A “wall” of standard cells forms a flexible block.

II. Gate Array Based ASICs:

❖ **Explain gate array based ASICs with diagrams. (April 2008, May 2009)**

- Gate array is known as GA.
- In GA based ASIC, the transistors are predefined on the silicon wafer.
- *Base array*: the predefined pattern of transistors on a gate is known as base array.
- *Base cell:* the small element which is replicated to make the base array is known as base cell or primitive cell.
- *Masked Gate array*: Interconnect is defined by using top few layers of metal.
- This type of gate array is known as masked gate array.
- Gate array library is provided by ASIC Company.
- The designer can choose the predefined logic cells from a gate array library. These logic cells are known as **Macros**.
- Cell-layout is same for each logic cell. But interconnect is customized.
- It is also called as pre-diffused array because the transistors are diffused at first.

Types of MPGAs (Mask Programmable Gate Arrays):

- ✓ Channeled Gate Array
- ✓ Channel less Gate Array
- ✓ Structured Gate Array

(a) Channeled Gate Array:

- It is similar to CBIC (cell based ASIC).
- In the both types, rows of cells are separated by channels. These channels are used for interconnect.
- Space between rows of cells is fixed in a channeled gate array. But space between rows of cells may be adjusted in a CBIC.

Features:

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

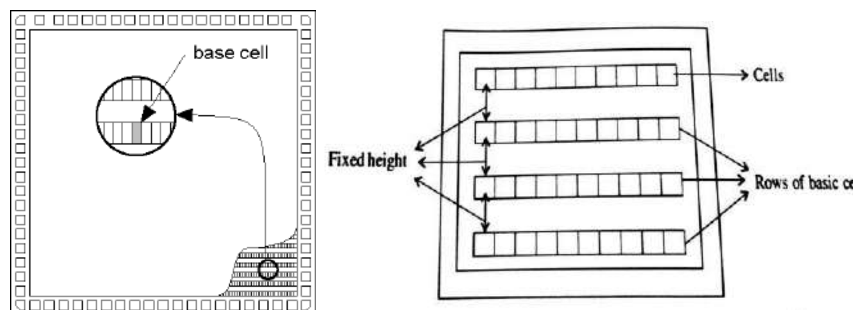


Figure: Channel Gate Array

(b) Channel less Gate Array:

- Channel less Gate Array is also called as channel free GA.
- In this array, there is no predefined space between rows for routing.
- Top few layers are used for defining interconnect connections.
- There are no predefined areas set aside for routing - routing is over the top of the gate-array devices.
- Achievable logic density is higher than for channeled gate arrays.
- Each logic cell or macro in a gate-array library is predesigned using fixed tiles of transistors known as the gate-array base cell (or just base cell).

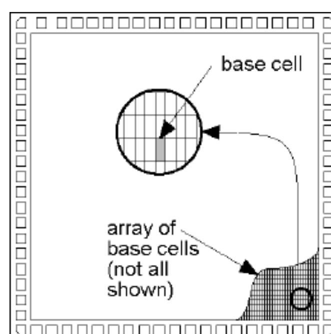


Figure: Channel less Gate Array

- Channeled and channelless gate arrays may use either gate isolation or oxide isolation.
- Isolate the transistors on a gate array from one another either with thick field oxide or by using other transistors that are wired permanently off.

(c) Structured Gate Array:

- Structured Gate Array is also called as embedded gate array or master slice or master image gate.
- It combines some of the features of CBIC and Masked gate array (MGA).
- In this array, some of the area is used for implementation of specially designed embedded block.
- Embedded area either can contain a different base cell that is more suitable for building memory cells, or a complete circuit block, such as a microcontroller.

Special features:

- o Only the interconnect is customized
- o Custom block can be embedded
- o Manufacturing lead time is 2 days to 2 weeks
- o Area efficiency is increased
- o Performance is increased with low cost

Disadvantages: the embedded function is fixed.

- For ex: if embedded block has 32K bit memory. But the customer needs only 18K bit, the 16K memory is wasted.

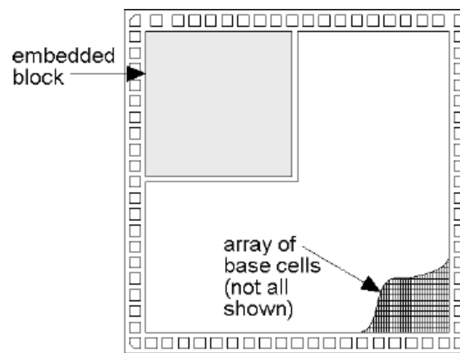


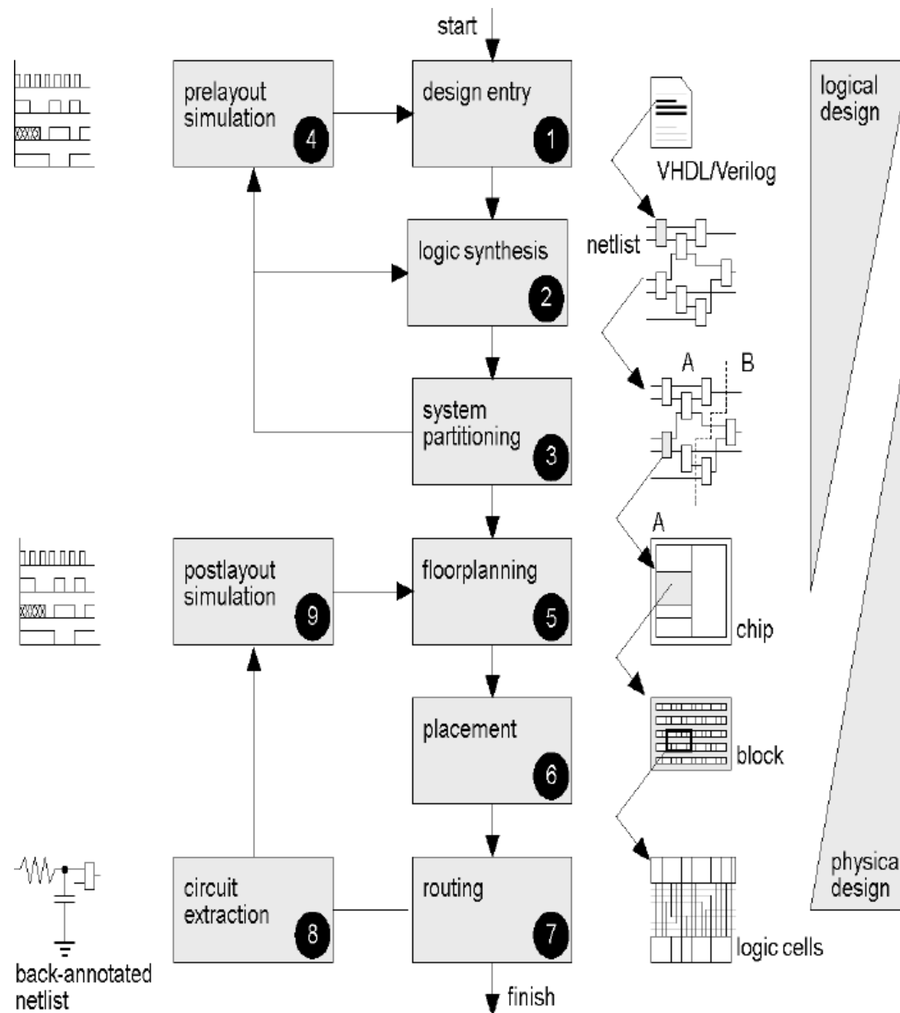
Figure: Structured Gate Array

****

5.2 : ASIC Design Flow / Cycle

Explain the ASIC design flow with a neat diagram. (Nov 2007, April 2008, Nov 2008)
Draw the flowchart of digital circuit design techniques. (NOV 2018)

1. **Design entry** - Using a hardware description language (HDL) or schematic entry
2. ***Logic synthesis*** - Produces a netlist - logic cells and their connections
3. ***System partitioning*** - Divide a large system into ASIC-sized pieces
4. **Prelayout simulation** - Check to see if the design functions correctly
5. ***Floorplanning*** - Arrange the blocks of the netlist on the chip
6. ***Placement*** - Decide the locations of cells in a block
7. ***Routing*** - Make the connections between cells and blocks
8. ***Extraction*** - Determine the resistance and capacitance of the interconnect
9. **Post layout simulation** - Check to see the design still works with the added loads of the interconnect



****

5.3 : ASIC Cell Libraries

Explain the standard Cell libraries in ASIC.
Give a note on standard cell design. (Nov 2019)

- The cell library is the key part of ASIC design.
- For a programmable ASIC the FPGA Company supplies a library of logic cells in the form of a design kit.
- For MGAs and CBICs you have three choices:
- ASIC vendor will supply a cell library, or you can buy a cell library from a third-party library vendor, or you can build your own cell library.
- The first choice, using an ASIC-vendor library, requires you to use a set of design tools approved by the ASIC vendor to enter and simulate your design.
- Some ASIC vendors (especially for MGAs) supply tools that they have developed in-house.
- An ASIC vendor library is normally a phantom library the cells are empty boxes, or phantoms, but contain enough information for layout.
- After you complete layout you hand off a netlist to the ASIC vendor,
- The second and third choices require making a buy-or-build decision.
- If complete an ASIC design using a cell library that you bought, you also own the masks that are used to manufacture your ASIC. This is called customer-owned tooling.

- A library vendor normally develops a cell library using information about a process supplied by an ASIC foundry.
- An ASIC foundry only provides manufacturing, with no design help. If the cell library meets the foundry specifications, we call this a qualified cell library.
- These cell libraries are normally expensive, but if a library is qualified at several foundries.
- The third choice is to develop a cell library in-house. Many large computer and electronics companies make this choice.
- However, created each cell in an ASIC cell library must contain the following:
  - ✓ A physical layout
  - ✓ A behavioral model
  - ✓ A Verilog/VHDL model
  - ✓ Detailed timing models.
  - ✓ A test strategy
  - ✓ A circuit schematic
  - ✓ A cell icon
  - ✓ A wire-load model
  - ✓ A routing models.

- The ASIC designer needs a high-level behavioral model for each cell.
- Because simulation at the detailed timing level takes too long for a complete ASIC design.
- The designer may require Verilog and VHDL models in addition to the models for a particular logic simulator.

****

5.4 : Library-Cell Design

Explain the important of Library –cell design in detail.

- Design rules for each ASIC vendor are slightly different even for the same generation of technology.
- For example, two companies may have very similar 0.35 nm CMOS process technologies, but the third-level metal spacing might be slightly different.
- A library constructed in this fashion may not be competitive with one that is constructed specifically for each process.
- ASIC vendors prize their design rules as secret, it turns out that they are similar except for a few details.
- We would like all vendors to agree on a common set of design rules.
- The reason that most vendors have similar rules is because most vendors use the same manufacturing equipment and a similar process.
- Layout of library cells is either hand-crafted or uses some form of symbolic layout.
- Symbolic layout is usually performed in one of two ways: using either interactive graphics or text layout language.
- Shapes are represented by simple lines or rectangles, known as sticks, in a symbolic layout.
- The actual dimensions of the sticks are determined after layout is completed in a Post processing step.
- Graphical symbolic layout uses a text layout language, like a programming language such as C that directs a program to assemble layout.

****

CMOS Testing

Explain about Microchip design process.

Microchip design process:

The microchip design process involves several stages from conceptualization to production. Here is an overview of the typical steps involved:

1. **Specification:** In this stage, the requirements and functionality of the microchip are defined. Designers work closely with stakeholders to understand the application and performance targets.

2. **Architecture Design:** The chip's high-level architecture is developed, including the selection of components, interconnections, and overall system design. This stage focuses on defining the chip's functionality and how different components will interact.

3. **RTL Design:** The Register Transfer Level (RTL) design is created, describing the chip's behavior using hardware description languages like Verilog or VHDL. RTL design forms the basis for later stages.

4. **Functional Verification:** The RTL design is extensively tested to ensure it behaves as intended. Various verification techniques, such as simulation, formal verification, and hardware emulation, are employed to catch design bugs and issues.

5. **Synthesis and Physical Design:** The RTL code is synthesized into a gate-level netlist, which represents the chip's physical implementation. The physical design phase involves floor planning, placement, routing, and optimization to meet timing and area constraints.

6. **Design for Testability (DFT):** Techniques like scan chains, built-in self-test (BIST) structures, and boundary scan are added to make the chip more testable during manufacturing and in the field.

7. **Manufacturing:** The final design is sent to a semiconductor foundry for fabrication. This process involves photolithography and other steps to create the actual silicon chip.

8. **Testing and Quality Assurance:** After manufacturing, the chips undergo various testing methodologies to ensure they meet the desired specifications and are free from defects.

****

Explain the Issues in Test and Verification of Complex Chips, Embedded Cores, and SoCs:

Issues in Test and Verification of Complex Chips, Embedded Cores, and SoCs:

1. **Complexity:** As chips and systems-on-chip (SoCs) become more complex, the verification effort increases exponentially. Ensuring all possible scenarios and corner cases are covered in testing becomes challenging.

2. **Verification Time and Cost:** With the growing complexity, the time and cost required for functional verification can become substantial.

3. **Integration Testing:** Integrating various IP cores and subsystems onto a single chip or SoC introduces new challenges in testing the interactions between these components.

4. **Power and Clock Domains:** Handling multiple power domains and clock domains in a chip requires careful verification to ensure proper functionality and minimize power consumption.

5. **Performance Verification**: Ensuring that the chip operates at the desired performance levels under all conditions and workloads is crucial, especially for high-performance chips.

6. **Test Generation:** Generating effective and efficient test patterns to cover various fault models is a non-trivial task, especially for complex designs.

7. **Debugging:** Identifying and debugging issues in large and complex designs can be time-consuming and requires advanced debugging techniques.

Fault Models:

Fault models are representations of potential defects that can occur in a chip or design. Common fault models include:

1. **Stuck-at Faults:** These faults assume that a particular node in the circuit is stuck at either '0' or '1'.

2. **Transition Delay Faults:** These faults model timing-related issues, where a signal changes too slowly or too fast.

3. **Path Delay Faults:** These faults model delays along specific paths in the circuit.

4. **Bridge Faults:** These faults represent a short circuit between two nets or nodes.

5. **Cell-Aware Faults**: These are specific to certain types of cells and are critical for nanometer-scale technologies.

Test Coding:

Test coding involves writing test patterns to test the functionality and detect faults in a chip. Various methods and languages can be used for test coding, such as:

1. **ATPG (Automatic Test Pattern Generation):** ATPG tools automatically generate test patterns based on fault models.

2. **BIST (Built-In Self-Test):** BIST structures are embedded within the chip to facilitate self-testing.

3. **Scan Chains:** These enable efficient testing by serially scanning in test data and capturing results.

4. **Testbenches:** Testbenches are used for simulation-based verification, where test stimuli are applied to the design, and responses are analyzed.

5. **High-Level Test Languages:** Some specialized languages and tools are used for high-level test descriptions, which can be automatically converted to lower-level test patterns.

- In conclusion, designing and testing complex chips, embedded cores, and SoCs require a thorough understanding of various verification techniques, fault models, and test coding methods.
- As technology continues to advance, the challenges in test and verification continue to evolve, demanding innovative solutions and methodologies.

****

Explain about the test benches.
Introduction to test benches:

- Test benches are an essential part of digital hardware and software development, especially in the field of electronic design automation (EDA). They play a crucial role in verifying and validating the functionality of digital circuits, integrated circuits (ICs), and other electronic systems.
- A test bench serves as a virtual environment in which designers can simulate the behavior of their design, apply test stimuli, and observe the responses to ensure the correctness and functionality of the design before it is physically implemented or manufactured.

Here's an introduction to test benches:

1. **Purpose:** The primary purpose of a test bench is to thoroughly test the functionality of a digital design or electronic system before its physical implementation. It allows designers to catch and fix design errors and functional bugs in a controlled, simulated environment, reducing the risk of costly and time-consuming errors in the final product.

2. **Simulation Environment:** A test bench is created as a separate entity from the actual design being tested. It provides an environment that emulates the behavior of the design under test (DUT) and contains the necessary stimuli to drive inputs and monitor outputs.

3. **Simulation Types:** Test benches are used in various types of simulations, such as functional simulation, timing simulation, and power analysis. Each type of simulation focuses on different aspects of the design and provides valuable insights into its behavior.

4. **Test Stimuli:** In a test bench, test stimuli are applied to the inputs of the DUT to simulate different scenarios and conditions. These stimuli can be pre-defined patterns, random data, or specific corner cases to test the design's robustness.

5. **Output Monitoring:** The test bench also includes monitors that observe and record the DUT's outputs during the simulation. This allows designers to compare the expected outputs with the actual outputs to check for correctness.

6. **Debugging and Analysis:** Test benches facilitate debugging by providing detailed information about the DUT's behavior during simulation. Designers can analyze the waveform results to pinpoint errors and verify that the design meets the required specifications.

7. **Languages and Tools:** Test benches are typically written using hardware description languages (HDLs) like Verilog or VHDL. There are also higher-level verification languages, like SystemVerilog,

which provide more advanced features for test bench creation. EDA tools such as simulation tools (e.g., ModelSim, VCS) and hardware description and verification languages make the process of test bench creation more efficient and manageable.

8. **Coverage Analysis:** Test benches are instrumental in evaluating the functional coverage, code coverage, and other metrics to assess the effectiveness and completeness of the tests.

9. **Regression Testing:** As designs evolve, test benches can be used for regression testing, ensuring that any new changes or optimizations do not introduce new errors or regressions in the design.

- In summary, test benches are an integral part of the hardware and software development process, enabling designers to validate and verify digital designs through simulation.
- They are crucial for achieving high-quality, bug-free, and robust designs, leading to reduced development time and costs while ensuring the functionality and reliability of the final product.

****

1. **Explain the following: (NOV 2012)**

(i) Silicon debug principles. (MAY 2013, MAY 2014)

(ii) Fault models

Silicon debug principles

- A major challenge in silicon debugging is when the chip operates incorrectly. There are several techniques for directly accessing silicon.
- Specific signals can be brought to the top of the chip as probe points.
- These are small squares of top-level metal that connect to key points in the circuit.
- The designer has included before debug.
- The over glass cut mask should specify a hole in the passivation over the probe pads so the metal can be reliably contacted.
- The exposed squares can be probed with a Pico probe in a fixture under a microscope.
- The die can be probed electrically or optically if mechanical contact is not feasible.
- An electron beam (ebeam) probe uses a scanning electron microscope to produce a tightly focused beam of electrons to measure on-chip voltages.
- Laser Voltage Probing (LVP) involves shining a laser at a circuit and observing the reflected light. The reflections are modulated by the electric fields so switching waveforms can be deduced.
- Picosecond Imaging Circuit Analysis (PICA) captures faint light emission naturally produced by switching transistors.
- Silicon is partially transparent to infrared light, so both LVP and PICA can be performed through the substrate from the backside of a chip in a flip-chip package.
- Infrared (IR) imaging can be used to examine “hot spots” (a resistive short between power rails) in a chip.
- There are liquid crystal materials, which can be “painted on” to a die to indicate temperature problems. If the location of the fault is known, a Focused Ion Beam (FIB) can be used to cut wires or lay new conductors down.

(a) Types of failures:

- There are three types of failures, manufacturing, functional and electrical failure.
- Manufacturing failures occur when a chip has a defect or is outside of parametric specifications.
- Functional failures are logic bugs or physical design errors that cause the chip to fail under all conditions.
- Electrical failures occur when the chip is logically correct, but malfunctions under certain conditions such as voltage, temperature, or frequency.

(b) Shmoo Plot

- Shmoo plot is used to debug electrical failures in silicon.
- A shmoo plot is a plot with voltage on one axis and speed on the other.
- The test vectors are applied at each combination of voltage and clock speed, and the success of the test.
  is recorded.
- A shmoo can also plot operating speed against temperature.
- At cold temperatures, FETs are faster, have lower effective resistance, and have higher threshold voltages. Failures at low temperature could indicate coupling or charge sharing noise.
- Failures at high temperatures could indicate excessive leakage or noise problems.

2. Explain the manufacturing test principles in detail. (NOV 2011, NOV 2012, NOV 2013) Explain the chip level test techniques. (NOV 2007, MAY 2008, NOV 2021)

Manufacturing test principles

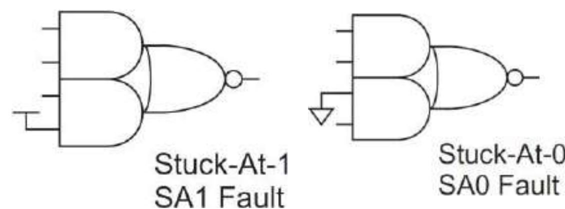
- The purpose of the manufacturing test is to screen out the defective parts before they are shipped to customers.

(a) **Fault models**

- To deal with the existence of good and bad parts, it is necessary to propose a fault model, i.e., a model of how faults occur and their impact on circuits.

(i) Stuck at faults.

- In the Stuck-At model, a faulty gate input is modeled as a stuck at zero (Stuck-At-0, S-A-0) or stuck at one (Stuck-At-l, S-A-l).
- These faults most frequently occur due to gate oxide shorts (the nMOS gate to GND or the pMOS gate to VDD) or metal-to-metal shorts.



(ii) Short circuit fault

- Two bridging or shorted faults are shown in Figure.
- The short S1 results in an S-A-0 fault at input A, while short S2 modifies the function of the gate.
- For instance, in the case of a simple NAND gate, the intermediate node between the series nMOS transistors is hidden by the schematic.
- The probable bridging fault in CMOS circuit can be grouped into three categories. There are metal polysilicon short, polysilicon n-diffusion short and polysilicon p-diffusion short.

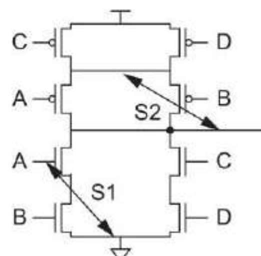


Figure: Bridging fault / Short circuit fault

(iii) Open circuit fault

- A fault can convert a combinational circuit into a sequential circuit. This is illustrated in Figure for the case of a 2-input NOR gate.
- If nMOS transistor A is stuck open, then the function displayed by the gate will be $Z = \overline{A+B} + \overline{B}Z'$ where Z' is the previous state of the gate.
- Stuck closed states can be detected by observing the static VDD current (IDD) while applying test vectors.

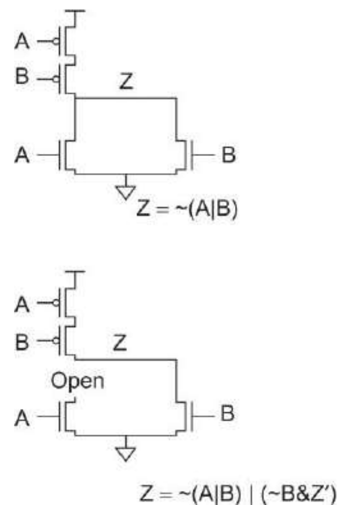


Figure: Open circuit fault

(b) Observability

- The observability of a particular circuit node is the degree to which can observe that node at the outputs of an integrated circuit (i.e., the pins).
- This metric is relevant, when want to measure the output of a gate within a larger circuit to check that it operates correctly.
- Given the limited number of nodes that can be directly observed, it is the aim of good chip designers to have easily observed gate outputs.

(c) Controllability

- The controllability of an internal circuit node within a chip is a measure of the ease of setting the node to a 1 or 0 state.
- This metric is of importance when assessing the degree of difficulty of testing a particular signal within a circuit.
- An easily controllable node would be directly settable via an input pad.

(d) Repeatability

- The repeatability of system is the ability to produce the same outputs given the same inputs.

(e) Survivability

- The survivability of a system is the ability to continue function after a fault. For example, error-correcting codes provide survivability in the event of soft errors.

(f) Fault coverage

- A measure of goodness of a test program depend the amount of fault coverage by the test program.
- The fault coverage of a set of test vectors is the percentage of the total nodes that can be detected as faulty when the vectors are applied.
- Each circuit node is taken in the sequence and held to S_a_0, and then simulation started. The chip's outputs are compared with outputs of good machine.

- If the outputs of IC are not matched with the outputs of good, and then fault is marked and the simulation is stopped.
- The same procedure is repeated to set the node to logic 1. This method is known as sequential fault grading.
- Fault coverage is defined as ratio of the number of nodes detected as faults and total number of nodes in the circuit.

(g) Automatic Test Pattern Generation (ATGP)

- If want to test the gate which is embedded in large logic circuit, use existing circuit to create a specific path from the location of gate which is going to be checked finding fault.
- This technique is known as path sensitization. This process of creating the path is known as propagation.

****

3. Explain with diagram the design strategies for testing the CMOS devices. (NOV 2008, NOV 2009)

Write briefly about different test strategies of testing digital circuits. (MAY 2009)

Explain any two approaches of DFT (Design for Testability) in brief with example. (MAY 2010, NOV 2009, MAY 2013)[Apr/May 2022]

Explain the three main approaches commonly used for design for testability (DFT). [May 2021]

Design for Testability

- The keys to designing circuits that are testable are controllability and observability.
- Controllability is the ability to set (to 1) and reset (to 0) every node internal to the circuit.
- Observability is the ability to observe, either directly or indirectly, the state of any node in the circuit.
- Good observability and controllability reduce the cost of manufacturing testing because they allow high fault coverage with relatively few test vectors.
- Three types of testing are,
  - ➢ Adhoc testing
  - ➢ Scan based approaches
  - ➢ Built -in self -test (BIST)

4. Describe the adhoc testing to design for testability in detail. (NOV 2011) [Nov/Dec 2022]

(a) Adhoc Testing

- Ad hoc test techniques are collections of ideas aimed at reducing the combinational explosion of testing.
- It is useful for small designs where scan, ATPG, and BIST are not available.
- A complete scan-based testing methodology is recommended for all digital circuits. Common techniques for ad hoc testing are
  - ➢ Partitioning large sequential circuits
  - ➢ Adding test points
  - ➢ Adding multiplexers
  - ➢ Providing for easy state reset

- A technique classified in this category is the use of the bus in a bus-oriented system for test purposes.
- Each register has been made loadable from the bus and capable of being driven onto the bus. The internal logic values that exist on a data bus are enabled onto the bus for testing purposes.
- The tester can access all the subsystems which are connected by the buses. The tester can disconnect any functional unit from the bus by setting its output into high impedance state.
- Test pattern for each subsystem can be applied separately.

- Multiplexers can be used to provide alternative signal paths during testing. In CMOS, transmission gate multiplexers provide low area and delay overhead.
- Any design should always have a method of resetting the internal state of the chip within a single cycle or at most a few cycles.
- Apart from making testing easier, this also makes simulation faster as a few cycles are required to initialize the chip.

5. Explain in detail Scan based test techniques. (NOV 2009)[Nov/Dec 2022]
Describe the scan based approaches to design for testability in detail. (NOV 2011)

(b) Scan based approaches

- The scan-design strategy for testing provides observability and controllability at each register. In this method, the registers operate in one of two modes are scan mode and normal mode.
- In normal mode, registers behave as expected.
- In scan mode, registers are connected to form a giant shift register called a scan chain spanning the whole chip.
- By applying N clock pulses in scan mode, all N bits of state in the system can be shifted out and new N bits of state can be shifted in.
- Therefore, scan mode gives easy observability and controllability of every register in the system.
- Modern scan is based on the use of scan registers, as shown in Figure.

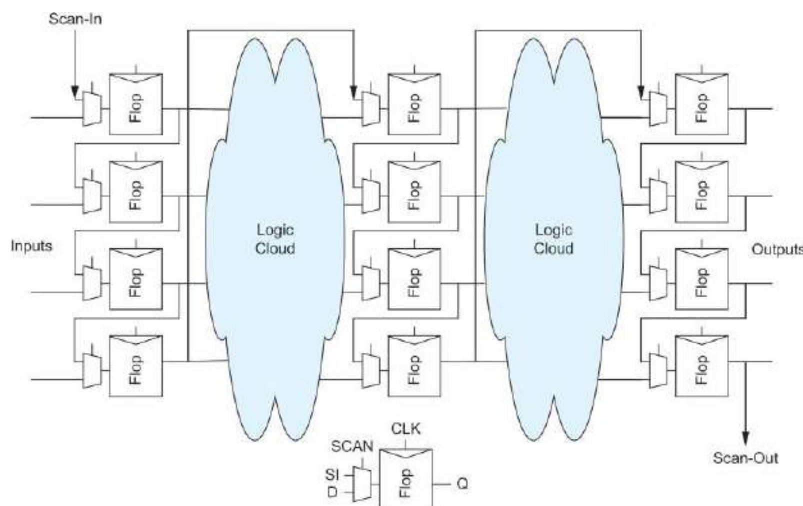


Figure: Scan based testing

- The scan register is a D flip-flop preceded by a multiplexer.
- When the SCAN signal is deasserted, the register behaves as a conventional register, storing data on the D input.
- When SCAN is asserted, the data is loaded from the SI pin, which is connected in shift register fashion to the previous register Q output in the scan chain.
- For the circuit to load the scan chain, SCAN is asserted and CLK is pulsed eight times to load the first two ranks of 4-bit registers with data.
- SCAN is deasserted and CLK is asserted for one cycle to operate the circuit normally with predefined inputs.
- SCAN is then reasserted and CLK asserted eight times to read the stored data out. At the same time, the new register contents can be shifted in for the next test.
- **The disadvantage** is the area and delay impact of the extra multiplexer in the scan register.

(i) Parallel scan approach

- The serial scan chains can become quite long, and the loading and unloading can dominate testing time.
- A simple idea is to split the chains into smaller segments.
- This can be done on a module-by-module basis or completed automatically to some specified scan length.
- This is similar to that used inside FPGAs to load and read the control RAM.
- The figure shows a two-by-two register section. Each register receives a column (column<m>) and row (row<n>) access signal along with a row data line (data<n>).

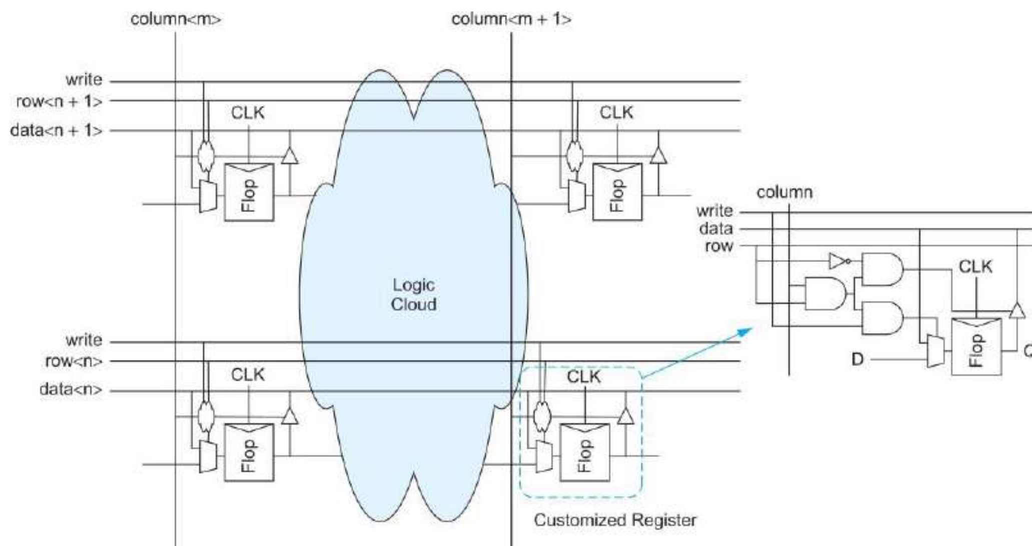


Figure: Parallel scan testing

- A global write signal (write) is connected to all registers.
- By asserting the row and column access signals in conjunction with the write signal, any register can be read or written as conventional RAM.
- Implementing the logic required at the transistor level can reduce the overhead for each register.

(ii) Scannable register design

- An ordinary flip-flop can be made scannable by adding a multiplexer on the data input, as shown in Figure (a).
- Figure (b) shows a circuit design for such a scan register using a transmission-gate multiplexer.
- The setup time increases by the delay of the extra transmission gate in series with the *D* input as compared to the ordinary static flip-flop.

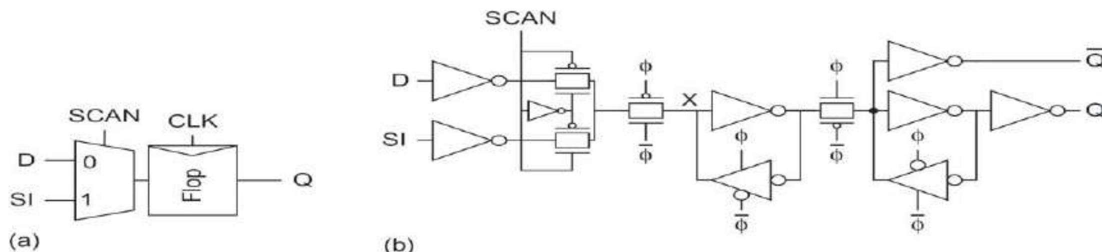


Figure: Scannable flip-flop

(iii) Level sensitive scan design (LSSD)

- During scan mode, the flip-flops are connected back-to-back. Clock skew can lead to hold time problems in the scan chain.
- These problems can be overcome by adding delay buffers on the SI input to flip-flops.
- Another approach is to use non-overlapping clocks to ensure hold times.

- The Level Sensitive Scan Design (LSSD) methodology developed at IBM uses flip-flops with two-phase non-overlapping clocks.
- During scan mode, a scan clock φs is toggled in place of φ2.
- The non-overlapping clocks also prevent hold time problems in normal operation, but increase the sequencing overhead of the flip-flop.

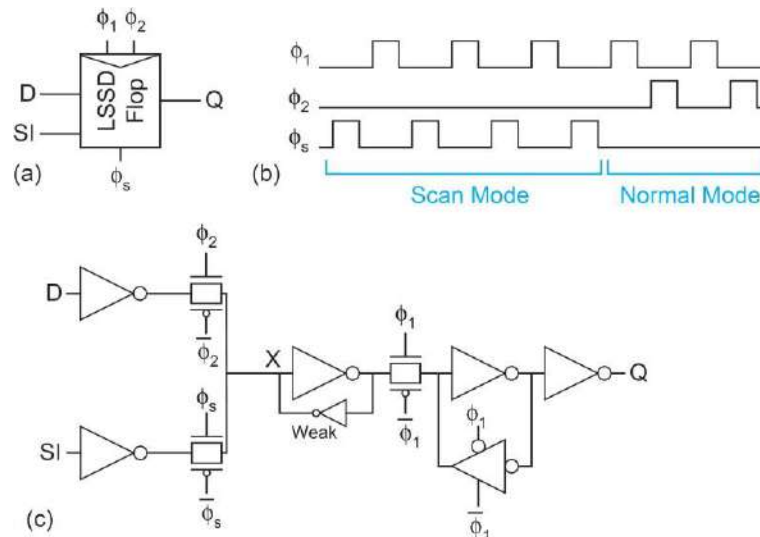


Figure: LSSD Flipflop

- Alternatively, φ1 and φ2 can be complementary clocks, but φs can be nonoverlapping to prevent races.
- Figure (c) shows a conventional design using a weak feedback inverter on the master latch that can be overpowered when either the φ2 or φs transmission gates are on.

6. Explain Built in self-test. (MAY 2008, NOV 2012, NOV 2021)

(c) Build in- Self –Test

(i) Pseudo-random sequence generator

- Self-test and built-in test techniques allow the circuit to perform operations upon themselves that prove correct operation.
- These techniques add area to the chip for the test logic, but reduce the test time required and thus can lower the overall system cost.
- One method of testing a module is to use signature analysis or cyclic redundancy checking.
- This involves using a pseudo-random sequence generator (PRSG) to produce the input signals for a section of combinational circuitry and a signature analyzer to observe the output signals.
- A PRSG of length n is constructed from a linear feedback shift register (LFSR), which in turn is made of n flip-flops connected in a serial fashion, as shown in Figure (a).
- The XOR of particular outputs are fed back to the input of the LFSR.
- An n-bit LFSR will cycle through 2^n–1 states before repeating the sequence. LFSRs are described by a characteristic polynomial indicating which bits are fed back.
- A complete feedback shift register (CFSR), shown in Figure (b), includes the zero state that may be required in some test situations.
- n-bit LFSR is converted to an n-bit CFSR by adding an n – 1 input NOR gate connected to all but the last bit.
- When in state 0…01, the next state is 0…00. When in state 0…00, the next state is 10…0. Otherwise, the sequence is the same.
- The bottom n bits of an n + 1-bit LFSR can be used to cycle through the all zeros state without the delay of the NOR gate.

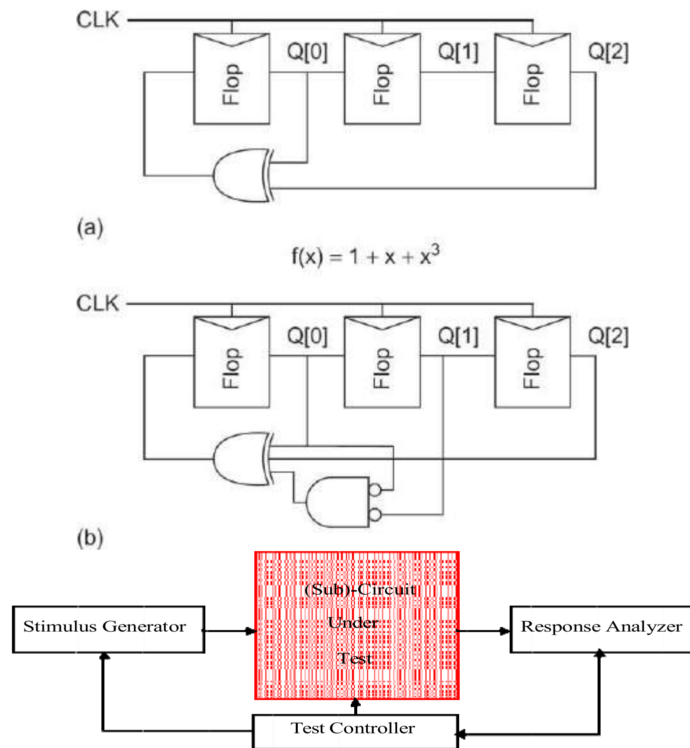


Figure: Pseudo-random sequence generator

- A signature analyzer receives successive outputs of a combinational logic block and produces a syndrome that is a function of these outputs.
- The syndrome is reset to 0, and then XORed with the output on each cycle.
- The syndrome is swizzled each cycle so that a fault in one bit is unlikely to cancel itself out.
- At the end of a test sequence, the LFSR contains the syndrome that is a function of all previous outputs.
- This can be compared with the correct syndrome to determine whether the circuit is good or bad.

(ii) Build in- Self –Test (BIST) or Built –In Logic Block Observation (BILBO)

- The combination of signature analysis and the scan technique creates a structure known as BIST—for Built-In Self-Test or BILBO—for Built-In Logic Block Observation.
- The 3-bit BIST register shown in Figure is a scannable, resettable register that also can serve as a pattern generator and signature analyzer.

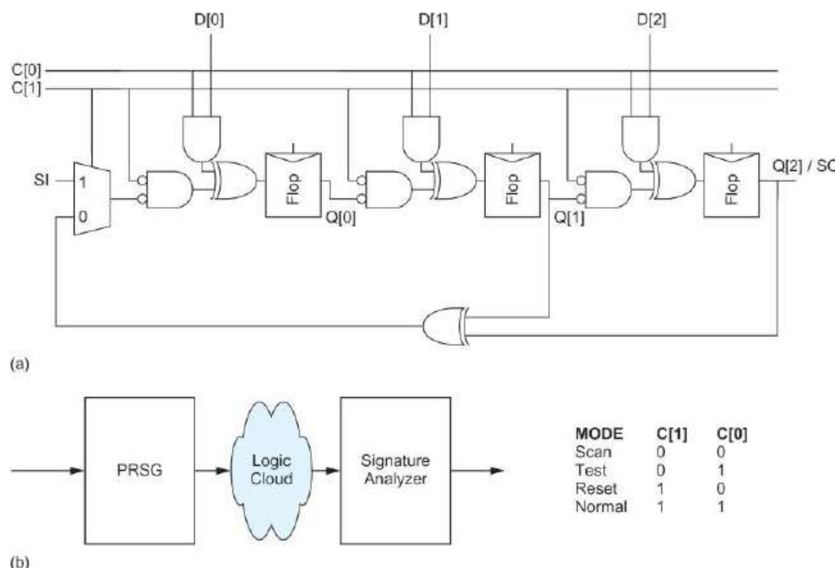


Figure: BIST (a) 3-bit register, (b) use in a system

- C[1:0] specifies the mode of operation.
- In the reset mode (10), all the flip-flops are synchronously initialized to 0.
- In normal mode (11), the flip-flops behave normally with their D input and Q output.
- In scan mode (00), the flip-flops are configured as a 3-bit shift register between SI and SO.
- In test mode (01), the register behaves as a pseudo-random sequence generator or signature analyzer.
- If all the D inputs are held low, the Q outputs loop through a pseudo-random bit sequence, which can serve as the input to the combinational logic.
- If the D inputs are taken from the combinational logic output, they are swizzled with the existing state to produce the syndrome.

7. Explain how to detect a stuck at fault with examples. (NOV 2012)
IDDQ Testing

- A method of testing for bridging faults is called IDDQ test (VDD supply current Quiescent) or supply current monitoring.
- When a CMOS logic gate is not switching, it draws no DC current (except for leakage).
- When a bridging fault occurs, then for some combination of input conditions, a measurable DC IDD will flow.
- Testing consists of applying the normal vectors, allowing the signals to settle, and then measuring IDD.
- In addition, to be effective, any circuits that draw DC power such as pseudo-nMOS gates or analog circuits have to be disabled.
- Dynamic gates can also cause problems. As current measuring is slow, the tests must be run slower than normal, which increases the test time.
- IDDQ testing can be completed externally to the chip by measuring the current drawn on the VDD line or internally using specially constructed test circuits.
- This technique gives a form of indirect massive observability at little circuit overhead.
- However, as subthreshold leakage current increases, IDDQ testing ceases to be effective because variations in subthreshold leakage exceed currents caused by the faults.

8. Explain the system level test techniques. (NOV 2007, MAY 2008, NOV 2008)
Explain in detail boundary – scan test. (MAY 2008, MAY 2013, NOV 2013, MAY 2014)

System level testing (Boundary scan testing)

- System defects occur at the board level, including open or shorted printed circuit board traces and incomplete solder joints.
- At the board level, "bed-of-nails" testers used to test boards.
- In this type of a tester, the board-under-test is lowered onto a set of test points (nails) that probe points of interest on the board.
- These can be sensed (the observable points) and driven (the controllable points) to test the complete board.
- At the chassis level, software programs are frequently used to test a complete board set.
- System designers agreeing on a unified scan-based methodology called boundary scan for testing chips at the board (and system) level.
- Boundary scan was originally developed by the Joint Test Access Group (JTAG)
- Boundary scan has become a popular standard interface for controlling BIST features as well.
- The IEEE 1149 boundary scan architecture is shown in Figure.
- All of the I/O pins of each IC on the board are connected serially in a standardized scan chain accessed through the Test Access Port (TAP)

- So that every pin can be observed and controlled remotely through the scan chain.
- At the board level, ICs obeying the standard can be connected in series to form a scan chain spanning the entire board.
- Connections between ICs are tested by scanning values into the outputs of each chip and checking that those values are received at the inputs of the chips they drive.
- Moreover, chips with internal scan chains and BIST can access those features through boundary scan to provide a unified testing framework.

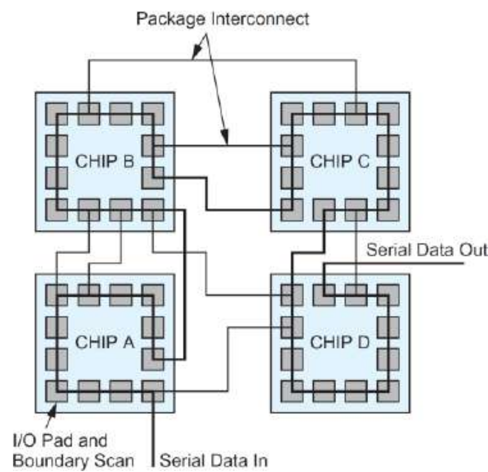


Figure: Boundary scan architecture

- The below figure shows a complete implementation of boundary scan for a chip with four inputs and four outputs.
- It consists of the TAP controller state machine and state decoder, a 3-bit instruction register with instruction decode, the bypass register, four boundary scan input pads, and four boundary scan output pads.
- The other pads comprise the test access port. The boundary scan register control signals (UpdateDR, ClockDR, ShiftDR, mode_in, and mode_out) are shown as the Control bus.

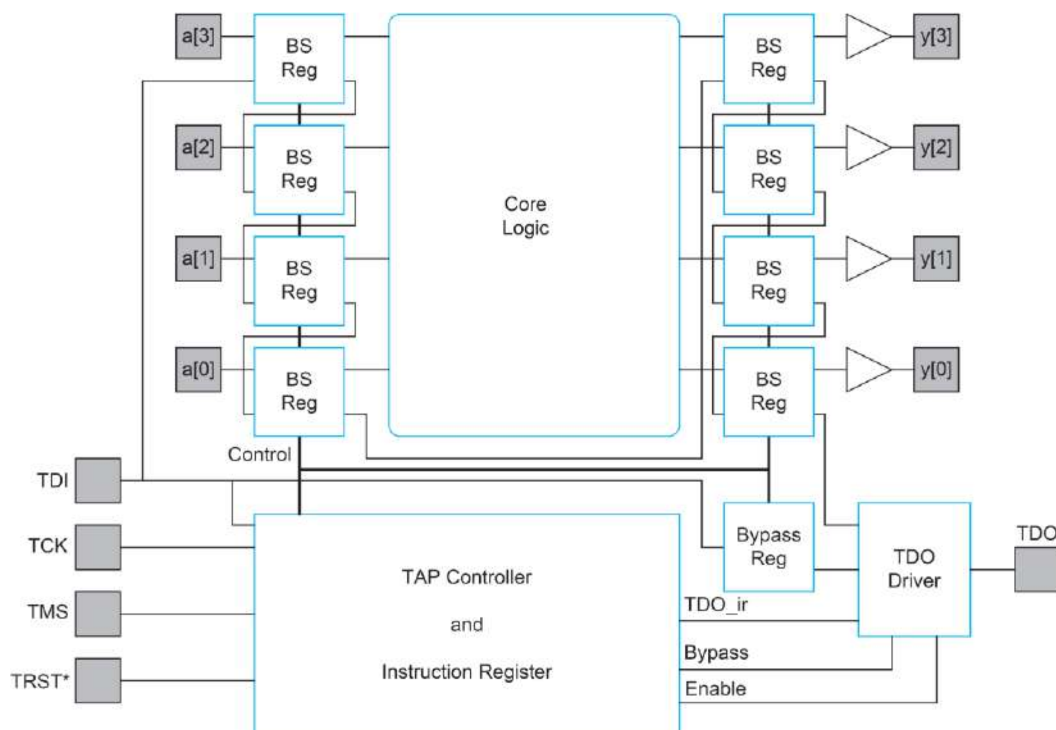


Figure: Complete Boundary scan implementation

- Boundary scan testing typically begins with the SAMPLE/PRELOAD instruction. Then, a data value is preloaded into the boundary scan registers.
- Next, the EXTEST or INTEST instruction is applied to activate the loaded value. Subsequent data values are shifted into the boundary scan registers and the results of the tests are shifted out.
- The TAP controller is initially reset. At this point, the core logic operates normally with an input pattern of 0000 and an output pattern of 0001. Then the IR is loaded with 101 (SAMPLE/PRELOAD).
- The data pattern 0111 is shifted in. The IR is loaded with 1000 (INTEST).
- This sends the 0111 pattern to the core logic, producing an output pattern of 0110.
- Finally, the data pattern 1111 is shifted in and the old output 0110 is shifted out.
- Because the INTEST is still active, the 1111 is applied to the core, producing a new output of 1100.
- It provides a uniform interface to single- and multiple-chip testing and circuit-board testing.

The Test Access Port (TAP):

- The Test Access Port has four or five single-bit connections:
  - TCK Test Clock Input Clocks tests into and out of the chip
  - TMS Test Mode Select Input Controls test operations
  - TDI Test Data In Input Test data into the chip
  - TDO Test Data Out Output Test data out of the chip; driven only when TAP controller is shifting out test data.
  - TRST* Test Reset Signal Input Optional active low signal to asynchronously reset the TAP controller if no power-up reset signal is automatically generated by the chip.
- When the chip is in normal mode, TRST* and TCK are held low and TMS is held high to disable boundary scan.
- To prevent race conditions, inputs are sampled on the rising edge of TCK and outputs toggle on the falling edge.

The Test Logic Architecture and Test Access Port:

- The basic test architecture is shown in Figure. It consists of the following:
  - The TAP interface pins
  - A set of two or more test-data registers (DR) to collect data from the chip
  - An instruction register (IR) specifying the type of test to perform
  - A TAP controller, which controls the scan of bits through the instruction and testdata registers

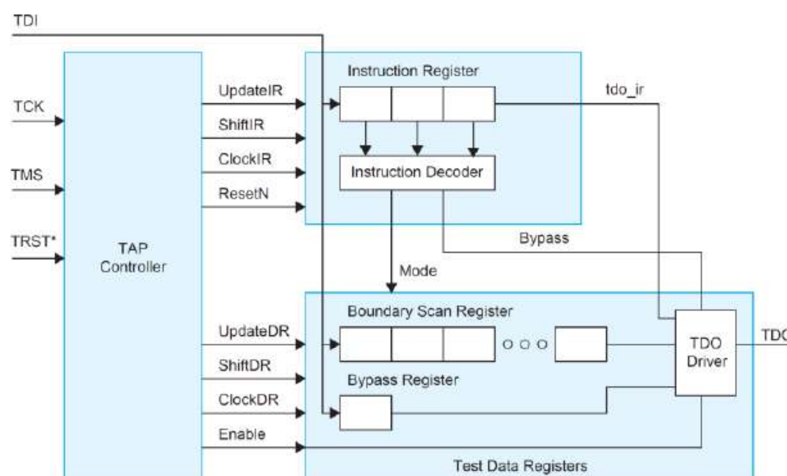


Figure: TAP Architecture

- The specification requires at least two test-data registers are the boundary scan register and the bypass register.
- The boundary scan register is associated with all the inputs and outputs on the chip so that boundary scan can observe and control the chip I/Os.
- The bypass register is a single flip-flop used to accelerate testing by avoiding shifting data into the boundary scan registers of idle chips.
- When only a single chip on the board is being tested. Internal scan chain, BIST, or configuration registers can be treated as optional additional data registers controlled by boundary scan.

The TAP Controller:

- The TAP controller is a 16-state FSM that proceeds from state to state based on the TCK and TMS signals.
- It provides signals that control the test-data registers and the instruction register. These include serial shift clocks and update clocks.
- The state transition diagram is shown in Figure. The TAP controller is initialized to Test-Logic-Reset on power-up by TRST* or an internal power-up detection circuit.
- It moves from one state to the next on the rising edge of TCK based on the value of TMS.

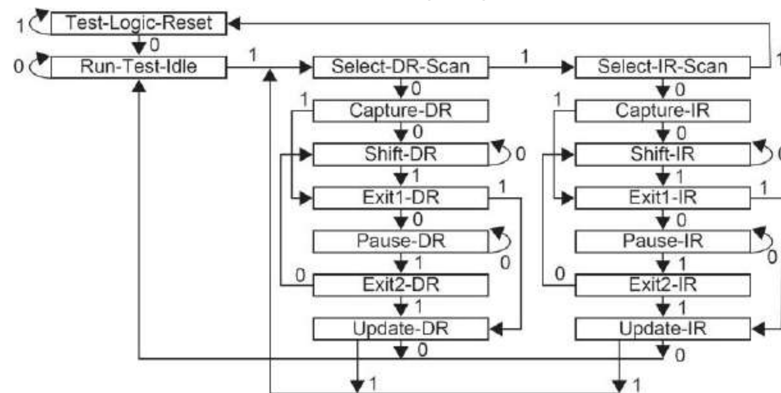


Figure: TAP controller state diagram

The Instruction Register (IR):

- The instruction register has to be at least 2 bits long. The instruction register specifies which data register will be placed in the scan chain when the DR is selected.
- It also determines from where the DR will load its value in the Capture-DR state, and whether the values will be driven to output pads or core logic.
- The following three instructions are required to be supported:
  - ➢ BYPASS—This instruction places the bypass register in the DR chain so that the path from TDI to TDO involves only a single flip-flop.
  - ➢ SAMPLE/PRELOAD—This instruction places the boundary scan registers in the DR chain. In the Capture-DR state, it copies the chip’s I/O values into the DRs.
  - ➢ EXTEST—This instruction allows for the testing of off-chip circuitry.
- In addition to these instructions, the following are also recommended:
  - ➢ INTEST— This instruction allows for single-step testing of internal circuitry via the boundary scan registers. It is similar to EXTEST.
  - ➢ RUNBIST— This instruction is used to activate internal self-testing procedures within a chip.

The Data Register:

- The test data registers are used to set the inputs of modules to be tested and collect the results of running tests.
- The simplest data register configuration consists of a boundary scan register and a bypass register (1-bit long).

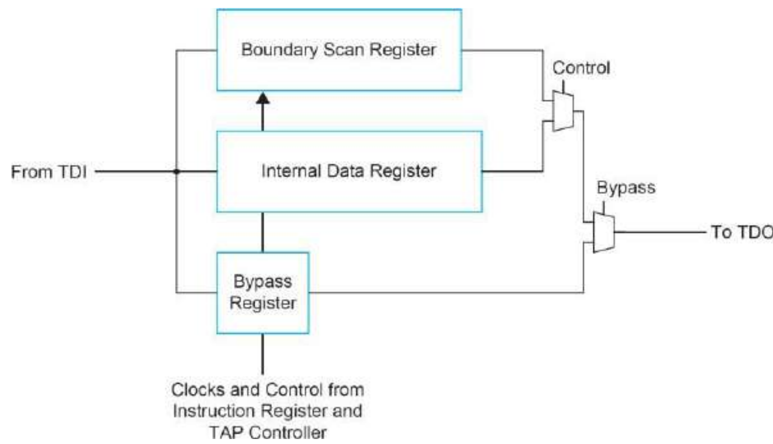


Figure: Test Data Register

The Boundary scan register:

- The boundary scan register connects to all of the I/O circuitry.
- It internally consists of a shift register for the scan chain and an additional bank of flip-flops to update the outputs in parallel.
- An extra multiplexer on the output allows the boundary scan register to override the normal path through the I/O pad so it can observe and control inputs and outputs.
- The schematic and symbol for a single bit of the boundary scan register are shown in Figure.

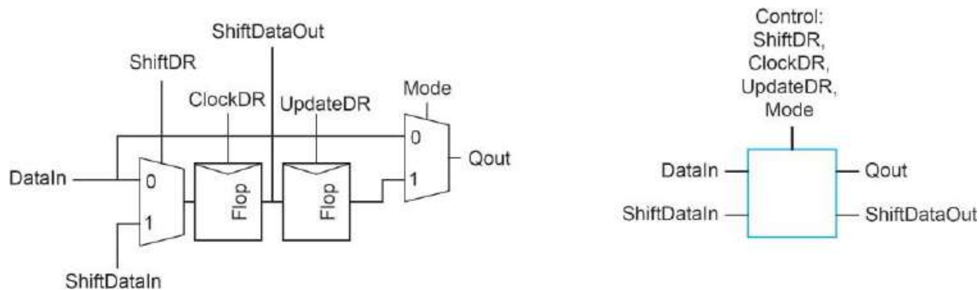


Figure: Boundary scan register bit

The Bypass Register:

- When executing the BYPASS instruction, the single-bit Bypass register is connected between TDI and TDO.
- It consists of a single flip-flop that is cleared during Capture-DR, and then scanned during Shift-DR.

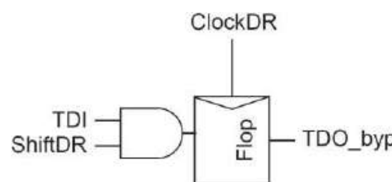


Figure: Bypass register

****

Explain the Writing test benches in Verilog HDL.

Writing test benches in Verilog HDL:

Writing test benches in Verilog HDL involves creating a module that emulates the environment in which the Design Under Test (DUT) will be tested. Test benches provide stimuli to the DUT's inputs and monitor its outputs to verify its functionality.

Here's a step-by-step guide on how to write a basic test bench in Verilog:

1. Module Declaration: Begin by declaring the test bench module, including the module name and any ports that need to be connected to the DUT.

```verilog
module tb_example;
  // Declare the DUT inputs and outputs
  // (e.g., input ports and output ports)
  // ...

  // Declare any internal signals or wires (optional)
  // ...

  // Instantiate the DUT
  // ...

  // Test bench code goes here
  // ...

endmodule
```

2. **Instantiate the Design Under Test (DUT):** In the test bench, instantiate the module representing the DUT. Connect the DUT's input and output ports to the corresponding signals or wires in the test bench.

```verilog
module tb_example;
  // Declare the DUT inputs and outputs
  // ...

  // Declare any internal signals or wires (optional)
  // ...

  // Instantiate the DUT
  DUT_module DUT_inst (
    // Connect DUT inputs to test bench signals/wires
    .input_signal_1(input_signal_1),
    .input_signal_2(input_signal_2),
    // ...

    // Connect DUT outputs to test bench signals/wires
    .output_signal_1(output_signal_1),
    .output_signal_2(output_signal_2),
    // ...
  );

  // Test bench code goes here
  // ...

endmodule
```

3. **Test Stimuli:** Within the test bench, apply test stimuli to the DUT's inputs. This can be done using initial blocks or always blocks.

```verilog
module tb_example;
  // Declare the DUT inputs and outputs
  // ...

  // Declare any internal signals or wires (optional)
  // ...

  // Instantiate the DUT
  // ...

  // Apply test stimuli
  initial begin
    // Initialize input signals with test values
    input_signal_1 = 0;
    input_signal_2 = 1;
    // ...

    // Add delays between input changes (optional)
    #5; // Delay of 5 time units

    // Change input values during simulation
    input_signal_1 = 1;
    #10; // Delay of 10 time units
    input_signal_2 = 0;

    // Add more test cases as needed
    // ...

    // End the simulation after all test cases have been executed
    $finish;
  end

  // Test bench code goes here
  // ...

endmodule
```

4. **Output Monitoring:** Use `initial` or `always` blocks to monitor and check the DUT's outputs during simulation. You can use `$display`, `$monitor`, or assertion-based methods for this purpose.

```verilog
module tb_example;
  // Declare the DUT inputs and outputs
  // ...

  // Declare any internal signals or wires (optional)
  // ...
```

```
  // Instantiate the DUT
  // ...

  // Apply test stimuli
  // ...

  // Monitor the DUT outputs
  always @(posedge clock) begin
    // Display or check DUT outputs during simulation
    $display("Output signal 1: %b, Output signal 2: %b", output_signal_1, output_signal_2);

    // Add assertions to check specific conditions (optional)
    // assert (output_signal_1 == expected_output_1) else $error("Output signal 1 mismatch");
    // assert (output_signal_2 == expected_output_2) else $error("Output signal 2 mismatch");
    // ...
  end

  // Test bench code goes here
  // ...

endmodule
```

5. **Simulating the Test Bench:** To simulate the test bench, use a Verilog simulator such as ModelSim, VCS, or Questa. The simulation will execute the test cases defined in the test bench and display the results and any assertion failures.

That's a basic outline of writing a test bench in Verilog HDL. Keep in mind that test benches can become more complex, depending on the complexity of the DUT and the desired test scenarios. Advanced test benches may include random stimulus generation, coverage analysis, and other verification methodologies to thoroughly validate the DUT's functionality.

****

Explain about Automatic Test Pattern Generation

Automatic Test Pattern Generation:

Automatic Test Pattern Generation (ATPG) is a crucial technique in semiconductor testing and verification. It involves generating a set of test patterns automatically to detect and diagnose faults or defects in a digital integrated circuit or design. The generated test patterns are applied to the design under test (DUT) during manufacturing testing or in-field testing to ensure its functionality and reliability. Here's an overview of how ATPG works:

1. **Fault Modeling:** The first step in ATPG is to create a fault model that represents the potential defects or faults in the DUT. Common fault models include stuck-at faults, transition faults, path delay faults, and bridging faults. Each fault model describes a specific type of fault that can occur in the DUT.

2. **Design Representation:** The DUT's design is represented at the gate level, typically in the form of a gate-level netlist. The netlist contains information about the gates, their connections, and the logical behavior of the design.

3. **Test Cube:** A test cube represents the inputs and outputs of the DUT that are relevant to testing a specific fault. It specifies the input patterns needed to activate the fault and the expected output responses. The ATPG tool generates test patterns based on these test cubes.

4. **Algorithmic Generation:** ATPG algorithms employ different techniques to automatically generate test patterns. Some popular algorithms include the D-algorithm, the Path Sensitization algorithm, and the Boolean Satisfiability (SAT) solver-based algorithms. These algorithms consider the fault models and the design's structure to generate effective test patterns.

5. **Test Generation Flow:**
a. The ATPG tool reads the DUT's gate-level netlist and the fault model.
b. It identifies target faults that need to be tested based on the fault model.
c. ATPG applies different algorithms to identify test cubes that activate each target fault.
d. For each test cube, the ATPG tool generates input patterns and expected output responses, creating test patterns for the target faults.
e. The generated test patterns are saved in a test vector format (e.g., STIL, WGL) or other formats suitable for test equipment.

6. **Simulation and Verification:** The generated test patterns are then applied to the DUT in a simulation environment or during manufacturing testing. The DUT's responses are compared with the expected outputs to detect any faults. If the DUT fails the test, the specific fault(s) that caused the failure can be identified for further diagnosis and debugging.

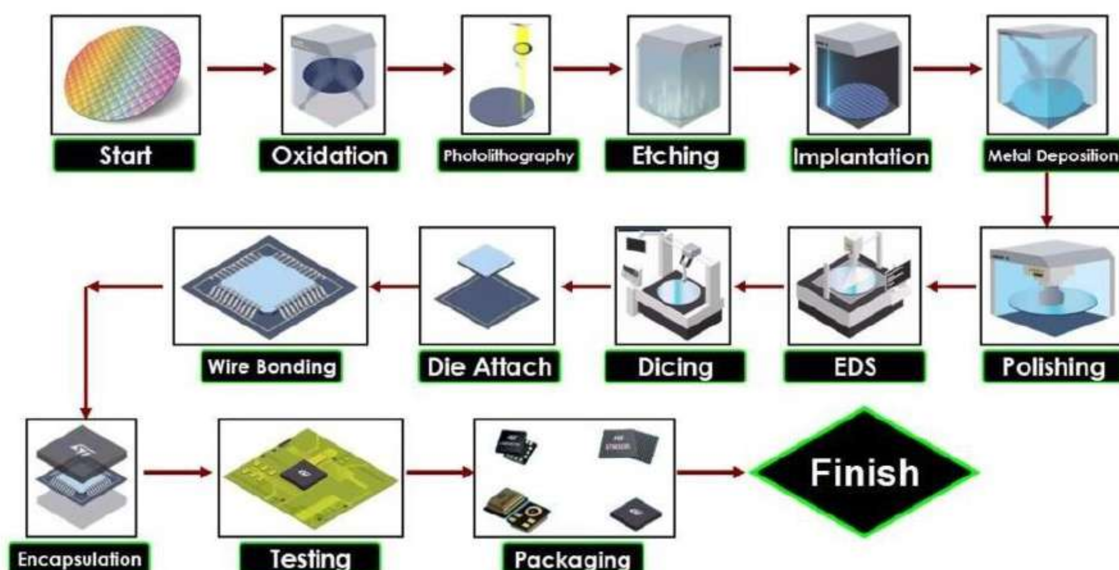
ATPG is a powerful technique that significantly improves the efficiency and coverage of semiconductor testing. It helps ensure the quality and reliability of integrated circuits, enabling the detection of manufacturing defects and design errors, thereby enhancing the overall product yield and performance.

****

Explain the Process involved in wafer to chip fabrication.
Introduction to wafer to chip fabrication process flow

- Wafer to chip fabrication, also known as semiconductor manufacturing, is the process of transforming a silicon wafer into individual semiconductor chips or integrated circuits (ICs).
- The process involves a series of steps that are meticulously executed in a cleanroom environment to ensure the highest possible quality and yield.

Here's an overview of the typical wafer to chip fabrication process flow:



1. **Wafer Ingot Growth:** The process begins with the growth of a silicon ingot. The silicon ingot is sliced into thin, circular wafers using a diamond-tipped saw. These wafers serve as the base material for manufacturing chips.

2. **Wafer Cleaning**: The wafers undergo rigorous cleaning processes to remove any contaminants or particles that might have accumulated during handling or previous steps. Cleanliness is crucial to ensure defect-free manufacturing.

3. **Oxidation:** The wafers are exposed to high-temperature oxygen or steam to create a thin layer of silicon dioxide (SiO_2) on their surface. This layer serves as an insulating material and also provides a base for subsequent processes.

4. **Photolithography:** In this step, a photoresist material is applied to the wafer's surface. Light is then shone through a photomask that contains the pattern of the desired circuit. The photoresist is exposed to this patterned light, creating a mask on the wafer. This process defines the circuit pattern for the subsequent steps.

5. **Etching:** The exposed parts of the wafer's surface are either removed or modified using chemical or physical etching processes. This step transfers the pattern from the photomask onto the wafer, defining the circuit layout.

6. **Doping:** Dopants (impurity atoms) are selectively introduced into specific areas of the wafer to modify its electrical properties. This process creates regions with either excess or deficient electrons, forming the various components of transistors (source, drain, gate, etc.).

7. **Thin Film Deposition:** Thin films of various materials, such as metal, polysilicon, or insulators, are deposited onto the wafer surface using techniques like chemical vapor deposition (CVD) or physical vapor deposition (PVD). These films serve as conductors or insulators in the circuit.

8. **Chemical Mechanical Polishing (CMP):** CMP is used to planarize the wafer's surface, making it smooth and even. This is essential for accurate layering and subsequent processing steps.

9. **Annealing:** The wafer is heated in a controlled environment to activate dopants, repair crystal damage, and improve the electrical properties of the fabricated components.

10. **Chemical Mechanical Polishing (CMP):** CMP is used to planarize the wafer's surface, making it smooth and even. This is essential for accurate layering and subsequent processing steps.

11. **Annealing:** The wafer is heated in a controlled environment to activate dopants, repair crystal damage, and improve the electrical properties of the fabricated components.

12. **Testing:** Throughout the process, various tests are conducted to ensure the quality of the chips being manufactured. These tests help identify defects and ensure that the chips meet the required specifications.

13. **Packaging:** Once all the chips on the wafer are deemed functional, they are separated and assembled into their respective packages. The packages provide protection and electrical connections to the chips, enabling them to be mounted on printed circuit boards (PCBs).

14. **Final Testing:** After packaging, the chips undergo final testing to verify their functionality and performance. Defective chips are discarded, and only fully functional chips are sent for distribution and integration into electronic devices.

- It's important to note that the above process is a simplified overview, and the actual fabrication process can be much more complex, involving multiple iterations of the steps to create multiple layers and intricate circuitry on a single chip.
- Semiconductor manufacturing is a continuously evolving field, with advancements in technology and miniaturization constantly pushing the boundaries of what is possible.

TWO MARK QUESTIONS & ANSWERS

UNIT V - ASIC DESIGN AND TESTING

1. List out the Implementation technologies in ASIC.

The implementation technologies used in ASIC are:
TTL – Transistor Transistor Logic
ECL – Emitter Coupled Logic
MOS – Metal Oxide Semiconductor (NMOS, CMOS)

2. What are the types of ASICs?

The ASICs are classified as follows:

I. Full-Custom ASICs
II. Semi-custom ASICs
   a. Standard-Cell–Based ASICs (CBIC)
   b. Gate-Array–Based ASICs (MPGA)
      - Channeled Gate Array
      - Channel less Gate Array
      - Structured Gate Array
III. Programmable ASICs
   a. Complex Programmable Logic Devices (CPLD)
   b. Field-Programmable Gate Arrays (FPGA)

3. What is meant by Full-Custom design? (May 2009)

- All mask layers are customized in a full-custom ASIC
- Generally, the designer lays out all cells by hand
- Some automatic placement and routing may be done
- Critical (timing) paths are usually laid out completely by hand.

4. What are the features of Full-Custom ASICs? (May 2016)
List the advantages of Full-Custom ASICs.

- Full-custom design offers the highest performance, minimizes its area and lowest part cost (smallest die size) for a given design.

5. What are the disadvantages of Full-Custom ASICs?

The disadvantages of full-custom design include increased design time, complexity, design expense, and highest risk.

6. Give examples of Full-Custom ASICs.

- Microprocessors (strategic silicon) were exclusively full-custom, but designers are increasingly turning to semicustom ASIC techniques in this area as well.
- Other examples of full-custom ICs or ASICs are requirements for high-voltage (automobile), analog/digital (communications), sensors and actuators, and memory (DRAM)

7. What are Semi-custom ASICs?

- Semi-custom ASIC is a cell-based ASIC (CBIC — "sea-bick")
- It has Standard cells.
- Possibly megacells, megafunctions, full-custom blocks, system-level macros (SLMs), fixed blocks, cores, or Functional Standard Blocks (FSBs)
- All mask layers are customized - transistors and interconnect.

8. What is meant by standard cell?

A standard cell is a group of transistors and interconnect structure that provides a Boolean logic function (e.g., AND, OR, XOR, Inverter) or a function (flip-flop or latch).

9. What is the standard cell-based ASIC design? (Nov 2016)

✓ In semiconductor design, standard cell methodology is a method of designing application-specific integrated circuits with digital logic features.

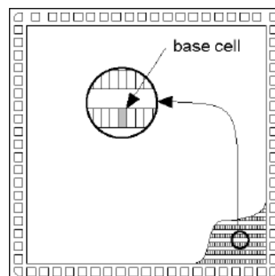
✓ A standard cell is a group of transistors and interconnect structure that provides a Boolean logic function (e.g., AND, OR, XOR, Inverter) or a function (flip-flop or latch).

10. What is meant by CBIC? (Nov 2009, April 2017)

Cell-Based Integrated Circuit consists of standard cells. CBIC based circuit is fixed and cannot be reconfigured.

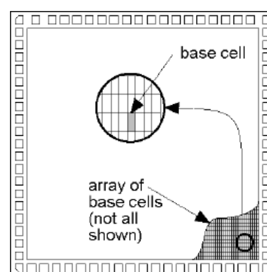
11. What is Channeled Gate Array?

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



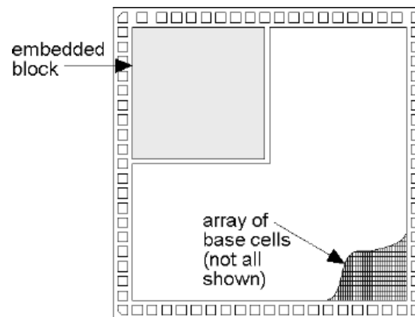
12. What is Channel less Gate Array?

- There are no predefined areas set aside for routing - routing is over the top of the gate-array devices
- Achievable logic density is higher than for channeled gate arrays
- Manufacturing lead time is between two days and two weeks



13. What is Structured Gate Array?

- Only the interconnect is customized
- Custom blocks (the same for each design) can be embedded
- These can be complete blocks such as a processor or memory array
- An array of different base cells better suited to implementing a specific function
- Manufacturing lead time is between two days and two weeks.



14. What is Ratio factor?

The *ratio factor* is the relative balance between the two partitions with respect to cell area. It is used to prevent all cells from clustering into one partition.

The ratio factor r is defined as

$$r = \frac{area(A)}{area(A) + area(B)}$$

where $area(A)$ and $area(B)$ are the total respective areas of partitions A and B

15. Write the Goals and Objectives of floor planning?

The goals of floor planning are to:

- ✓ arrange the blocks on a chip,
- ✓ decide the location of the I/O pads,
- ✓ decide the location and number of the power pads,
- ✓ decide the type of power distribution, and
- ✓ decide the location and type of clock distribution.

- The objectives of floor planning are to minimize the chip area and minimize delay.
- Measuring area is straightforward, but measuring delay is more difficult.

16. What is Channel Definition?

During the floorplanning step, we assign the areas between blocks that are to be used for interconnect. This process is known as channel definition or channel allocation.

17. What is placement?

- After completing a floorplan, we can begin placement of the logic cells within the flexible blocks.
- Placement is much more suited to automation than floor planning.
- Thus we shall need measurement techniques and algorithms.

18. What are the different placement algorithms in ASIC.

There are two classes of placement algorithms commonly used in commercial CAD tools: constructive placement and iterative placement improvement.

19. What is synthesis?

The initial synthesis contains little or no information on any interconnect loading. The output of the synthesis tool is the input to the floor planner.

20. What is Timing-driven placement?

After placement using constraints from the synthesis tool, the location of every logic cell on the chip is fixed and accurate estimates of interconnect delay can be passed back to the synthesis tool.

21. What is Routing?

Once the designer has floor planned a chip and the logic cells within the flexible blocks have been placed, it is time to make the connections by routing the chip.

22. What are the two different types of routing? (April 2018)

Two types of routing are Global routing and Hierarchical routing.

23. What is Global Routing?

A global router does not make any connections, it just plans them. We typically global route the whole chip before detail routing the whole chip (or the pieces).

24. What are the methods of Global Routing?

- Sequential routing
- Order-independent routing

25. What is hierarchical routing?

Hierarchical routing handles all nets at a particular level at once. Rather than handling all of the nets on the chip at the same time.

26. What is Reserved-layer routing?

Reserved-layer routing restricts all interconnects on each layer to flow in one direction in each routing area.

27. What is Special Routing?

The routing of nets that require special attention, clock, and power nets for example, is normally done before detailed routing of signal nets.

28. What is meant by standard cell library? (NOV 2016)

What is the role of cell libraries in ASIC design? (April 2018)

- The cell library is the key part of ASIC design.
- For a programmable ASIC the FPGA Company supplies a library of logic cells in the form of a design kit.

29. What are the models should have in a cell library?

Each cell in an ASIC cell library must contain the following:

- A physical layout
- A behavioral model
- A Verilog/VHDL model
- A detailed timing model
- A test strategy
- A circuit schematic
- A cell icon
- A wire-load model
- A routing model

30. What is meant by library cell design?

- Layout of library cells is either hand-crafted or uses some form of symbolic layout.
- Symbolic layout is usually performed in one of two ways: using either interactive graphics or a text layout language.
- Shapes are represented by simple lines or rectangles, known as sticks or logs , in symbolic layout.

31. List out the basic elements of the FPGA structure.
State the three important blocks in FPGA architecture. (April 2019)
State the building blocks of FPGA. (Nov 2019)

The basic elements of FPGA structure are:

1. Logic blocks
- Based on memories *(LUT – Lookup Table)* Xilinx
- Based on multiplexers *(Multiplexers)* Actel
- Based on PAL/PLA *(PAL - Programmable Array Logic, PLA – Programmable Logic Array)* Altera
- Transistor Pairs

2. Interconnection Resources
- Symmetrical FPGA-s
- Row-based FPGA-s
- *Sea-of-gates* type of FPGA-s
- Hierarchical FPGA-s (*CPLD)*

3. Input-output cells (*I/O Cell)*

32. Name the elements in a Configuration Logic Block. (April 2017)

Configuration Logic blocks:

- Memories *(LUT – Lookup Table, Flip-flop)* - Xilinx
- Multiplexers -Actel
- PAL/PLA *(PAL - Programmable Array Logic, PLA – Programmable Logic Array)* - Altera
- Transistor Pairs

33. Write the features of Xilinx FPGA. (April 2008)

The features of Xilinx FPGA are:

✓ High-performance

-5 ns pin-to-pin logic delays on all pins

- ✓ Large density range - 36 to 288 macrocells with 800 to 6,400 usable gates
- ✓ 5 V in-system programmable
- ✓ Endurance of 10,000 program/erase cycles

34. Give the functions of Input Output Block.

- ✓ The I/O Block (IOB) interfaces between the internal logic and the device user I/O pins.
- ✓ Each IOB includes an input buffer, output driver, output enable selection multiplexer, and user programmable ground control.

35. Write about various ways of routing procedure. (Nov 2017)

- ✓ Hierarchical Routing Architecture
- ✓ Island-Style Routing Architecture
- ✓ Xilinx Routing Architecture
- ✓ Altera Routing Architecture
- ✓ Actel Routing Architecture

36. What is VLSI and ULSI? (Nov 2017)

Very large-scale Integration:

Very large-scale Integration (VLSI) with gates counting upto lakhs.

Ultra large-scale Integration:

Ultra large-scale integration (ULSI) is the process of integrating millions of transistors on a single silicon semiconductor microchip.

37. What is feed through cells? State their uses. (May 2016)

A Feed through is a connection that needs to cross over a row of standard cells.

Feed through cells needed for vertical routing for routing using the same metal layer(s) as within cells.

38. What is a programmable logic device?

A programmable logic device (PLD) is an electronic component used to build reconfigurable digital circuits.

Unlike a logic gate, which has a fixed function, a PLD has an undefined function at the time of manufacture.

39. What are the types of programmable logic device (Programmable ASIC)?

- PLA
- PAL
- FPGA

40. What is meant by ASIC?

Application Specific Integrated Circuit is an Integrated Circuit (IC) designed to perform a specific function for a specific application.

41. What is an antifuse? State its merits and demerits. (Nov 2016)

- ✓ Antifuse is nothing high resistance (>100 MΩ) is changed into low resistance(200-500Ω) by applying programming voltage.

- ✓ Merit: Antifuses separate interconnect wires on the FPGA chip and the programmer blows an antifuse to make a permanent connection.
- ✓ Demerit: Once an antifuse is programmed, the process can't be reversed.

42. What is PLA?

Programmable logic arrays (PLAs) is a type of fixed architecture logic devices with *programmable AND gates followed by programmable OR array.*

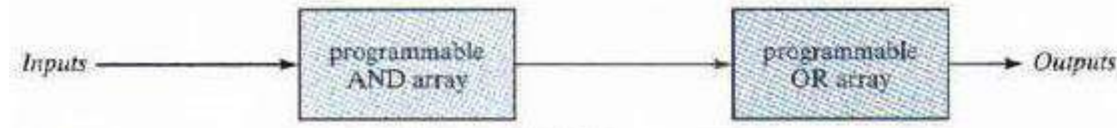


Figure: Programmable logic array

43. What is PAL?

The PAL is a programmable logic device with a *fixed OR array and a programmable AND* array.

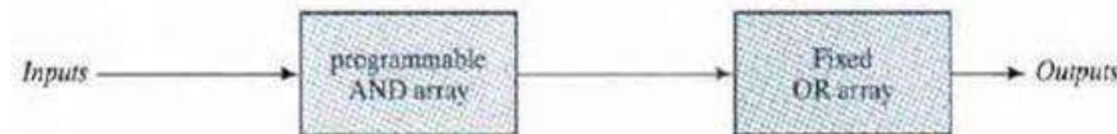


Figure: Programmable Array Logic

44. What are the types FPGA programming technologies?
What are the different types of programming structure available in PAL? (Nov 2008)

There are three types of programming technology.

- Fusible link programming (Anti fuse)
- SRAM Programming
- EPROM and EEPROM programming

45. What is meant by Reprogrammable Gate array (FPGA)?
What is the significance of field programmable gate arrays? [May 2021][Apr/May 2022]

A field programmable gate array (FPGA) is a VLSI circuit that can be programmed at the user's location.

A typical FPGA consists of an array of millions of logic blocks, surrounded by programmable input, and output blocks and connected together via programmable interconnections.

46. Differentiate between Full custom and Cell based ASICs. (April 2008)

Full Custom ASICs:

- All mask layers are customized in a full-custom ASIC
- Generally, the designer lays out all cells by hand
- Some automatic placement and routing may be done

Cell based ASICs:

- It has Standard cells, Possibly mega cells, mega functions, full-custom blocks, system-level macros (SLMs), fixed blocks, cores or Functional Standard Blocks
- All mask layers are customized - transistors and interconnect.

47. What is needed for testing? [Nov/Dec-2013] [Apr/May-2008]

Physical defects are likely in manufacturing.

– Missing connections (opens)

– Bridged connections (shorts)

– Imperfect doping, processing steps

– Packaging

Need to weed out bad die before assembly. Need to test during operation – Electromagnetic interference, mechanical stress, electromigration, alpha particles.

48. What are different stages of testing on a chip? [Nov/Dec-2012]

Different stages of testing are wafer level, circuit level, chip level, board level, field level, logic level.

49. What is meant by tester in VLSI testing?

A tester is a device that can apply a sequence of stimuli to a chip or system under test and record the results.

50. Distinguish testers and test fixtures. [Nov/Dec-2012]

Testers: equipment used for testing.

Test fixtures: components through which the equipment's are connected with testing block.

51. State all the test vectors to test 3 input NAND gate. [May/June-2009]

Three inputs test vector are 000, 001, 010, 011, 100, 101, 110 and 111.

52. What are the test fixtures required to test a chips? [Nov/Dec-2011]

To test a chip, various types of test fixtures may be required. These are

- Probe card: It is used to test at the wafer level or unpackaged die level with a chip tester.
- Load board: It is used to test a packaged part with a chip tester.
- Printed circuit board (PCB): It is used for bench-level testing (with or without a tester).
- PCB with the chip in situ: It is used for demonstrating the system application for which the chip is used.

53. What is meant by test program?

The tester requires a test program. This program is written in a high-level language that supports a library of primitives for a particular tester.

54. What is handler?

Used for feeding Ics to a test fixture which is attached to a tester? It has mechanical positioning equipment. Status is indicated at the top, to check if the handler is functioning or not. It can handle 2 to 4 chips at a time.

55. How is testing classified?

List out the basic types of CMOS testing (or) What are the different types of CMOS testing? [May/June-2013] [Nov/DEC-2008]

The basic types of CMOS testing are functionality testing and Manufacturing testing.

56. What is functionality test?

State the objective of functionality test. [Nov/Dec-2011]

Functionality test is to check whether logic block works with correct logic. It leads to imperfection of logic function. It is done before fabrication process.

57. What is manufacturability test?

Manufacturing test is to check is there any defects occurred in circuit after fabrication process. It leads to nodes to float, shorted to power or ground. It is done after fabrication process.

58. Distinguish functionality test and manufacturing test. [Nov/Dec-2007]

Functionality: Functionality test is to check whether logic block works with correct logic. It leads to imperfection of logic function. It is done before the fabrication process.

Manufacturing: Manufacturing test is to check is there any defects occurred in circuit after fabrication process. It leads to nodes to float, shorted to power or ground. It is done after fabrication process.

59. What is the principle behind logic verification? [Nov/Dec-2013]

Verifying the logical principles of the circuit by the following ways:
Test benches & Harness, Regression testing, Version control and Bug tracking**.**

60. Define test benches and harness.

A verification test bench or harness is a piece of HDL code.
In the simplest test bench, input vectors are applied to the module under test and at each cycle, the outputs are taken and compared with the output of another model.

61. What is shmooing?

The ability of varying the voltage and timing on a per-pin basis with a tester allows a process known as shmooing.

62. What is shmooing plots?

In this plot, voltage is taken in x-axis and speed is taken in y-axis. The test vectors are applied and the output is recorded.

63. What is VCD?

Vector Change Description - It is used to compact stimulation results.

64. What is version control?

It is used in regression testing. It is orderly management of different design iterations.

65. What is regression testing?

Stimulation performed automatically. Verify that no functionality has changed in a model. Regression conducted after design activities every day.

66. What is meant by Bug tracking?

Bug tracking is nothing but allows the management of wide variety of Bugs(error checks).

67. What is meant by silicon debugging principles and name some probes used for it?

Silicon debugging principles are those technique which can directly access the silicon for testing. LVP-Laser Voltage Probing, PICA-Picasecond Imaging Circuit Analysis.

68. What is hotspot and how it is examined?

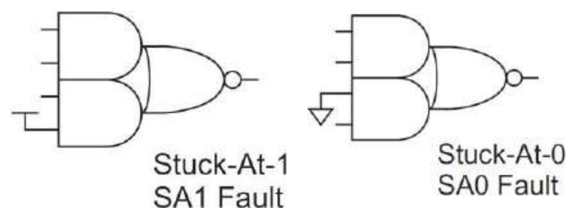
Hotspots are examined using infrared imaging techniques.

69. How the temperature is examined in a chip?
Liquid Crystal material can be painted to a die to identify the temperature related problems.

70. What is FIB?
Focused Ion Beam (FIB): If the fault locations are identified, then FIB can be used to cut wires (or) lay new conductors down.

71. What is electrical failure?
Electrical failur occurs if the chip is logically correct but malfunctions occur due to temperature.

72. What is struck at fault? [Nov/DEC-2008]
If any input line is struck at logic '0' or logic '1' permanently called as struck at fault. If it struck at logic '1' then called as struck-at-1 or s-a-1 and if it struck at logic '0' then called as struck-at-0 or s-a-0.



73. What is struck open fault? [Nov/DEC-2008]
Due to defects while manufacturing leads to permanent disconnect of drain to source terminal called as struck open fault.

74. What is bridging fault or short circuit fault?
Due to defects while manufacturing leads to shorting of inputs between themselves or shorting of inputs to outputs(feedback) occurs called as bridging fault.

75. List any two faults that occur during manufacturing. [Nov/DEC-2008]
Struck at faults and struck open fault.

76. Define fault coverage?
Fault coverage is defined as ratio of the number of nodes detected as faults and total number of nodes in the circuit.

$$\text{Fault coverage} = \frac{\text{Total no. of nodes in which fault identified}}{\text{Total no. of nodes in circuit}}$$

77. How is the bridging fault categorized?
Bridging faults are categorized as

Input bridging
Feedback bridging
Non-feedback bridging

78. What is observability? [Nov/Dec 2022]
The observability of a particular circuit node is the degree to which we can observe that node at the outputs of an integrated circuit.

79. What is controllability? [Nov/Dec 2022]

The measure of ease of forcing/setting a node to 0 or 1 by driving input pins of the chip is called controllability.

80. What is meant by ATPG?

Automatic Test Pattern Generation (ATPG) – Block generates input patterns automatically by itself for testing its own logic block and stores the output pattern and compare it with defined pattern for error identification.

81. What are the 3 approaches in design for testability? (or) List out design required for testing in CMOS chip design. [Apr/May-2008]

Three approaches in design for testability are

Adhoc testing
Scan based testing
BIST- Built In Self Test

82. What do you mean by DFT? [Nov/Dec-2009]

Good observability and controllability reduce the cost of manufacturing testing because they allow high fault coverage with relatively few test vectors.
Three main approaches to what is commonly called *Design for Testability*
(DFT). These may be categorized as follows:

Adhoc testing
Scan based testing
BIST- Built In Self Test

83. What is serial scan & parallel scan?

In serial scan based approaches, logic is connected to form a giant shift register called as a scan chain spanning the whole chip.
In parallel scan based approaches, logic is split the chain into smaller segments. This can be done on a module –by-module basis or completely automatically to some specified scan length.

84. List the common techniques for ad hoc testing. (NOV 2021)

1. Buddy testing 2. Pair testing 3. Monkey testing

85. What is signature analyzer?

Signature analyzer is a block which observes the output signal.

86. What is the drawback of scan based approaches?

Drawbacks of scan based approaches are area and delay impact of the extra multiplexer in the scan register.

87. What is the aim of Adhoc test techniques?[Nov/Dec-2007] [Apr/may-2010]

Ad hoc test techniques are collections of ideas aimed at reducing the combinational explosion of testing. They are only useful for small designs where scan, ATPG, and BIST are not available. A complete scan-based testing methodology is recommended for all digital circuits.

88. What is the MUX test technique?

Multiplexers can be used to provide alternative signal paths during testing. In CMOS, transmission gate multiplexers provide low area and delay overhead.

89. Write a note on partition and MUX technique

What are common techniques used in adhoc testing? [Apr/May-2011]

(i) Partitioning large sequential circuits

(ii) Adding test points

(iii) Adding multiplexers (iv) Storing output datas

90. What are scannable elements for circuit design?

Mutliplexer, Flipflop and CMOS transmission gates are scannable elements in circuit design.

91. What is the necessary for non-overlapping clocks?

The non-overlapping clocks in LSSD prevents hold time problems in normal operation, but increase the sequencing overhead of the flipflop.

92. What is syndrome?

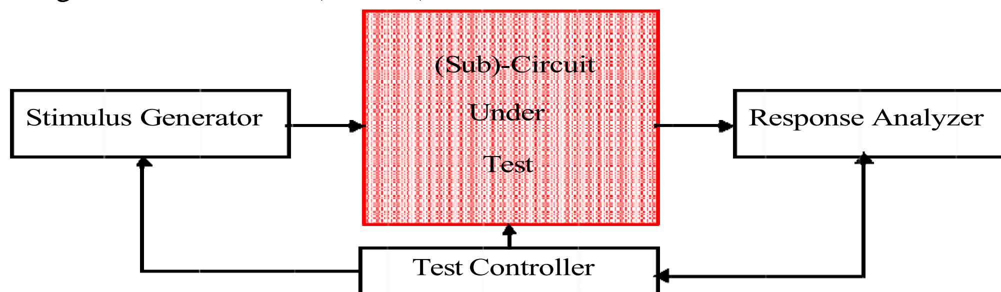
Syndrome is output pattern generated for the applied input pattern in testing.

93. What is LSSD?

Level Sensitive Scan Design : In this method, flipflops with two phase non-overlapping clocks are used in testing circuit.

94. What is BIST or BILBO?

The Combination of scan technique with PRSG & signature analysis creates a structure called as Built-in Logic Block Observer (BILBO).



95. What is needed for IDDQ testing? [Apr/May-2011] [Apr/May-2010]

List out design guidelines for IDDQ testing.

A method of testing for bridging faults is called IDDQ test (*VDD* supply current Quiescent) or supply current monitoring. This relies on the fact that when a CMOS logic gate is not switching, it draws no DC current (except for leakage).

When a bridging fault occurs, then for some combination of input conditions, a measurable DC *IDD* will flow. Testing consists of applying the normal vectors, allowing the signals to settle, and then measuring *IDD*.

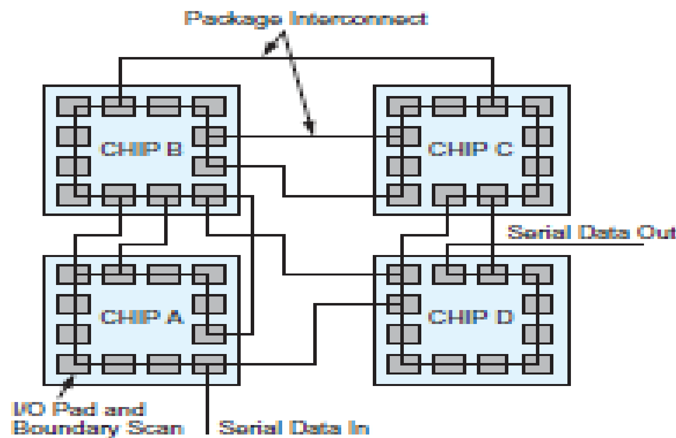
96. What are the limitations of IDDQ testing?(NOV 2021)[Apr/May 2022]

Compared to scan chain testing, IDDQ testing is time consuming and more expensive.

97. What is meant by system level (Boundary scan) testing?

System designers agree on a unified scan-based methodology called boundary scan for testing chips at the board (and system) level.

98. Draw the boundary scan input logic diagram. [Nov/Dec-2009]



99. What is TAP?

All the input and output pins of each IC on the board are connected serially in a standardized scan chain accessed through the TAP (Test Access Port), so that every pins can be observed and controlled remotely through the scan chain.

100. What are the signals used in Tap Access port (TAP)?

The Test Access Port has four or five single-bit connections:

TCK Test Clock Input Clocks tests into and out of the chip

TMS Test Mode Select Input Controls test operations

TDI Test Data In Input Test data into the chip

TDO Test Data Out Output Test data out of the chip; driven only when TAP controller is shifting out test data.

TRST* Test Reset Signal Input Optional active low signal to asynchronously reset the TAP controller if no power-up reset signal is automatically generated by the chip.

101. Draw the Tap Access port (TAP) architecture.

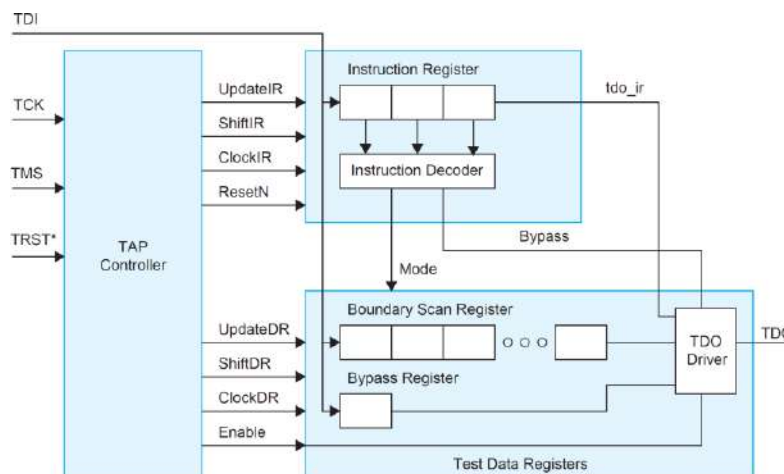


Figure: TAP Architecture

102. What is meant by TAP controller?

The TAP controller is a 16-state FSM that proceeds from state to state based on the TCK and TMS signals.

103. What is bypass register?

Single bit register connected between Test Data In (TDI) and Tata Data Out (TDO). It is cleared during capture- DR, and then scanned during shift Data Read (DR).

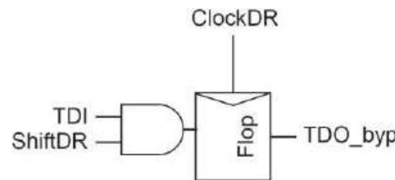


Figure: Bypass register

104. What is instruction register?

Instruction register is a 2 bit register which specifies which data register will be placed in the scan chain when DR – data register is selected.

105. What is Data Register (DR)?

The test data registers are used to set the inputs of modules to be tested and collect the results of running tests.

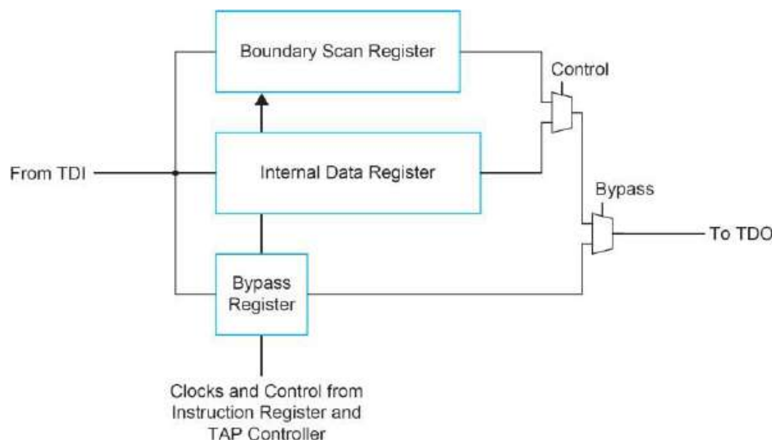


Figure: Test Data Register

106. What is boundary scan register?

The boundary scan register connects to all of the I/O circuitry. It internally consists of a shift register for the scan chain and an additional bank of flip-flops to update the outputs in parallel.

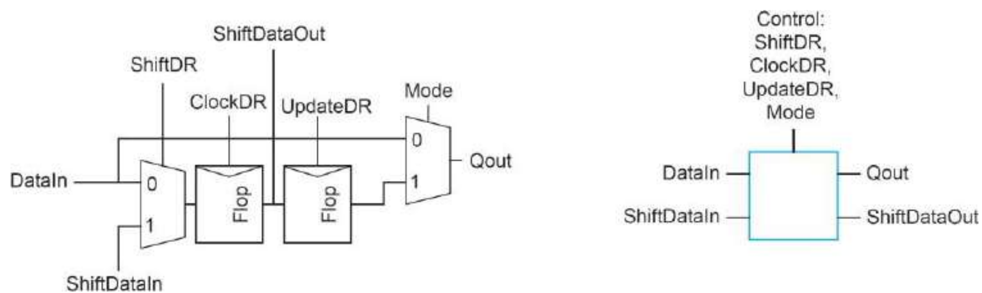


Figure: Boundary scan register bit

107. What are logic verification principles? [May 2013, Nov 2013]

Verifying the logical principles of the circuit by the following ways test benches & Harness, regression testing, version control and bug tracking.

106. Compare full custom and semi-custom design. (Nov 2019)

Full-Custom design

- All mask layers are customized in a full-custom ASIC
- Generally, the designer lays out all cells by hand
- Some automatic placement and routing may be done
- Critical (timing) paths are usually laid out completely by hand.

Semi-custom ASICs

- Semi-custom ASIC is a cell-based ASIC (CBIC —"sea-bick")
- It has Standard cells
- Possibly megacells , megafunctions , full-custom blocks , system-level macros (SLMs), fixed blocks, cores, or Functional Standard Blocks (FSBs)
- All mask layers are customized - transistors and interconnect

107. Identify the ways to optimize the manufacturability, to increase yield. [May 2021]

Examine workflow

Invest in employee training.

Modernize your business process.

Invest in smart machining equipment.

Develop realistic expectations.

Stay organized.

Create a culture of collaboration.

Invest in preventative maintenance.

108. What are the advantages and disadvantages of BIST? [Nov/Dec 2022]

Advantages of implementing BIST include:

1) Lower cost of test, since the need for external electrical testing using an ATE will be reduced, if not eliminated
2) Better fault coverage, since special test structures can be incorporated onto the chips
3) Shorter test times if the BIST can be designed to test more structures in parallel
4) Easier customer support
5) Capability to perform tests outside the production electrical testing environment. The last advantage mentioned can actually allow the consumers themselves to test the chips prior to mounting or even after these are in the application boards.

Disadvantages of implementing BIST include:

1) Additional silicon area and fab processing requirements for the BIST circuit
2) Reduced access times
3) Additional pin (and possibly bigger package size) requirements, since the BIST circuitry need a way to interface with the outside world to be effective
4) Possible issues with the correctness of BIST results, since the on-chip testing hardware itself can fail.

109. What is configurable logic block meant?

The programmable logic blocks of FPGAs are called Configurable Logic Blocks (CLBs). CLBs contain LUT, FF, logic gates and Multiplexer to perform logic functions.

UNIT V
ASIC DESIGN AND TESTING
Question bank

1. **Explain about wafer to chip fabrication process flow.**
2. Describe about Microchip design process.
3. Explain the issues in test and verification of complex chips.
4. Write short notes on embedded cores and SOCs.
5. **Explain the Fault models and Test coding.**
6. **Explain the ASIC Design Flow.**
7. **Explain the writing test benches in Verilog HDL**
8. **Explain the Automatic test pattern generation.**
9. **Explain the Design for testability.** (Scan design: Test interface and boundary scan)

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

EC3501- Wireless Communication

Semester - 05

Notes

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

Vision

To excel in providing value based education in the field of Electronics and Communication Engineering, keeping in pace with the latest technical developments through commendable research, to raise the intellectual competence to match global standards and to make significant contributions to the society upholding the ethical standards.

Mission

- ✓ To deliver Quality Technical Education, with an equal emphasis on theoretical and practical aspects.
- ✓ To provide state of the art infrastructure for the students and faculty to upgrade their skills and knowledge.
- ✓ To create an open and conducive environment for faculty and students to carry out research and excel in their field of specialization.
- ✓ To focus especially on innovation and development of technologies that is sustainable and inclusive, and thus benefits all sections of the society.
- ✓ To establish a strong Industry Academic Collaboration for teaching and research, that could foster entrepreneurship and innovation in knowledge exchange.
- ✓ To produce quality Engineers who uphold and advance the integrity, honour and dignity of the engineering.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

1. To provide the students with a strong foundation in the required sciences in order to pursue studies in Electronics and Communication Engineering.
2. To gain adequate knowledge to become good professional in electronic and communication engineering associated industries, higher education and research.
3. To develop attitude in lifelong learning, applying and adapting new ideas and technologies as their field evolves.
4. To prepare students to critically analyze existing literature in an area of specialization and ethically develop innovative and research oriented methodologies to solve the problems identified.
5. To inculcate in the students a professional and ethical attitude and an ability to visualize the engineering issues in a broader social context.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO1: Design, develop and analyze electronic systems through application of relevant electronics, mathematics and engineering principles.

PSO2: Design, develop and analyze communication systems through application of fundamentals from communication principles, signal processing, and RF System Design & Electromagnetics.

PSO3: Adapt to emerging electronics and communication technologies and develop innovative solutions for existing and newer problems.

EC 3501 – WIRELESS COMMUNICATION

UNIT I

THE CELLULAR CONCEPT-SYSTEM DESIGN FUNDAMENTALS

Introduction - Frequency Reuse - Channel Assignment Strategies - Handoff Strategies: Prioritizing Handoffs, Practical Handoff Considerations. Interference and System Capacity: Co-Channel Interference and System Capacity - Channel Planning for Wireless Systems, Adjacent Channel Interference, Power Control for Reducing Interference, Trunking and Grade of Service. Improving Coverage and Capacity in Cellular Systems: Cell Splitting, Sectoring.

PART A

1. What do you mean by mobile-assisted handoff? (May/Jun 2019)
2. What is channel assignment? What are the types?
3. What are the techniques used to expand the capacity of cellular system?

   (Or)

   How is the frequency reuse distance measured in cellular system? (Nov/Dec 2018)
4. Define frequency reuse. (Nov/Dec 2017) (May/Jun 2018) (May/Jun 2023)

   (Or)
5. How is frequency reuse distance measured in cellular system? (Nov/Dec 2018)
6. What is Multiple Access?
7. Define Grade of service. (Nov/ Dec 2015) (Nov/ Dec 2016)
8. In a cellular network, among a handoff call and a new call, which one is given priority? Why? (May/ Jun 2017) (May/ Jun 2023)
9. Define cell splitting.
10. What is cell sectoring?
11. Why is cellular concept used for mobile telephony? (May/ Jun 2017)

    (Or)
12. List the feature of Cellular concept. (Nov/Dec 2018)
13. Define co-channel reuse ratio. (Nov/ Dec 2015)
14. Write is microcell zone concept?
15. Define co-channel interference.

PART B

1. Describe Channel assignment strategies and Hand-Off strategies.

   (10) (May/ Jun 2017) (13) (May/Jun 2018)
2. Explain Handoff and interference systems.

(Or)

Explain in detail in handoff scenario at cell boundary.

(10) (Nov/ Dec 2016) (7) (May/Jun 2019)

(Or)

How is Hand-off in a cellular system implemented? Explain the different types of Hand-offs

(6) (Nov/Dec 2018)

(Or)

Explain in detail the handoff strategies. (10) (Nov/Dec 2019)

3. Explain cell sectoring and cell splitting in detail.
4. Explain “repeaters for range extension” and “microcell zone” concept.
5. A cellular service provider decides to use a digital TDMA scheme which can tolerate a signal to interference ratio of 15 dB in the worst case. Find the optimal value of N for
   (1) Omni directional antennas (3)
   (2) 120° sectoring (3)
   (3) 60° sectoring (3)
   (4) Should sectoring be used? If so, which case (60°or 120°) should be used? (Assume a path loss exponent of n=4 and consider trucking efficiency)

   (3) (Nov/ Dec 2015)

   If a signal-to-interference ratio of 15 dB is required for satisfactory forward channel performance of a cellular system, what is the frequency reuse factor and cluster size that should be used for maximum capacity if the loss exponent is i) n=4 ii) n=3 ?

   (4) (Nov/ Dec 2015)

6. Write short note on (i) Trunking (ii) Grade of service of cell system. (16) (Nov/Dec 2017)

   (Or)

   Write a brief note on (i) Trunking (ii) Cell splitting (8) (May/Jun 2019)

   (Or)

   Write a note on grade of service. (3) (Nov/Dec 2019)

7. If a total of 33 MHz of bandwidth is allocated to a particular FDD cellular telephone system which uses two 25 kHz simplex channels to provide full duplex voice and control channels, compute the number of channels available per cell if a system uses (i) four - cell reuse (2) seven- cell reuse and (3) twelve - cell reuse. If 1 MHz of the allocated spectrums is dedicated to control channels, determine the equitable distribution of control channels and voice channels in each cell of each of three systems. (6) (May/ Jun 2017)
8. A hexagonal cell within a four-cell system has a radius of 1.387 km. A total of 60 channels are used within in the entire system. If the load per user is 0.029 Erlangs and λ =1 call/hour, Compute the following for an Erlang C system that has a 5% probability of a delayed call:
   i) How many users per square kilometer will this system support? (5) (Nov/ Dec 2015)
   ii) What is the probability that a delayed call will have to wait for more than 10 sec.?

   (5) (Nov/ Dec 2015)

   iii) What is the probability that a call will be delayed for more than 10 sec?

   (6) (Nov/ Dec 2015)

   [Data: From Erlang C chart, for 5% probability of delay with C=15, traffic intensity = 9.0 Erlangs]

9. Explain about co-channel interference and adjacent channel interference. Describe the techniques to avoid interference. (16) (Nov/ Dec 2016)

(Or)

Discuss the impact of interference in a cellular system and system capacity. (7) (Nov/Dec 2018)

(Or)

Derive the interference experienced by a cell edge user in the cellular architecture. (13) (May/Jun 2023)

10. A certain city has an area of 1300 square miles and is covered by a cellular system using a 7- cell reuse pattern. Each cell has a radius of 4 miles and the city is allocated 40 MHz of spectrum with a full duplex channel bandwidth of 60 kHz. Assume that a GOS is 2% and traffic intensity per cell A=84 Erlangs/cell. If the offered per user is 0.03 Erlangs, compute.
    (1) The number of cells in the service area,
    (2) The number of channels per cell,
    (3) The maximum carried traffic
    (4) The total number of users that can be served for 2 % GOS
    (5) The number of mobiles per channel, and
    (6) The theoretical maximum number of users that could be served at one time by the system. (13) (Nov/Dec 2019)
11. Explain any two ways of improving the covering and capacity in cellular systems. (10) (Nov/Dec 2019)
12. Consider a time invariant frequency selective block fading channel consisting of 3 subchannels of B = 1 MHz. The frequency response associated with each channel is $H_1 = 1$, $H_2 = 2$, $H_3 = 3$. The transmit power constraint is p = 10 mw and noise power spectral density is N_0 = 10-9 W/Hz. Find the Shannon capacity of the channel and optimal power allocation that achieves this capacity. (13) (May/Jun 2018)
13. A hexagonal cell within a four cell system has a radius of 1.387km. A total of 60 channel are used within the entire system. If the load per user is 0.029 Erlangs, and λ=1 call/hour, compute the following for an Erlang C system that has a 5% probability of a delayed call and determine the following, (i) How many users per square kilometre will this system support? (ii) What is the probability that a delayed call will have to wait for more than 10sec? (iii) What is the probability that a call will be delayed for more than 10sec?
14. If a total of 33 MHz of bandwidth is allocated to a particular FDD cellular telephone system which uses two 25 kHz simplex channels to provide full duplex voice and control channels, compute the number of channels available per cell if a system uses
    (1) 4- cell reuse
    (2) 7- cell reuse (6) (May/Jun 2019)
15. A spectrum of 30MHZ is allocated to a wireless FDD cellular system which uses two 25KHZ simplex channels to provide full duplex voice and control channels, compute the number of channels available per cell if a system uses (a)four-cell reuse (b)seven-cell reuse ,and (c)12-cell reuse. If 1MHZ of the allocated spectrum is dedicated to control channels, determine an equitable distribution of control channels and voice channels in each cell for each of three systems.

16. An urban area has a population of two million residents. Three competing trunked mobile networks (systems A,B and C) provide cellular service in this area. System A has 394 cells with 19 channels each, system B has 98 cells with 57 channels each, and system C has 49 cells, each with 100 channels. Find the number of users that can be supported at 2% blocking if each user averages two calls per hour at an average call duration of three minutes. Assuming that all three trunked systems are operated at maximum capacity, compute the percentage market penetration of each cellular provider.
17. (i) Discuss about Grade of service of cell system. (ii) Estimate that how many users can be supported for 0.5% blocking probability for the following number of trunked channels in a blocked calls cleared system? (a) 1, (b) 5, (c) 10, (d) 20, (e) 100. Assume each user generates 0.1 Erlangs of traffic.

UNIT-II MOBILE RADIO PROPAGATION

Large Scale Path Loss: Introduction to Radio Wave Propagation - Free Space Propagation Model – Three Basic Propagation Mechanism: Reflection – Brewster Angle- Diffraction Scattering. Small Scale Fading and Multipath: Small Scale Multipath Propagation, Factors Influencing Small-Scale Fading, Doppler Shift, Coherence Bandwidth, Doppler Spread and Coherence Time. Types of Small- Scale Fading: Fading Effects due to Multipath Time Delay Spread, Fading Effects due to Doppler Spread.

PART A

1. What are propagation models and its types?
2. What are the two types of path loss model?
3. What is the necessity of Link budget?
4. Explain basic link budget equation.
5. What is free space propagation model? (May/Jun 2023)
6. Explain friis free space equation.
7. What is meant by multipath propagation? (Nov/Dec 2017)
8. Explain path loss equation of a free space propagation model.
10. Define EIRP
11. What is path loss model?
12. What is flat fading? (Nov/Dec 2017)

    (Or)

    Define fading. (May/Jun 2023)
13. Explain received power of a Two Ray ground reflection model.
14. Explain path loss equation of a Two Ray ground reflection model.
15. What are the factors influencing small scale fading?
16. What flat fading? Define small scale fading.
17. What is frequency selective fading?
18. What is fast fading? (Nov/Dec 2018)
19. Define coherence time. (Nov/Dec 2018)

20. Define Doppler shift & Doppler spread.
21. What are the two factors that contribute to the rapid fluctuations of the signal amplitude? (May/Jun 2019)
22. Define Small Differentiate constructive interference and destructive interference.
23. Define Power delay profile.
24. Define mean excess delay.
25. Define RMS delay spread.
26. Define maximum excess delay.
27. What is the major advantage of wireless communication? (May/ Jun 2017)
28. Find the far-field distance for an antenna with maximum dimension of 2 m and operating frequency of 1 GHz. (Nov/Dec 2015)
29. Define Coherence time and coherence bandwidth. (Nov/Dec 2015)

    (Or)

    Define Coherence time. In what way does this parameter decide the behaviour of wireless channel? (May/ Jun 2017)
30. Give the equation for average large-scale path loss between the transmitter and receiver as a function of distance. (Nov/ Dec 2016)
31. What is frequency selective fading? (Nov/ Dec 2016)
32. Compare fast and slow fading. (May/Jun 2018)

    (Or)

    Differentiate small from large scale fading. (May/Jun 2019)
33. Give the difference between frequency flat and frequency selective fading. (May/Jun 2018)
34. What is shadow fading? Why it is called so? (Nov/Dec 2019)
35. What factors does diffraction depend on at high frequencies? (Nov/Dec 2019)

PART B

1. a.) What you mean by path loss model? Explain in detail about log - distance path loss model. (8) (Nov/Dec 2017)

   b). What is the need for link calculation? Explain with suitable example. (8) (Nov/Dec 2017)

   (Or)

   Explain the various path loss models for large scale fading. (7) (Nov/Dec 2018)
2. Explain following

   a). Explain friis free space equation. b). Explain d^{-4} law? c). Explain path loss equation for a free space propagation model.

   (Or)

   Describe briefly about free space propagation model. (3) (May/Jun 2018)

   (Or)

   Describe the free space propagation model and derive the loss in signal strength. (7) (May/Jun 2019)

   (Or)

   Explain the free space path loss and derive the gain expression. (8) (Nov/Dec 2019)
3. a). Explain small scale fading and, what are the factors affecting the small-scale fading.

   (Or)

What are the factors influencing small scale fading? (6) (Nov/ Dec 2016)

(Or)

Describe small scale fading and derive expressions for parameters of mobile multipath channels. (13) (Nov/Dec 2018)

(Or)

Write down the three small scale fading effects and also name the techniques that are used to mitigate the effects of small scale fading. (5) (Nov/Dec 2019)

b) A mobile is located at 5 kms away from base station and uses a vertical $\lambda/4$ monopole antenna with a gain of 2.55 dB to receive cellular radio signals. The E- field at 1km from transmitter is measured to 10^{-3}V/m the carrier frequency is 900 MHz. Find the length and effective aperture of the effective the receiving antenna.

4. Explain fading due to multipath delay spread.

(Or)

Explain fading due to Doppler spread and coherence time.

(Or)

Explain fading effects due to multipath time delay spread and fading effects due to Doppler Spread. (10) (Nov/ Dec 2016)

(Or)

Compare and contrast fast fading and slow fading. "In practice fast fading only occurs for very low data rate (communications)". Why? (5) (May/ Jun 2017)

(Or)

Distinguish fast fading and slow fading in wireless channel and explain in detail. (16) (Nov/Dec 2017)

(Or)

Given that the coherence bandwidth is approximated by equation Bc = 1/5σr, show that flat fading channel occurs when TS>=10σr (6) (May/Jun 2018)

(Or)

What is frequency selective fading? Explain. (6) (Nov/Dec 2018)

(Or)

Discuss the flat fading channel characteristics with relevant diagrams. (8) (May/Jun 2019)

Discuss the classification of small scale fading with respect to Doppler spread. (5) (May/Jun 2019)

(Or)

Examine the effectiveness of flat fading and frequency selective fading. (15) (Nov/Dec 2019)

(Or)

Discuss your understanding on types of fading in the perspective of Doppler spread and coherence time. (13) (May/Jun 2023)

5. Derive the Impulse response model of a multipath channel and also obtain the relationship between bandwidth and Received power. (16) (Nov/ Dec 2015)
6. If a transmitter produces 50 W of power, which is applied to a unity gain antenna with a 900 MHz Carrier frequency, find the receive power in dBm at a free space distance of 100 m from the antenna.

What is received power at a distance of 10 km? Assume unity gain for the receiver antenna. (5) (May/ Jun 2017)

7. Describe in detail, the parameters of mobile multipath channel with their significance. (6) (May/ Jun 2017)
8. Determine the proper spatial sampling interval required to make small - scale propagation measurements which assume that consecutive samples are highly correlated in time. How many samples will be required over 10 m travel distance if fc = 1900 MHz and v = 50 m/s. How long would it take to make these measurements, assuming they could be made in real time from a moving vehicle? What is the Doppler spread BD for the channel? (5) (May/ Jun 2017)
9. Consider a transmitter which radiates a sinusoidal carrier frequency of 1850 MHz for a vehicle moving 60 mph, computer they received carrier frequency if the mobile is moving directly toward the transmitter. (4) (May/Jun 2018)
10. If the transmit power is 1 W and carrier frequency is 2.4 GHz and the receiver is ata distance of 1 mile from the transmitter. Assume that the transmitter and receiver antenna gains are 1.6
    (1) What is the received power in dBm in the free space of a signal?
    (2) What is the path loss in dB?
    (3) What is the transmission delay in ns? (6) (May/Jun 2019)
11. Determine the proper spatial sampling interval required to make small scale propagation measurements which assume that consecutive samples are highly correlated in time. How many samples will be required over 10 m travel distance if fc = 1900 MHz and v = 50 in/s. How long would it take to make these measurements, assuming they could be made in real time from a moving vehicle? What is the Doppler spread BD for the channel? (8) (Nov/Dec 2019)
12. Determine the proper spatial sampling interval required to make small scale propagation measurements which assume that consecutive samples are highly correlated in time. How many samples will be required over 10 m travel distance if fc = 1900 MHz and v = 50 in/s. How long would it take to make these measurements, assuming they could be made in real time from a moving vehicle? What is the Doppler spread BD for the channel? (8) (Nov/Dec 2019)
13. Compare and contrast wired and wireless communication. (5) (May/Jun 2019) (Nov/Dec 2019)
14. Assume a transmitter produces 50W of power, express the transmit power in units (a) dBm and (b) dBw. If 50W is applied to a unity gain antenna with a 900MHz carrier frequency. Find the received power in dBm at a free space distance of 100m from the antenna. What is received power (10 Km)? assume unity gain receiver antenna.
15. Explain in brief about the three propagation mechanisms which have impact on propagation in a mobile environment. (8) (Nov/Dec-2012)
16. Find the far field distance for an antenna with maximum dimension of 2m and operating frequency 1GHz. (6)

(Or)

17. Find the far field distance for an antenna with maximum dimension of 1m and operating frequency 900 MHz. (6) (Nov/Dec-2012)

UNIT- III MODULATION TECHNIQUES AND EQUALIZATION AND DIVERSITY

Digital Modulation – An Overview: Factors that Influence the Choice of Digital Modulation, Linear Modulation Techniques: Minimum Shift Keying (MSK), Gaussian Minimum Shift Keying (GMSK), Spread Spectrum Modulation Techniques: Pseudo- Noise (PN) Sequences, Direct Sequence Spread Spectrum (DS-SS)- Modulation Performance in Fading and Multipath Channels- Equalization, Diversity and Channel Coding: Introduction-Fundamentals of Equalization- Diversity Techniques: Practical Space Diversity Considerations, Polarization Diversity, Frequency Diversity, Time Diversity.

PART A

1. Write the advantage of MSK over GMSK.
   (Or)
   State any two advantages of MSK. (May/ Jun 2017)
2. Draw the block diagram of GMSK transmitter using direct FM generator.
3. Find the 3 dB bandwidth of a Gaussian low pass filter used to produce 0.25 GMSK with a channel data rate of Rb = 270 kbps. What is the 90% power bandwidth in the RF channel? Specify the Gaussian filter parameter.
4. Define Gaussian filter co-efficient in terms of bandwidth.
   (Or)
   Give the function of Gaussian filter in GMSK. (Nov/ Dec 2016)
   (Or)
   State the advantages of GMSK. (Nov/Dec 2018)
5. What is the principle of cyclic prefix in frequency selective channels?
   (Or)
   What is cyclic prefix? (Nov/ Dec 2016) (May/Jun 2018)
6. List the features of offset QPSK. (May/Jun 2019)
7. Draw the structure of a generic optimum receiver.
8. What are the differences between zero-forcing and mean squared error equalizer?(May/Jun 2019)
9. Write the applications of MFSK and OFDM.
10. What is MSK? Why it is named so? (Nov/Dec 2019)
11. Mention the merits and demerits of nonlinear modulation.
12. Find the 3-dB bandwidth for a Gaussian low pass filter used to produce 0.25 GMSK with a channel data rate of Rb = 300 kbps. (Nov/ Dec 2015)

13. A 900 MHz carrier signal is frequency modulated using a 100 kHz sinusoidal modulating waveform. The peak deviation of the FM signal is 500 kHz. If this FM signal is received by a super heterodyne receiver having an IF frequency of 5 MHz, determine the IF bandwidth necessary to pass the signal. (Nov/ Dec 2015)
14. What you mean by transversal filter in linear equalizer?

    (Or)
15. What are linear equalizers and nonlinear equalizers? (Nov/ Dec 2016)

    (Or)

    Distinguish linear and non-linear equalization. (Nov/Dec 2018)
16. Write two advantage of lattice equalizer.
17. Draw the structure of maximum like hood sequence estimator (MLSE) in nonlinear equalizer.
18. What is the various factor involved in the adaptive equalizer algorithm?
19. Name the three techniques used to improve the received signal quality. (May/Jun 2019)
20. What you mean feed forward filter? Write the advantage of feed forward filter in nonlinear equation?
21. Why is an adaptive equalizer required? (May/ Jun 2017)
22. In digital cellular equalizer, if the carrier frequency is 900 MHz and maximum Doppler shift is 66.67 Hz, calculate the maximum mobile velocity for the given Doppler shift.
23. If digital signal processor chip can perform one million multiplications per second, determine the time required between each iteration.
24. Draw the predicative decision feedback nonlinear equalizer.
25. Define spatial diversity. (Nov/Dec 2017)
26. Define STCM. (Nov/Dec 2017)
27. Write the expression of folded frequency response of channel in zero force algorithms?
28. Define correlation coefficient of diversity.
29. What is use of automatic repeat request (ARQ) for temporal filter?
30. Define angle diversity in antenna.
31. What is diversity? Why is it employed? (May/ Jun 2017)
32. What is the function of selection and combining diversity?
33. Write the drawback of Bit-Error-rate-Driven diversity.
34. Write the basic principle of combining diversity.
35. What is Macro diversity? (Nov/ Dec 2016) (Nov/Dec 2018) (May/Jun 2019)
36. Draw the structure of a linear transversal equalizer. (Nov/ Dec 2015)
37. If a digital signal processing chip can perform one million multiplications per second, determine the time required between each iteration for the following adaptive equalizer algorithms. (Nov/ Dec 2015)

    (a) Square root RLS DFE

    (b) Gradient lattice DFE
38. Design a three tap zero forcing linear equalizer. (May/Jun 2018)
39. Distinguish between diversity gain versus array gain. (May/Jun 2018)
40. What is the use of equalization technique? Name the types. (Nov/Dec 2019)
41. Differentiate micro from macro diversity. (Nov/Dec 2019)
42. What is the significance of spread spectrum?

43. What is the use of special code in spread spectrum?
44. Where spread spectrum is used?
45. What are the two types of spread spectrum?
46. What is the Meaning of the word jamming and anti-jam?
47. What is jamming margin?
48. What is meant by PN sequence and what are the properties of PN sequence?
49. Define chip duration and chip rate.
50. What is the shape of auto-correlation function of PN sequence?

PART B

1. Derive for the mean square error for linear equalizer during training adaptive equalizer.
2. Describe in detail about i) Linear equalizers ii) Non - linear equalizers. (16) (Nov/Dec 2017)

   (Or)

   What is Equalization? Why is the equalization in wireless system required to be Adaptive? (6) (Nov/Dec 2018)

   (Or)

   With neat diagrams, explain and analyze linear equalization procedure. (15) (Nov/Dec 2018)

   (Or)

   Write down the purpose of algorithms for adaptive equalization and also state on what factors the performance of these algorithms depend on. (5) (Nov/Dec 2019)

   (Or)

   Explain the necessary mathematical model, the working mechanism of adaptive equalizer. (13) (May/Jun 2023)
3. Derive the expression for least mean square algorithm.

   (Or)

   Explain in detail the various factors to determine the algorithm for adaptive equalizer. Also derive the Least Mean Square Algorithm for adaptive equalizer. (16) (Nov/ Dec 2016)

   (Or)

   With valid statements, analytically prove that the adaptive equalizers exhibit superior performance over the conventional equalizers. (15) (Nov/Dec 2017)

   (Or)

   Draw and explain a simplified communication system using an adaptive equalizer at the receiver. (13) (May/Jun 2019)

   (Or)

   Draw the chart showing the classification of equalizers. (8) (May/Jun 2019)
4. Describe the role played by Equalization and diversity as Multipath mitigation techniques. Compare and contrast these two techniques. (10) (May/ Jun 2017)

5. Analyze various diversity techniques used in wireless communication. (16) (Nov/Dec 2017)

   (Or)

   Describe any two diversity combining techniques stating their respective merits. (7) (Nov/Dec 2018)

   (Or)

Write a brief note on categories of space diversity reception methods. (6) (May/Jun 2019)

(Or)

Derive an any one of the diversity combining technique and compare Macro and Micro diversity. (13) (May/Jun 2023)

6. Write detail operation of micro diversity in terms of Special, Temporal, Frequency, Angle polarization.
7. Explain the following
   a). Error probability in flat-fading channels
   b). Symbol error rate in frequency selective fading channel
8. Consider the design of the US Digital Cellular equalizer, where f = 900 MHz and the mobile velocity v = 80 km/hr, determine the maximum Doppler shift, the coherence time of the channel and the maximum number of symbols that could be transmitted without updating the equalizer assuming that the symbol rate is 24.3 k symbols/sec. (6) (May/ Jun 2017)
   i. Determine the mean of the Rayleigh fading signal as a referenced to the threshold. (5) (Nov/ Dec 2015)
   ii. Find the likelihood that a two-branch selection diversity receiver will be 6 dB below the mean SNR threshold. (2) (Nov/ Dec 2015)
   iii. Find the likelihood that a three-branch selection diversity receiver will be 6 dB below the mean SNR threshold. (2) (Nov/ Dec 2015)
   iv. Find the likelihood that a four-branch selection diversity receiver will be 6 dB below the mean SNR threshold. (2) (Nov/ Dec 2015)
   v. Based on your answer above, is there a law of diminishing returns when diversity is used? (5) (Nov/ Dec 2015)
12. Assume four branch diversity is used, where each branch receivers an independent Rayleigh fading signal. If the average SNR is 20 dB, determine the probability that the SNR will drop below 10 dB. Compare this with the case of a single receiver without diversity. (4) (May/ Jun 2017)
13. Consider uncoded spatial multiplexing over a MIMO channel with MR≥MT, Show that the ML, MMSE and ZF receivers perform equally well if the channel is orthogonal, i.e., HHH = ηIMR .Where η is a constant. What is the per-stream SNR? (16) (May/Jun 2018)
14. What is zero forcing equalizer algorithm? Explain. (6) (Nov/Dec 2018)
15. Determine the proper spatial sampling interval required to make small scale propagation measurements which assume that consecutive samples are highly correlated in time. How many samples will be required over 10 m travel distance if fc = 1900 MHz and v = 50 m/s. How long would it take to make these measurements, assuming they could be made in real time for a moving vehicle? What is the Doppler spread for the channel? (7) (May/Jun 2019)
16. Explain GMSK transmitter and receiver with signal spacing diagram and give an expression for spectral efficiency.

    (Or)

    Explain in detail about Gaussian Minimum Shift Keying (GMSK) Transmission and Reception with necessary block diagrams. (10) (Nov/ Dec 2015)

    (Or)

    Discuss the working mechanism of GMSK. Compare its performance with MSK.(13) (May/Jun 2023)
17. Describe with neat diagram, the modulation technique of QPSK. (8) (Nov/Dec 2017)
18. What is MSK? Explain with transmitter and receiver diagram. Explain the various types of demodulation of MSK.

(Or)

Derive the expression for MSK signal as a special type of continuous phase FSK signal. (16) (Nov/ Dec 2015)

(Or)

What is MSK? Also derive the expression of MSK signal as a special type of FSK signal and explain its power spectral density. (16) (Nov/ Dec 2016)

(Or)

Why is constant envelope modulation schemes such as MSK and GMSK used in a wireless communication system? Compare and contrast these two modulation techniques. (8) (May/ Jun 2017)

(Or)

Examine the principle of MSK modulation and derive the expression for power spectral density. (16) (Nov/Dec 2017)

(Or)

Explain the MSK system and its importance in a wireless communication system. (7) (Nov/Dec 2018)

19. What are PN sequences? What are the properties of PN sequences? (4) (Nov/Dec-2018)

(Or)

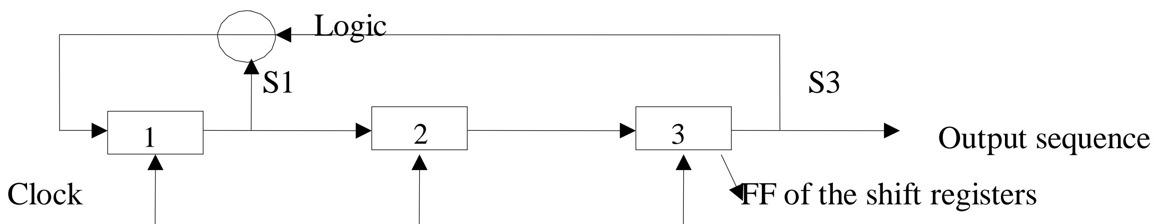
Explain the generation of PN sequences and its properties. (13) (Nov/Dec-2019)

(Or)

Explain the principle of operation of direct sequence spread spectrum with its n performance parameters. How pseudo noise is generated? (13) (May/Jun-2016)

20. For a linear feedback shift register with three stages (m=3), evaluate the maximum length PN sequence for feedback taps = (3, 1). Draw the schematic arrangement and verify all the properties of PN sequence in generated output. Sketch the sequence, its autocorrelation function if chip rate happens to be 10MHz.

21. (i) Explain the PN sequence generated by the feedback register shown in figure.



(ii) With neat block diagram and explain the direct sequence spread spectrum technique.

UNIT- IV MULTIPLE ACCESS TECHNIQUES

Introduction: Introduction to Multiple Access- Frequency Division Multiple Access (FDMA) – Time Division Multiple Access (TDMA)- Spread Spectrum Multiple Access-Code Division Multiple Access (CDMA)- Space Division Multiple Access (SDMA)- Capacity of Cellular Systems: Capacity of Cellular CDMA, Capacity of CDMA with Multiple Cells.

PART A

1. What is FDMA?
2. What are the basic units of a Cellular system?
3. How FDMA handles near-far problem? (May/Jun 2019)

4. What is MSC?
5. What do you mean by forward and reverse channel? (Nov/Dec 2017)
6. State the advantages of CDMA over FDMA. (Nov/ Dec 2016)
7. What is cyclic prefix? (Nov/ Dec 2016)
8. Differentiate between FDMA, TDMA, and CDMA technologies. (May/Jun 2018)
9. Write any three features of FDMA. (Nov/Dec 2019)
10. What are the disadvantages of TDMA? (Nov/Dec 2019)

PART-B

1. Compare FDMA, TDMA & CDMA.
2. Briefly explain the principle of cellular networks.
3. Explain the Multiple Access methods with neat diagrams.
4. Identify the channel capacity of TDMA in cell system. (16) (Nov/Dec 2017)
5. Calculate channel capacity of TDMA, FDMA, and CDMA in cell system.

   (Or)

   Derive the expressions for Cellular CDMA schemes for both noises limited an interference limited scenario. (10) (May/ Jun 2017)

   (Or)

   How can capacity of a cellular communication system be improved? Explain any two capacity expansion techniques. (7) (Nov/Dec 2018)

   (Or)

   Compare and contrast TDMA and CDMA. (6) (Nov/Dec 2018)

   (Or)

   Compare and contrast various multiple access techniques used in wireless communication. (6) (May/Jun 2023)

   (Or)

   Describe the principle of CDMA. (5) (May/Jun 2019)
6. If a total of 33 MHz of bandwidth is allocated to a particular FDD cellular telephone system which uses two 25 kHz simplex channels to provide full duplex voice and control channels, compute the number of channels available per cell if a system uses (i) four - cell reuse (2) seven - cell reuse and (3) twelve - cell reuse. If 1 MHz of the allocated spectrums is dedicated to control channels, determine the equitable distribution of control channels and voice channels in each cell of each of three systems. (6) (May/ Jun 2017)
7. Consider Global System for mobile, which is a TDMA/FDD system that uses 25 MHz for the forward link, which is broken in to radio channels of 200 MHz. If 8 speech signals are supported on a single radio channel of 200 MHz If a speech signal is supported on a single radio channel and of no guard band is assumed find the number of simultaneous users that can be accommodated in GSM. (2) (May/ Jun 2017)
8. If GSM uses a frame structure where each frame consists of eight time slots, and each time slot contains 156.25 bits, and data is transmitted at 270.833 kbps in the channel, find (1) the time duration of a bit (2)

the time duration of a slot (3) the timer duration of a frame and (4) how long must a user occupying a single time slot wait between two successive transmissions? (4) (May/ Jun 2017)

9. If W = 1.25 MHz, R = 9600 bps and a minimum acceptable Eb/No is found to be 10 dB, determine the maximum number of users that can be supported in a single-cell CDMA system using
   1) Omni-directional base station antennas and no voice activity detection, and
   2) 3-sectors at the base station and activity detection with $\alpha = 3/8$. Assume the system is interference limited. (5) (Nov/Dec 2019)

UNIT- V WIRELESS NETWORKING

Introduction: Difference between Wireless and Fixed Telephone Networks, The Public Switched Telephone Network (PSTN), Development of Wireless Networks: First Generation Wireless Networks, Second Generation Wireless Networks, Third Generation Wireless Networks, Fixed Network Transmission Hierarchy, Traffic Routing in Wireless Networks: Circuit Switching, Packet Switching- Personal Communication Services/ Networks (PCS/PCNs): Packet Vs Circuit Switching for PCN, Cellular Packet-Switched Architecture- Packet Reservation Multiple Access (PRMA)- Network Databases: Distributed Database for Mobility Management- Universal Mobile Telecommunication Systems (UMTS).

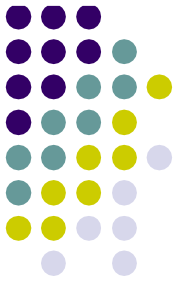
PART A

1. List the capabilities of third generation systems.
2. Differentiate wired and wireless telephone system.
3. What is PSTN?
4. How does PSTN work?
5. Define AMPS.
6. Define GSM.
7. What do you mean by MAHO?
8. What do you mean by connection oriented services?
9. Define circuit switching.
10. What is packet switching?
11. Differentiate circuit and packet switching.
12. Define PRMA.
13. Define UMTS.
14. What are the layers of UMTS?

PART B

1. Explain the services of UMTS in detail.
2. Explain the UMTS core network architecture with neat illustrations.
3. Draw the schematic diagram of PSTN and explain in detail.
4. Explain in detail about the protocols for network access.

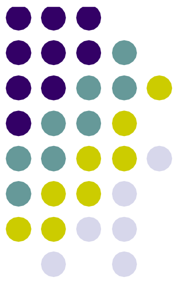
5. Discuss in detail about network databases.
6. With neat diagram and explain in detail about the PCS / PCNs.
7. Explain in detail about the first generation wireless networks.
8. Explain in detail about the second and third generation wireless networks.



UNIT-I

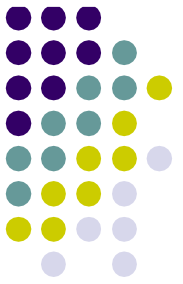
The Cellular Concept-system Design Fundamentals

Outline



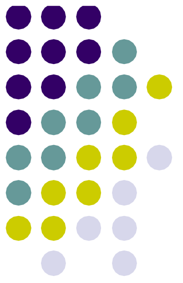
- Introduction
- Frequency Reuse
- Channel Assignment Strategies
- Handoff Strategies
- Interference and System Capacity
- Improving Capacity In Cellular Systems

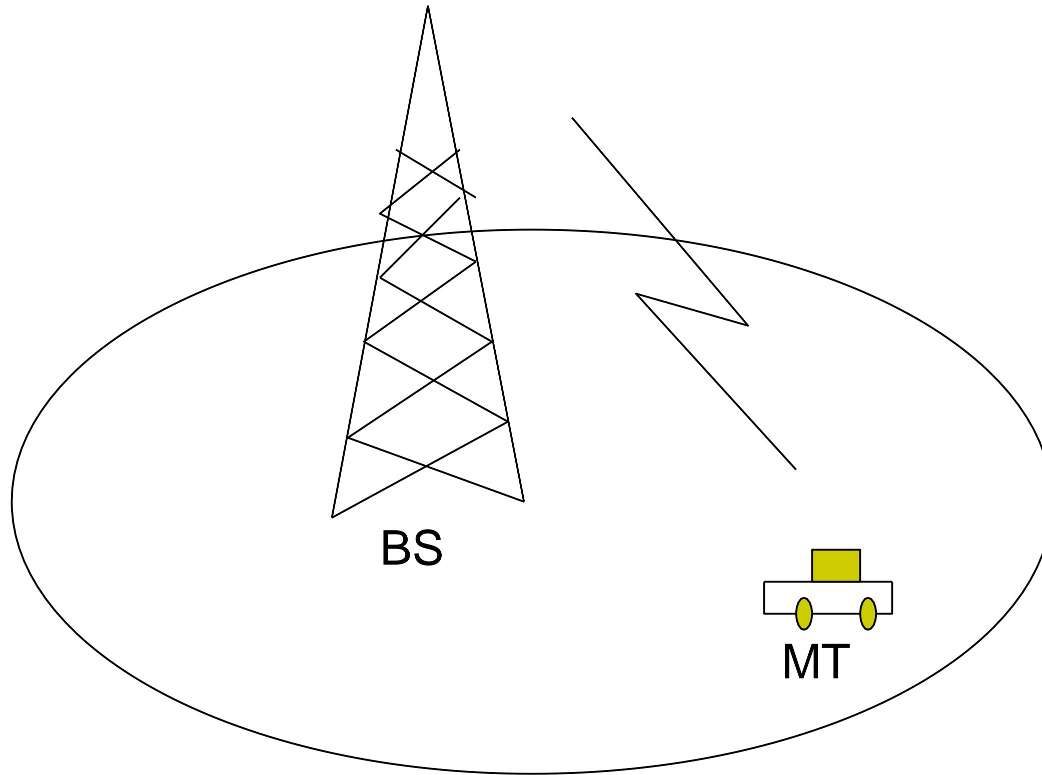
Introduction

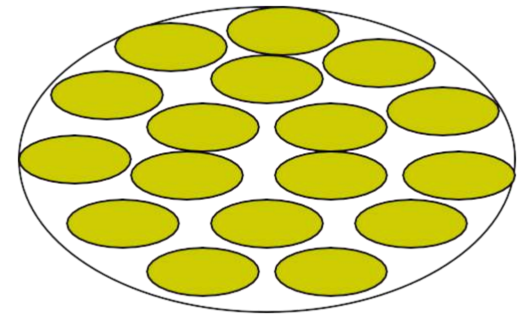


- Early mobile radio systems
  - A single high powered transmitter (single cell)
  - Large coverage area
  - Low frequency resource utility
  - Low user capacity
- The cellular concept
  - A major breakthrough in solving the problem of spectral congestion and user capacity
  - Many low power transmitters (small cells)
  - Each cell covers only a small portion of the service area.
  - Each base station is allocated a portion of the total number of channels
  - Nearby base stations are assigned different groups of channels so that the interference between base stations is minimized

Early mobile radio systems



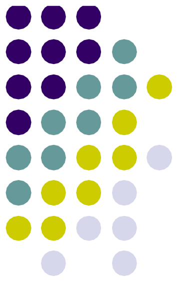




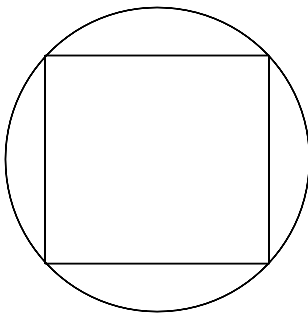
Frequency Reuse

- A service area is split into small geographic areas, called cells.
- Each cellular base station is allocated a group of radio channels.
- Base stations in adjacent cells are assigned different channel groups.
- By limiting the coverage area of a base station, the same group of channels may be reused by different cells far away.
- The design process of selecting and allocating channel groups for all of the cellular base stations within a system is called frequency reuse or frequency planning.

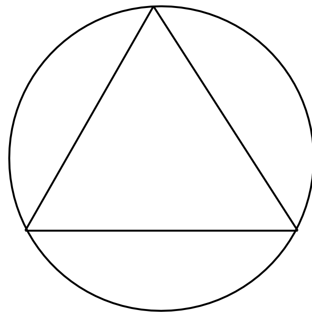
Frequency Reuse: Cell Shapes



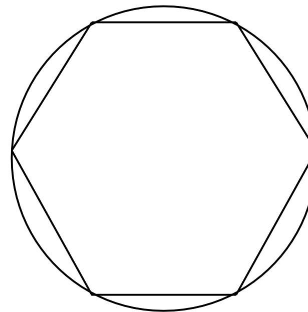
- Geometric shapes covering an entire region without overlap and with equal area.
- By using the hexagon, the fewest number of cells can cover a geographic region, and the hexagon closely approximates a circular radiation pattern which would occur for an omni-directional antenna.



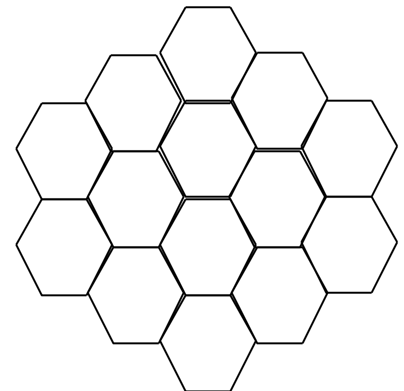
square



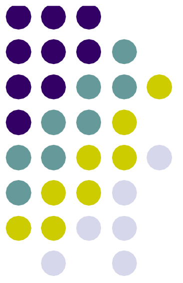
equilateral triangle



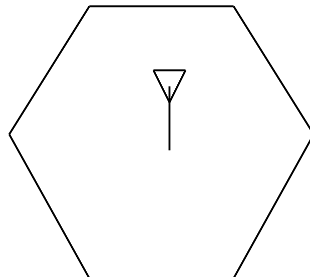
hexagon

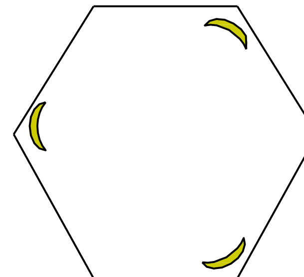


Frequency Reuse: Excitation modes

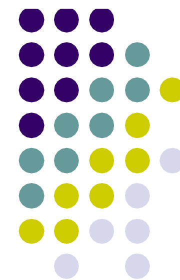


- Center-excited cell
  - Base station transmitter is in the center of the cell.
  - Omni-directional antennas are used.
- Edge-excited cell
  - Base station transmitters are on three of the six cell vertices.
  - Sectored directional antennas are used.

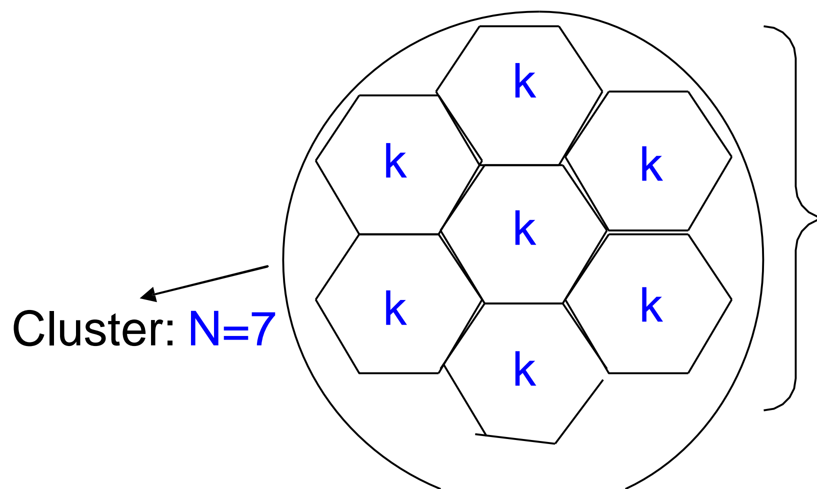




Frequency Reuse: The concept of Cluster



- Consider a cellular system which has a total of S duplex channels available for use.
  - The S channels are divided among N cells (cluster).
  - Each cell is allocated a group of k channels.
  - The total number of available radio channels can be expressed as S=kN.



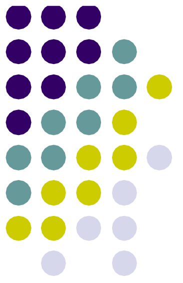
The N cells which collectively use the complete set of available frequencies is called a cluster.

Cluster size: N=4,7,12

Frequency reuse Factor: 1/N

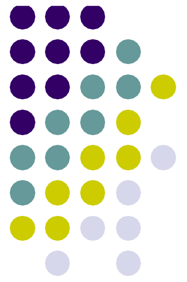
Totally S=kN duplex channels

Frequency Reuse: Reuse Planning



- If a cluster is replicated M times within the system, the total number of duplex channels, C, can be given as $C = MkN = MS$.
- Mathematically, $N = i^2 + ij + j^2$
  - Where i and j are non-negative integers.
  - The nearest co-channel neighbors of a particular cell can be found by doing what follows:
    - move i cells along any chain of hexagons;
    - turn 60 degrees counter-clockwise;
    - move j cells.

Frequency Reuse: Reuse Planning



- Examples

Cluster
N=7
i=2, j=1

Cluster

Cluster

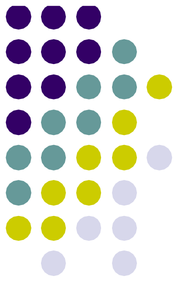
Cluster
N=4

Cluster
N=4
i=2, j=0

7-cell reuse

4-cell reuse

19-cell reuse example (N=19)



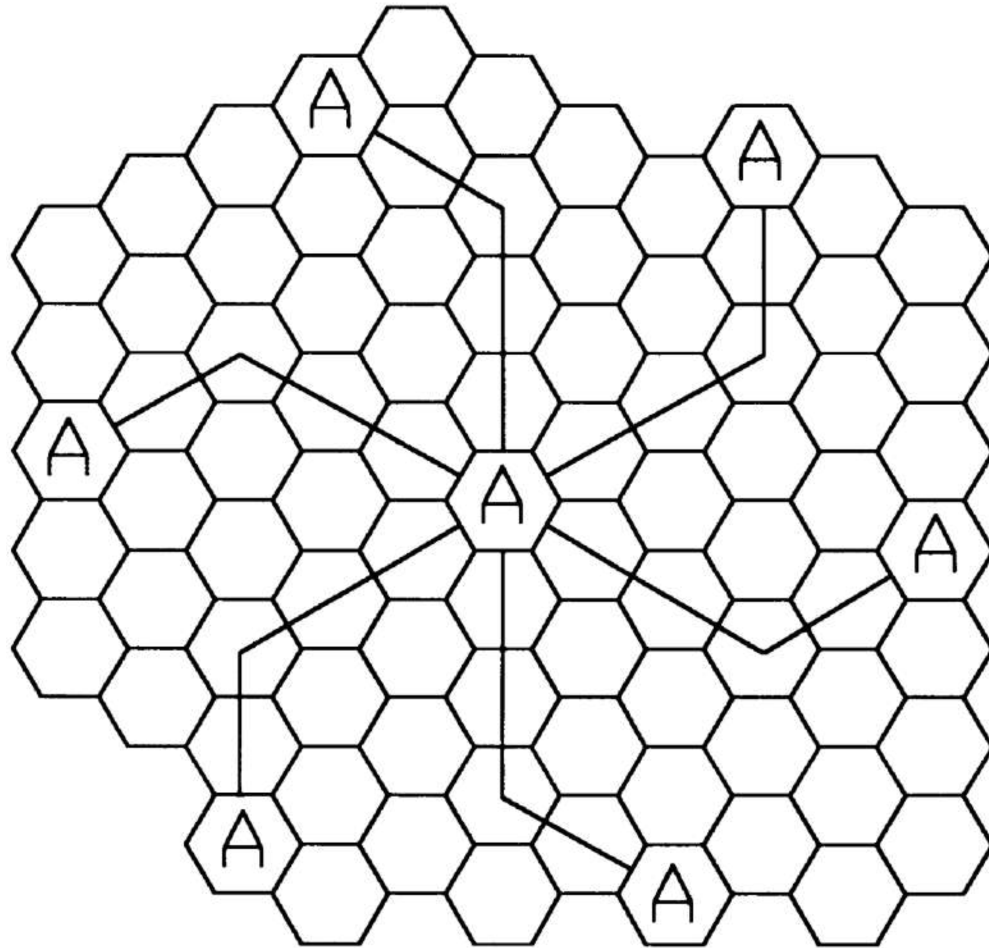
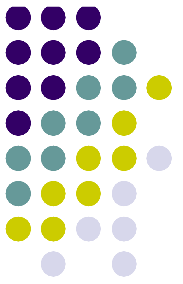


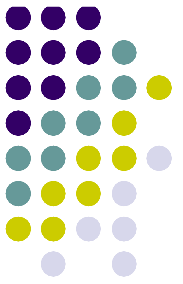
Figure 3.2 Method of locating co-channel cells in a cellular system. In this example, $N = 19$ (i.e., $I = 3$, $j = 2$). (Adapted from [Oet83] © IEEE.)

Channel Assignment Strategies



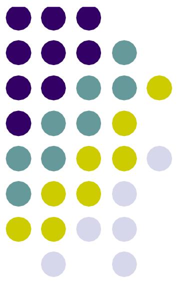
- Objectives:
  - Increasing capacity
  - Minimizing interference
- Classification:
  - Fixed channel assignment strategies
  - Dynamic channel assignment strategies

Fixed channel assignment



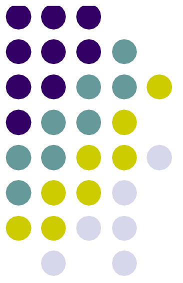
- Each cell is allocated a predetermined set of channels.
- Any call attempt within the cell can only be served by the unused channels in that particular cell.
- If all the channels in that cell are occupied, the call is blocked and the subscriber does not receive service.

Dynamic channel assignment strategies

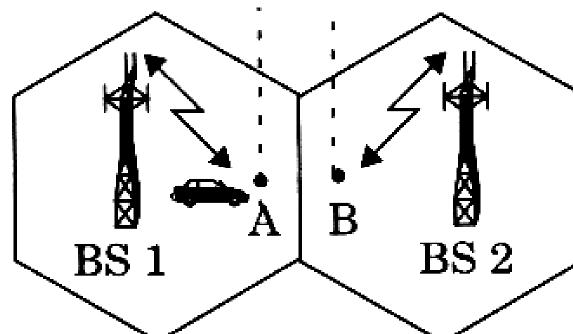


- Channels are not allocated to different cells permanently.
- Each time a call request is made, the serving base station requests a channel from the MSC.
- The switch then allocates a channel to the requested cell following an algorithm that takes into account:
  - the likelihood of fixture blocking within the cell
  - the frequency of use of the candidate channel
  - the reuse distance of the channel
  - other cost functions.

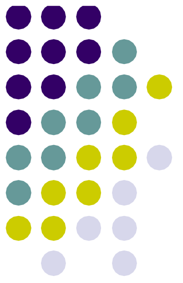
Handoff Strategies



- Handoff:
  - When a mobile moves into a different cell while a conversation is in progress, the MSC automatically transfers the call to a new channel belonging to the new base station.
- Processing handoffs is an important task in any cellular radio system.

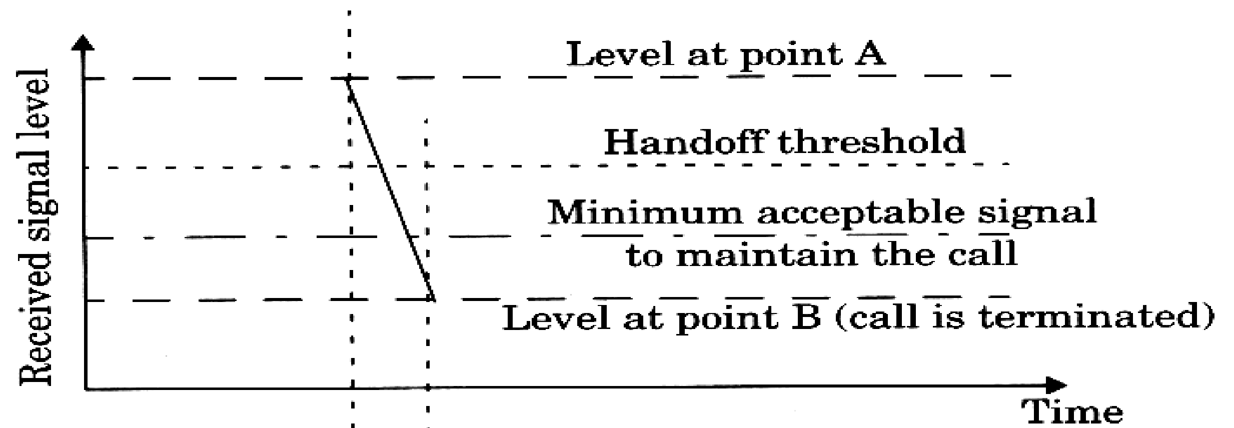


Handoff Strategies: Requirements

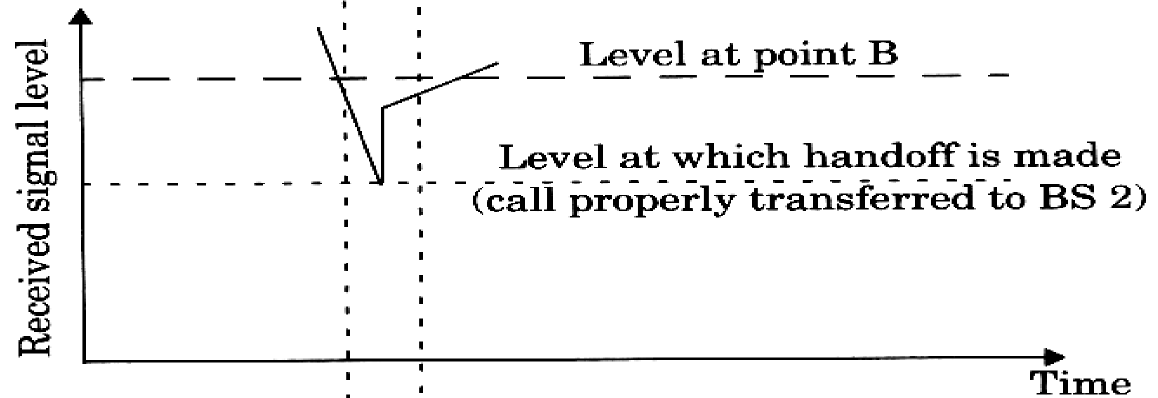


- Handoffs must be performed:
  - Successfully;
  - As infrequently as possible;
  - Imperceptible to the users.
- How to meet these requirements?
  - Specify an optimum signal level to initiate a handoff;
  - Decide optimally when to handoff;
  - Consider the statistics of dwell time.

(a) Improper handoff situation



(b) Proper handoff situation



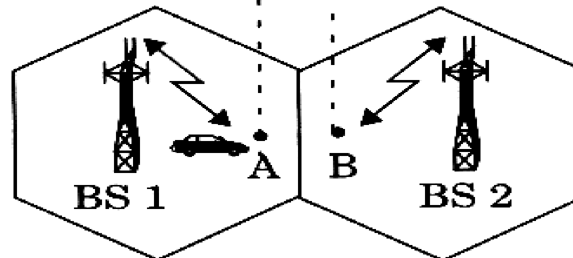
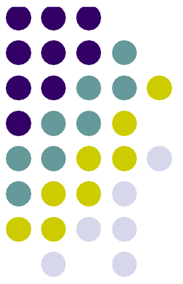


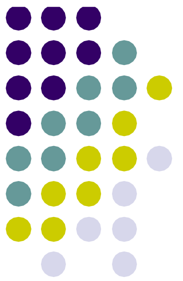
Figure 3.3 Illustration of a handoff scenario at cell boundary.

Handoff Strategies: Signal strength measurements



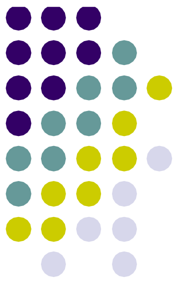
- First generation analog cellular systems:
  - Signal strength measurements are made by the base stations and supervised by the MSC.
- Second generation systems:
  - Handoff decisions are mobile assisted;
  - The MSC no longer constantly monitors signal strengths.

Handoff Strategies: Managing of handoffs



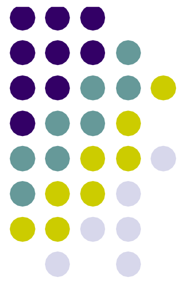
- Prioritizing Handoffs
  - Guard channel: a fraction of the total available channels in a cell is reserved exclusively for handoff requests from ongoing calls which may be handed off into the cell.
  - Queuing of handoff requests: to decrease the probability of forced termination of a call due to lack of available channels.
    - Queuing of handoffs is possible due to the fact that there is a finite time interval between the time the received signal level drops below the handoff threshold and the time the call is terminated due to insufficient signal level.

Practical Handoff Considerations



- Observations
  - High speed vehicles pass through the coverage region of a cell within a matter of seconds.
  - Pedestrian users may never need a handoff during a call.
  - Particularly with the addition of microcells to provide capacity, the MSC can quickly become burdened if high speed users are constantly being passed between very small cells.
  - It is difficult for cellular service providers to obtain new physical cell site locations in urban areas.
  - Another practical handoff problem in microcell systems is known as cell dragging.
- Solutions
  - The umbrella cell approach.
  - Newer cellular systems make handoff decisions based on a wide range of metrics other than signal strength.
  - Soft handoff (CDMA cellular networks): The ability to select between the instantaneous received signals from a variety of base stations is called soft handoff.

The umbrella cell approach

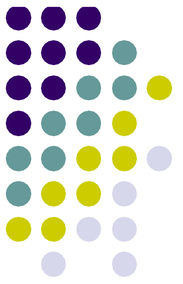


Small microcells for low speed traffic

Large “umbrella” cell for high speed traffic

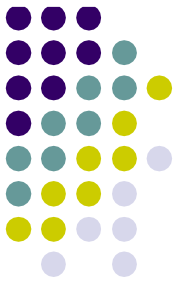
Figure 3.4 The umbrella cell approach.

Interference and System Capacity



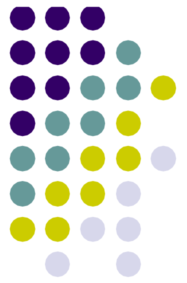
- Interference is the major limiting factor in the performance of cellular radio systems:
  - a major bottleneck in increasing capacity
  - often responsible for dropped calls
- The two major types of system-generated cellular interference are:
  - co-channel interference
  - adjacent channel interference
- Power Control for Reducing Interference

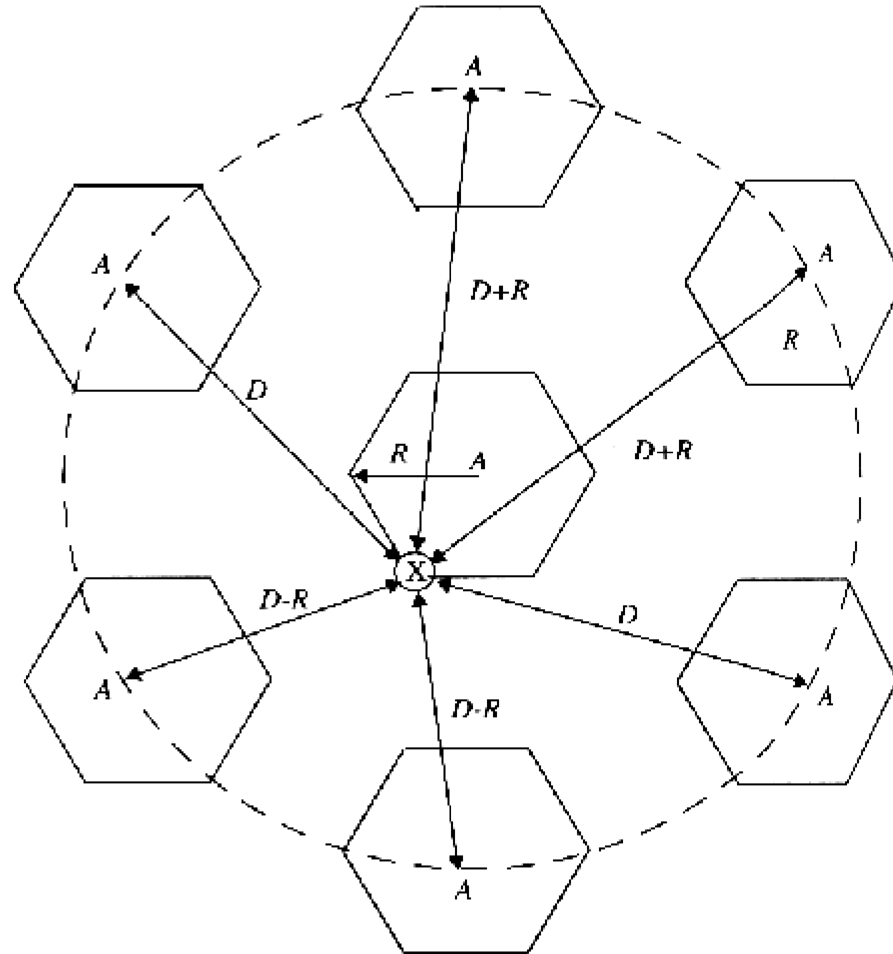
Co-channel Interference and System Capacity



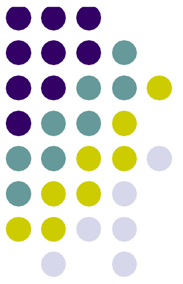
- **Co-channel Interference**
  - Cells using the same set of frequencies are called co-channel cells, and the interference between signals from these cells is called co-channel interference.
  - Unlike thermal noise which can be overcome by increasing the signal-to-noise ration (SNR), co-channel interference cannot be combated by simply increasing the carrier power of a transmitter. This is because an increase in carrier transmit power increases the interference to neighboring co-channel cells.
  - To reduce co-channel interference, co-channel cells must be physically separated by a minimum distance to provide sufficient isolation due to propagation.

Co-channel cells for 7-cell reuse



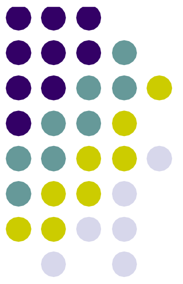


Co-channel Interference and System Capacity



- The co-channel interference ratio is a function of the radius of the cell (B) and the distance between centers of the nearest co-channel cells (D).
- By increasing the ratio of D/R, the spatial separation between co-channel cells relative to the coverage distance of a cell is increased. Thus interference is reduced.

Cochannel reuse ratio



- The parameter $Q = D/R$, called the cochannel reuse ratio, is related to the cluster size N.
  - When the size of each cell is approximately the same, and the base stations transmit the same power, we have

$$Q = D/R = (3N)^{1/3}$$

  - A small value of Q provides larger capacity since the cluster size N is small, whereas a large value of Q improves the transmission quality, due to a smaller level of co-channel interference.
  - A trade-off must be made between these two objectives in actual cellular design.

Smaller N is greater capacity

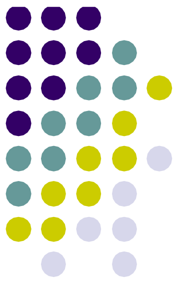
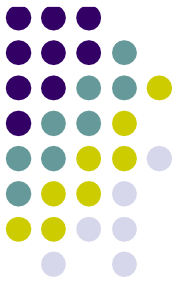


Table 3.1 Co-channel Reuse Ratio for Some Values of N

	Cluster Size (N)	Co-channel Reuse Ratio (Q)
$i = 1, j = 1$	3	3
$i = 1, j = 2$	7	4.58
$i = 2, j = 2$	12	6
$i = 1, j = 3$	13	6.24

Signal-to-interference ratio (SIR)

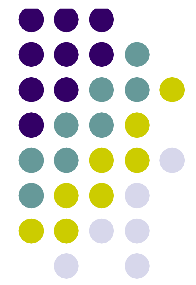


- The signal-to-interference ratio (SIR) for a mobile receiver can be expressed as

$$SIR = \frac{S}{\sum_{i=0}^{i_0} I_i}$$

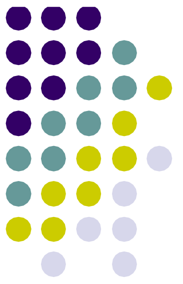
  - S denotes the desired signal power;
  - I_i is the interference power caused by the i-th interfering co-channel cell base station;
  - i_0 is the number of cochannel interfering cells.

Signal-to-interference ratio (SIR)



A
A
D+R
A
R
D
R
A
D+R
X
D-R
D
A
D-R
A
A
A

Signal-to-interference ratio (SIR)



- The average received power P at a distance d from the transmitting antenna is approximated by

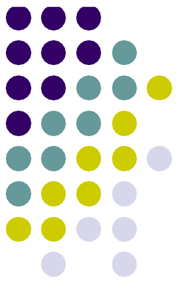
$$P_r = P_0 \left(\frac{d}{d_0} \right)^n$$

- If all base stations transmit at the same power level, the SIR can be given as

$$SIR = \frac{R^{-n}}{\sum_{i=0}^{i_0} D_i^{-n}}$$

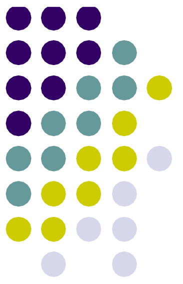
- In practice, measures should be taken to keep the SIR on a acceptable level.

Adjacent Channel Interference



- Interference resulting from signals which are adjacent in frequency to the desired signal is called adjacent channel interference.
- Adjacent channel interference results from imperfect receiver filters which allow nearby frequencies to leak into the passband.
- Near-far effect:
  - If an adjacent channel user is transmitting in very close range to a subscriber's receiver, the problem can be particularly serious.

Adjacent Channel Interference



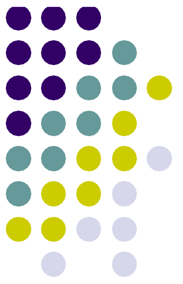
- Adjacent channel interference can be minimized through careful filtering and channel assignments:
  - By keeping the frequency separation between each channel in a given cell as large as possible, the adjacent channel interference may be reduced considerably.
  - Channel allocation schemes can also prevent a secondary source of adjacent channel interference by avoiding the use of adjacent channels in neighboring cell sites.
  - High Q cavity filters can be used in order to reject adjacent channel interference.

Table 3.2 AMPS Channel Allocation for A and B Side Carriers

1A	2A	3A	4A	5A	6A	7A	1B	2B	3B	4B	5B	6B	7B	1C	2C	3C	4C	5C	6C	7C
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126
127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147
148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189
190	191	192	193	194	195	196	197	198	199	20	201	202	203	204	205	206	207	208	209	210
211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231
232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252
253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273
274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294
295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	-	-	-
313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	667	668	669
670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690
691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711
712	713	714	715	716	-	-	-	-	991	992	993	994	995	996	997	998	999	1000	1001	1002

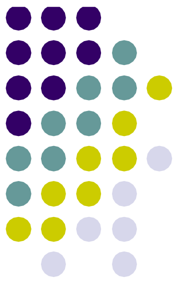
A SIDE

Power Control for Reducing Interference



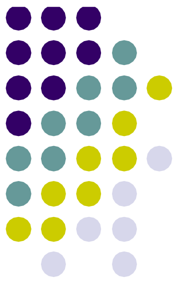
- In practical cellular radio and personal communication systems the power levels transmitted by every subscriber unit are under constant control by the serving base stations.
- This is done to ensure that each mobile transmits the smallest power necessary to maintain a good quality link on the reverse channel.
- Power control not only helps prolong battery life for the subscriber unit, but also dramatically improves the reverse channel S/I in the system.
- Power control is especially important for emerging CDMA spread spectrum systems that allow every user in every cell to share the same radio channel.

Improving Capacity In Cellular Systems



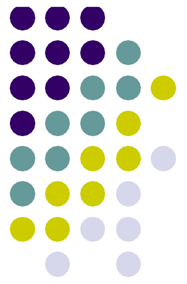
- As the demand for wireless service increases, the number of channels assigned to a cell eventually becomes insufficient to support the required number of users.
- Techniques to expand the capacity of cellular systems :
  - Cell splitting: increases the number of base stations in order to increase capacity.
  - Sectoring: relies on base station antenna placements to improve capacity by reducing co-channel interference.
  - Coverage zone: distributes the coverage of a cell and extends the cell boundary to hard-to-reach places.

Cell Splitting



- Cell splitting is the process of subdividing a congested cell into smaller cells, each with its own base station and a corresponding reduction in antenna height and transmitter power.
- Cell splitting increases the capacity of a cellular system since it increases the number of times that channels are reused.

Cells are split to add channels with no new spectrum usage



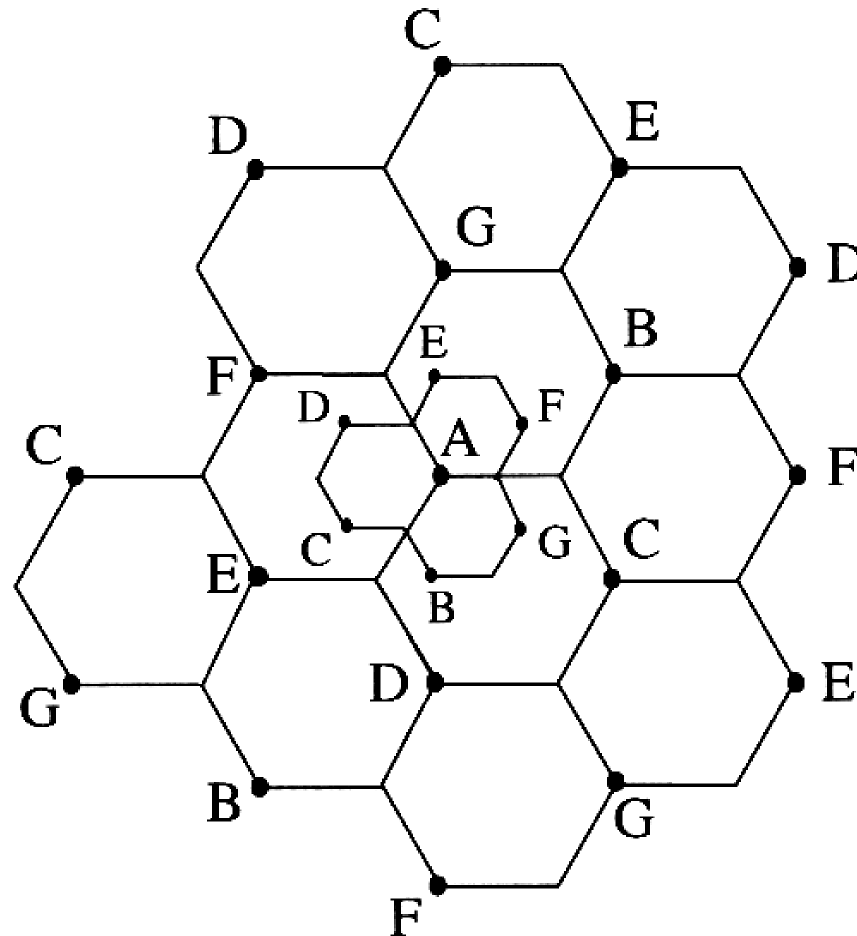
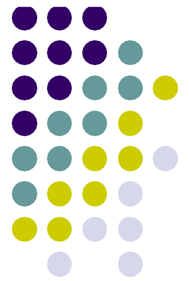


Figure 3.8 Illustration of cell splitting.

Cell Splitting increases capacity



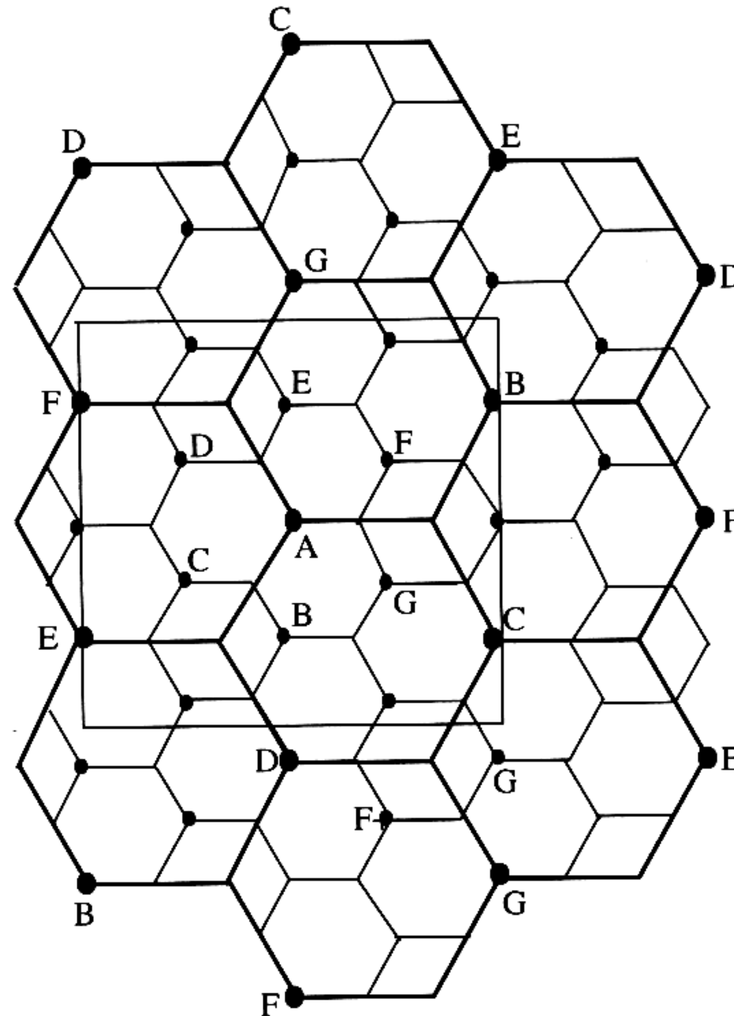
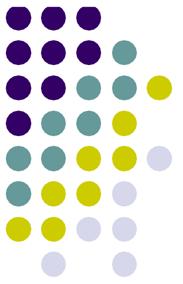


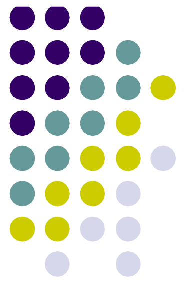
Figure 3.9 Illustration of cell splitting within a 3 km by 3 km square centered around base station A.

Sectoring



- The technique for decreasing co-channel interference and thus increasing system capacity by using directional antennas is called sectoring.
- The factor by which the co-channel interference is reduced depends on the amount of sectoring used.

Sectoring improves S/I



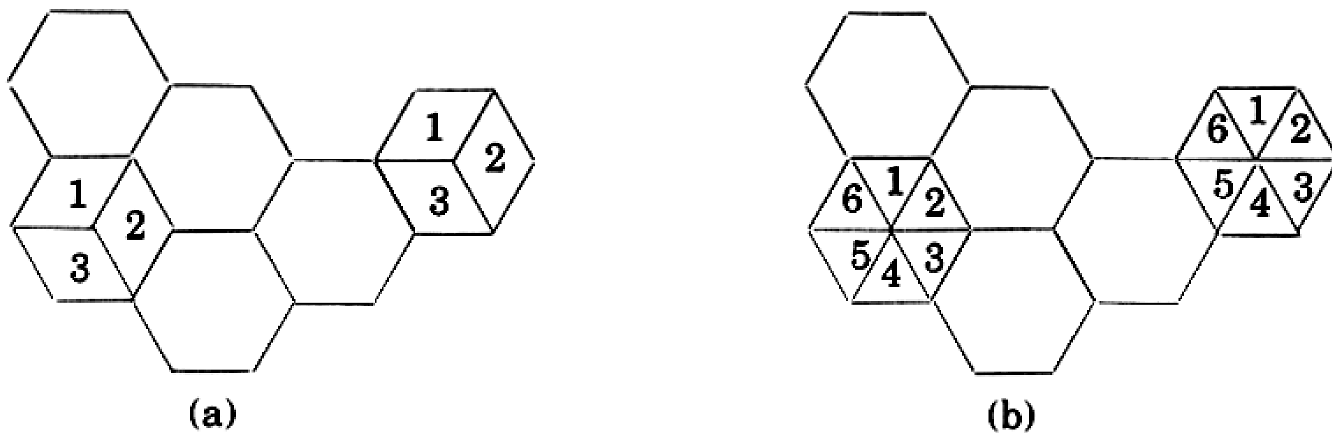
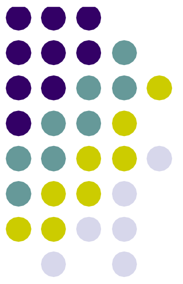


Figure 3.10 (a) 120° sectoring; (b) 60° sectoring.

Sectoring improves S/I



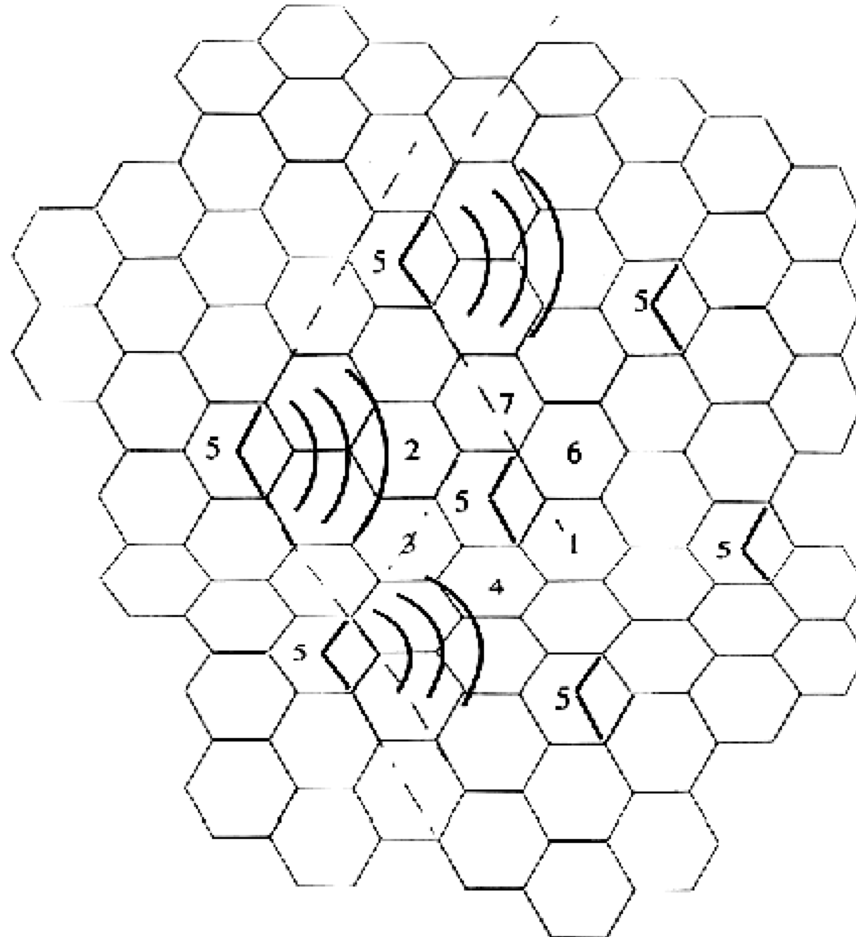
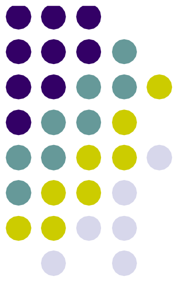


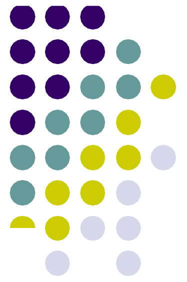
Figure 3.11 Illustration of how 120° sectoring reduces interference from co-channel cells. Out of the 6 co-channel cells in the first tier, only two of them interfere with the center cell. If omnidirectional antennas were used at each base station, all six co-channel cells would interfere with the center cell.

A Novel Microcell Zone Concept



- Zone Concept
  - Zone sites are connected to a single base station and share the same radio equipment.
  - The zones are connected by coaxial cable, fiberoptic cable, or microwave link to the base station.
  - Multiple zones and a single base station make up a cell.
  - As a mobile travels within the cell, it is served by the zone with the strongest signal.
  - This technique is particularly useful along highways or along urban traffic corridors.
- This approach is superior to sectoring since antennas are placed at the outer edges of the cell, and any base station channel may be assigned to any zone by the base station.
- In comparison with sectoring, the number of handoffs can be reduced significantly.

The Zone Cell Concept



Microwave or fiber optic link

Zone Selector

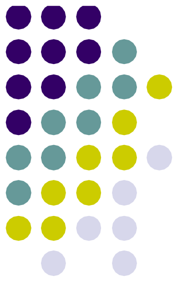
Base Station

Tx/Rx

Tx/Rx

Tx/Rx

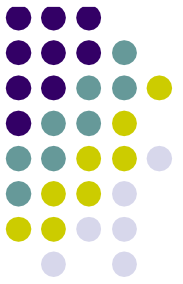
Figure 3.13 The microcell concept [adapted from [Lee91b] © IEEE].



UNIT- II

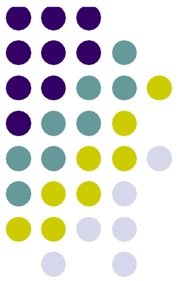
MOBILE RADIO PROPAGATION– Large-Scale Path Loss

Outline



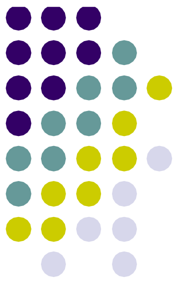
- Introduction to Radio Wave Propagation
- Three Basic Propagation Mechanisms
- Free Space Propagation Model
- Practical Link Budget Design using Path Loss Models
- Outdoor Propagation Models
- Indoor Propagation Models
- Signal Penetration into Buildings

Introduction to Radio Wave Propagation



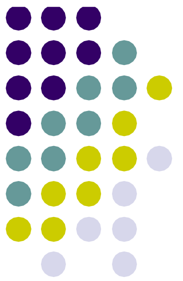
- The mobile radio channel places fundamental limitations on the performance of wireless communication systems.
- Radio channels are extremely random and do not offer easy analysis.
- Modeling radio channel is important for:
  - Determining the coverage area of a transmitter
    - Determine the transmitter power requirement
    - Determine the battery lifetime
  - Finding modulation and coding schemes to improve the channel quality
    - Determine the maximum channel capacity

Introduction to Radio Wave Propagation

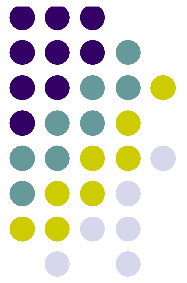


- The mechanisms behind electromagnetic wave propagation are diverse, but can generally be attributed to reflection, diffraction, and scattering.
- Propagation models have traditionally focused on predicting the average received signal strength at a given distance from the transmitter, as well as the variability of the signal strength in close spatial proximity to a particular location.

Introduction to Radio Wave Propagation



- Propagation models that predict the mean signal strength for an arbitrary transmitter-receiver (T-R) separation distance are useful in estimating the radio coverage area of a transmitter and are called large-scale propagation models.
- On the other hand, propagation models that characterize the rapid fluctuations of the received signal strength over very short travel distances (a few wavelengths) or short time durations (on the order of seconds) are called small-scale or fading models.



Received Power (dBm)

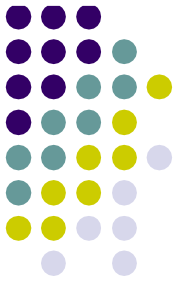
-30 -35 -40 -45 -50 -55 -60 -65 -70

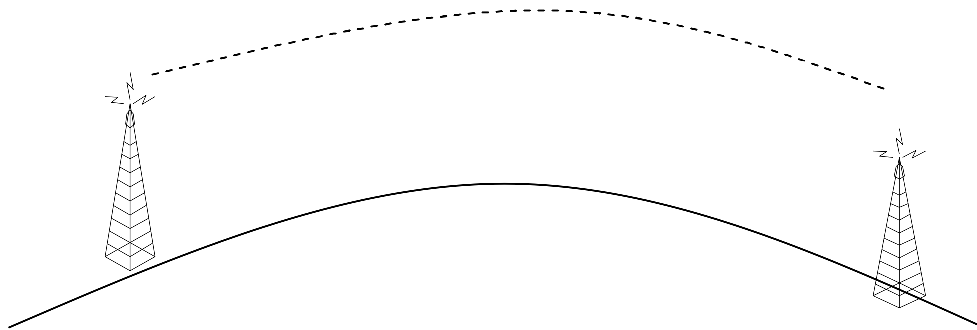
14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

T-R Separation (meters)

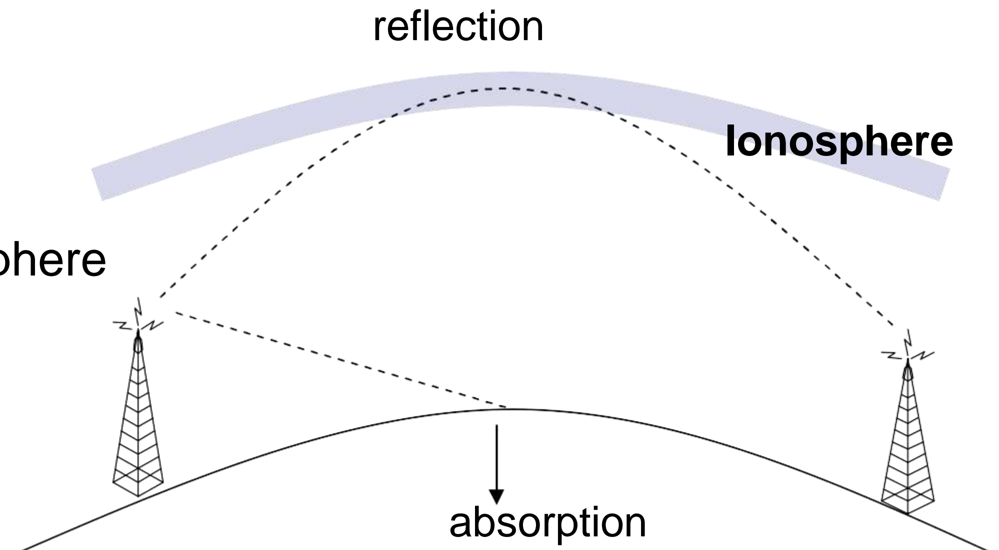
Figure 4.1 Small-scale and large-scale fading.

Basics - Propagation

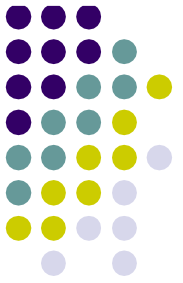




At **VLF, LF, and MF** bands, radio waves follow the ground. AM radio broadcasting uses MF band

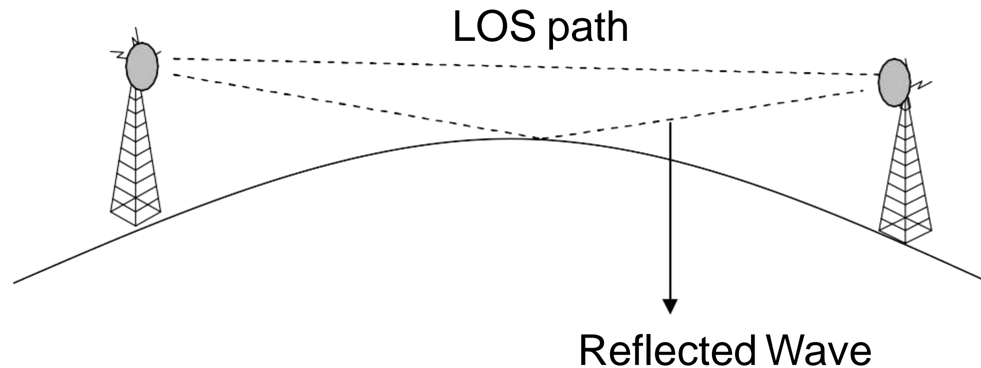


At **HF** bands, the ground waves tend to be absorbed by the earth. The waves that reach ionosphere (100-500km above earth surface), are refracted and sent back to earth.

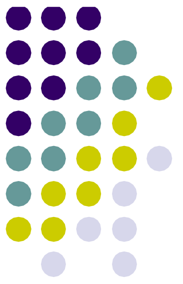


Basics - Propagation

VHF Transmission



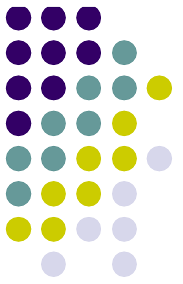
-Directional antennas are used
-Waves follow more direct paths
-LOS: Line-of-Sight Communication
-Reflected wave interfere with the original signal



Basics - Propagation

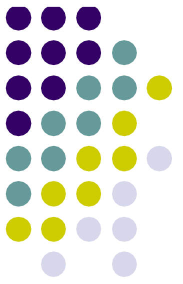
- Waves behave more like light at higher frequencies
  - Difficulty in passing obstacles
  - More direct paths
- They behave more like radio at lower frequencies
  - Can pass obstacles

Radio Propagation Models



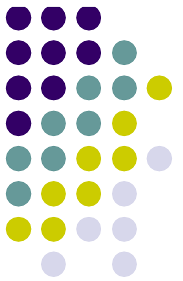
- Transmission path between sender and receiver could be
  - Line-of-Sight (LOS)
  - Obstructed by buildings, mountains and foliage
- Even speed of motion effects the fading characteristics of the channel

Three Radio Propagation Mechanisms



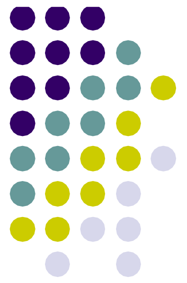
- The physical mechanisms that govern radio propagation are complex and diverse, but generally attributed to the following three factors
    1. Reflection
    2. Diffraction
    3. Scattering
  - Reflection
    - Occurs when waves impinges upon an obstruction that is much larger in size compared to the wavelength of the signal
    - Example: reflections from earth and buildings
    - These reflections may interfere with the original signal constructively or destructively

Three Radio Propagation Mechanisms



- Diffraction
  - Occurs when the radio path between sender and receiver is obstructed by an impenetrable body and by a surface with sharp irregularities (edges)
  - Explains how radio signals can travel urban and rural environments without a line-of-sight path
- Scattering
  - Occurs when the radio channel contains objects whose sizes are on the order of the wavelength or less of the propagating wave and also when the number of obstacles are quite large.
  - They are produced by small objects, rough surfaces and other irregularities on the channel
  - Follows same principles with diffraction
  - Causes the transmitter energy to be radiated in many directions
  - Lamp posts and street signs may cause scattering

Three Radio Propagation Mechanisms



transmitter

R

S

D

D

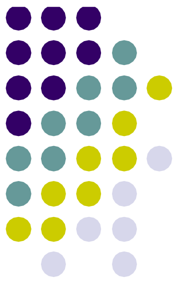
Street

receiver

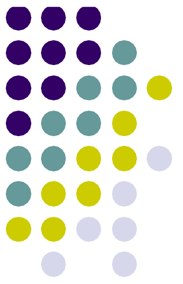
R: Reflection
D: Diffraction
S: Scattering

Building Blocks

Three Radio Propagation Mechanisms



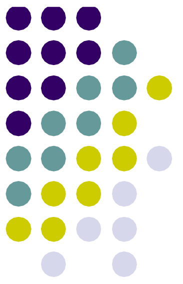
- As a mobile moves through a coverage area, these 3 mechanisms have an impact on the instantaneous received signal strength.
  - If a mobile does have a clear line of sight path to the base-station, than diffraction and scattering will not dominate the propagation.
  - If a mobile is at a street level without LOS, then diffraction and scattering will probably dominate the propagation.



Radio Propagation Models

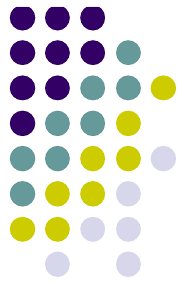
- As the mobile moves over small distances, the instantaneous received signal will fluctuate rapidly giving rise to small-scale fading
  - The reason is that the signal is the sum of many contributors coming from different directions and since the phases of these signals are random, the sum behave like a noise (Rayleigh fading).
  - In small scale fading, the received signal power may change as much as 3 or 4 orders of magnitude (30dB or 40dB), when the receiver is only moved a fraction of the wavelength.

Radio Propagation Models



- As the mobile moves away from the transmitter over larger distances, the local average received signal will gradually decrease. This is called large-scale path loss.
  - Typically the local average received power is computed by averaging signal measurements over a measurement track of 5λ to 40λ. (For PCS, this means 1m-10m track)
- The models that predict the mean signal strength for an arbitrary-receiver transmitter (T-R) separation distance are called large-scale propagation models
  - Useful for estimating the coverage area of transmitters

Small-Scale and Large-Scale Fading



Received Power (dBm)

-30

-40

-50

-60

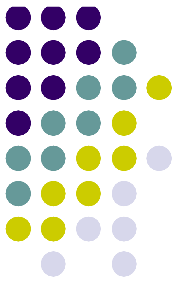
-70

14 16 18 20 22 24 26 28

T-R Separation (meters)

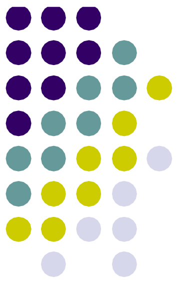
This figure is just an illustration
to show the concept. It is not based on read
data.

Free-Space Propagation Model



- Used to predict the received signal strength when transmitter and receiver have clear, unobstructed LOS path between them.
- The received power decays as a function of T-R separation distance raised to some power.
- Path Loss: Signal attenuation as a positive quantity measured in dB and defined as the difference (in dB) between the effective transmitter power and received power.

Free-Space Propagation Model

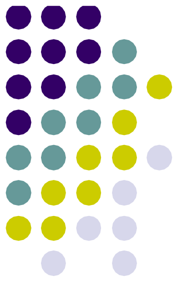


- Free space power received by a receiver antenna separated from a radiating transmitter antenna by a distance d is given by Friis free space equation:

$$P_r(d) = (P_t G_t G_r \lambda^2) / ((4\pi)^2 d^2 L) \qquad \text{[Equation 1]}$$

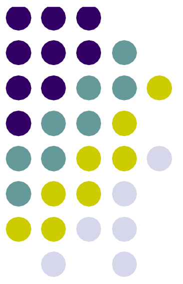
  - P_t is transmited power
  - $P_r(d)$ is the received power
  - G_t is the trasmitter antenna gain (dimensionless quantity)
  - G_r is the receiver antenna gain (dimensionless quantity)
  - d is T-R separation distance in meters
  - L is system loss factor not related to propagation (L >= 1)
    - L = 1 indicates no loss in system hardware (for our purposes we will take L = 1, so we will igonore it in our calculations).
  - λ is wavelength in meters.

Free-Space Propagation Model



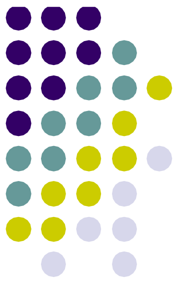
- The gain of an antenna G is related to its affective aperture A_e by:
  - $G = 4\pi A_e / \lambda^2$ [Equation 2]
    - The effective aperture of A_e is related to the physical size of the antenna,
  - λ is related to the carrier frequency by:
    - $\lambda = c/f = 2\pi c / \omega_c$ [Equation 3]
      - f is carrier frequency in Hertz
      - ω_c is carrier frequency in radians per second.
      - c is speed of light in meters/sec

Free-Space Propagation Model



- An *isotropic* radiator is an ideal antenna that radiates power with unit gain uniformly in all directions. It is as the reference antenna in wireless systems.
- The *effective isotropic radiated power* (EIRP) is defined as:
  - $EIRP = P_t G_t$ [Equation 4]
- Antenna gains are given in units of dBi (dB gain with respect to an isotropic antenna) or units of dBd (dB gain with respect to a half-wave dipole antenna).
  - Unity gain means:
    - G is 1 or 0dBi

Free-Space Propagation Model



- Path loss, which represents signal attenuation as positive quantity measured in dB, is defined as the difference (in dB) between the effective transmitted power and the received power.

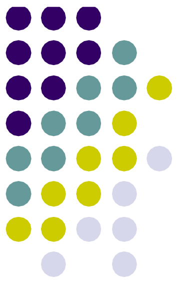
$$PL(dB) = 10 \log (P_t/P_r) = -10\log[(G_t G_r \lambda^2)/(4\pi)^2 d^2] \quad \text{[Equation 5]}$$

(You can drive this from equation 1)

- If antennas have unity gains (exclude them)

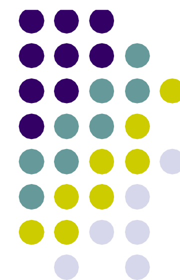
$$PL(dB) = 10 \log (P_t/P_r) = -10\log[\lambda^2/(4\pi)^2 d^2] \quad \text{[Equation 6]}$$

Free-Space Propagation Model



- For Friis equation to hold, distance *d* should be in the far-field of the transmitting antenna.
- The far-field, or Fraunhofer region, of a transmitting antenna is defined as the region beyond the far-field distance d_f given by:
  - $d_f = 2D^2/\lambda$ [Equation 7]
    - D is the largest physical dimension of the antenna.
  - Additionally, $d_f >> D$ and $d_f >> \lambda$

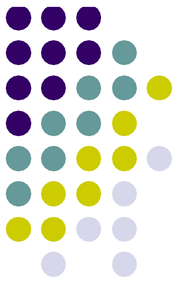
Free-Space Propagation Model – Reference Distance d_0



- It is clear the Equation 1 does not hold for $d = 0$.
- For this reason, models use a close-in distance d_0 as the receiver power reference point.
  - d_0 should be >= d_f
  - d_0 should be smaller than any practical distance a mobile system uses
- Received power $P_r(d)$, at a distance $d > d_0$ from a transmitter, is related to P_r at d_0, which is expressed as $P_r(d_0)$.
- The power received in free space at a distance greater than d_0 is given by:

$$P_r(d) = P_r(d_0)(d_0/d)^2 \quad d >= d_0 >= d_f \qquad \text{[Equation 8]}$$

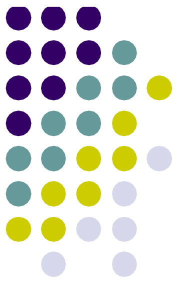
Free-Space Propagation Model



- Expressing the received power in dBm and dBW
  - $P_r(d)$ (dBm) = 10 log $[P_r(d_0)/0.001W]$ + 20log(d_0/d)
    where $d >= d_0 >= d_f$ and $P_r(d_0)$ is in units of watts.
    [Equation 9]
  - $P_r(d)$ (dBW) = 10 log $[P_r(d_0)/1W]$ + 20log(d_0/d)
    where $d >= d_0 >= d_f$ and $P_r(d_0)$ is in units of watts.
    [Equation 10]

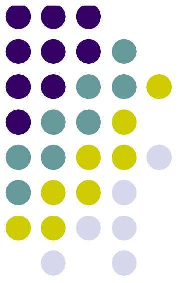
- Reference distance d_0 for practical systems:
  - For frequncies in the range 1-2 GHz
    - 1 m in indoor environments
    - 100m-1km in outdoor environments

Example Question



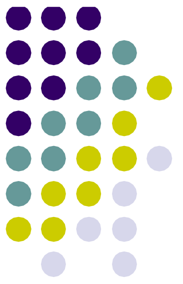
- A transmitter produces 50W of power.
  - A) Express the transmit power in dBm
  - B) Express the transmit power in dBW
  - C) If d_0 is 100m and the received power at that distance is 0.0035mW, then find the received power level at a distance of 10km.
    - Assume that the transmit and receive antennas have unity gains.

Solution



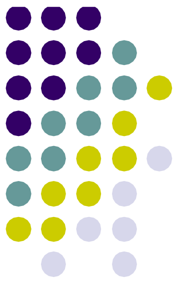
- A)
  - Pt(W) is 50W.
  - Pt(dBm) = 10log[Pt(mW)/1mW)]
    Pt(dBm) = 10log(50x1000)
    Pt(dBm) = 47 dBm
- B)
  - Pt(dBW) = 10log[Pt(W)/1W)]
    Pt(dBW) = 10log(50)
    Pt(dBW) = 17 dBW

Solution



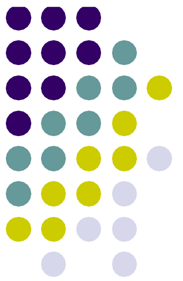
- $P_r(d) = P_r(d_0)(d_0/d)^2$
- Substitute the values into the equation:
  - $P_r(10km) = P_r(100m)(100m/10km)^2$
    $P_r(10km) = 0.0035mW(10^{-4})$
    $P_r(10km) = 3.5x10^{-10}W$
  - $Pr(10km)\ [dBm] = 10log(3.5x10^{-10}W/1mW)$
    $= 10log(3.5x10^{-7})$
    $= \underline{-64.5dBm}$

Two main channel design issues



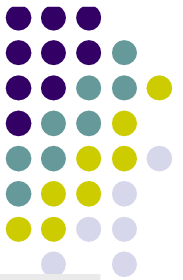
- Communication engineers are generally concerned with two main radio channel issues:
  - Link Budged Design
    - Link budget design determines fundamental quantities such as transmit power requirements, coverage areas, and battery life
    - It is determined by the amount of received power that may be expected at a particular distance or location from a transmitter
  - Time dispersion
    - It arises because of multi-path propagation where replicas of the transmitted signal reach the receiver with different propagation delays due to the propagation mechanisms that are described earlier.
    - Time dispersion nature of the channel determines the maximum data rate that may be transmitted without using equalization.

Link Budged Design Using Path Loss Models



- Radio propagation models can be derived
  - By use of empirical methods: collect measurement, fit curves.
  - By use of analytical methods
    - Model the propagation mechanisms mathematically and derive equations for path loss
- Long distance path loss model
  - Empirical and analytical models show that received signal power decreases logarithmically with distance for both indoor and outdoor channels

Long distance path loss model



- The average large-scale path loss for an arbitrary T-R separation is expressed as a function of distance by using a path loss exponent *n*:
    - The value of *n* depends on the propagation environment: for free space it is 2; when obstructions are present it has a larger value.

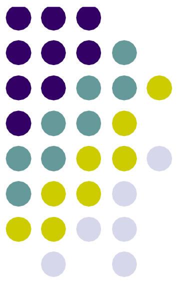
$$\overline{PL}(d) \propto \left(\frac{d}{d_0}\right)^n$$

$$\overline{PL}(dB) = \overline{PL}(d_0) + 10n\log\left(\frac{d}{d_0}\right)$$

Equation 11

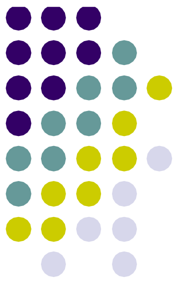
$\overline{PL}(d)$ denotes the average large - scale path loss at a distance d (denoted in dB)

Path Loss Exponent for Different Environments



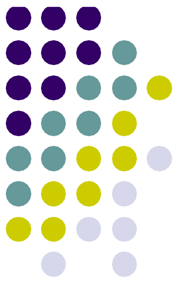
Environment	Path Loss Exponent, *n*
Free space	2
Urban area cellular radio	2.7 to 3.5
Shadowed urban cellular radio	3 to 5
In building line-of-sight	1.6 to 1.8
Obstructed in building	4 to 6
Obstructed in factories	2 to 3

Selection of free space reference distance



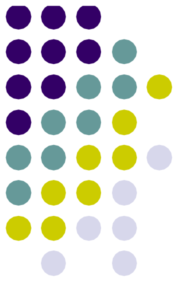
- In large coverage cellular systems
  - 1km reference distances are commonly used
- In microcellular systems
  - Much smaller distances are used: such as 100m or 1m.
- The reference distance should always be in the far-field of the antenna so that near-field effects do not alter the reference path loss.

Log-normal Shadowing



- Equation 11 does not consider the fact the surrounding environment may be vastly different at two locations having the same T-R separation
- This leads to measurements that are different than the predicted values obtained using the above equation.
- Measurements show that for any value d, the path loss PL(d) in dBm at a particular location is random and distributed normally.

Log-normal Shadowing- Path Loss



Then adding this random factor:

$$PL(d)[dB] = \overline{PL}(d) + X_o$$

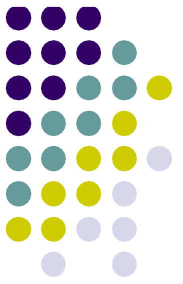
$$PL(d)[dB] = \overline{PL}(d_0) + 10n\log(\frac{d}{d_0}) + X_o$$

Equation 12

- $\overline{PL}(d)$ denotes the average large-scale path loss (in dB) at a distance d.
- X_o is a zero-mean Gaussian (normal) distributed random variable (in dB) with standard deviation σ (also in dB).
- $\overline{PL}(d_0)$ is usually computed assuming free space propagation model between transmitter and d_0 (or by measurement).

Equation 12 takes into account the shadowing affects due to cluttering on the propagation path. It is used as the propagation model for log-normal shadowing environments.

Log-normal Shadowing- Received Power



- The received power in log-normal shadowing environment is given by the following formula (derivable from Equation 12)

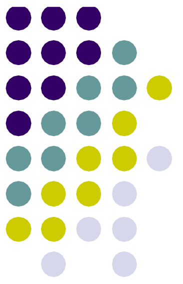
Equation 12

$$P_r(d)[dBm] = P_t[dBm] - PL(d)[dB]$$

$$P_r(d)[dBm] = P_t[dBm] - \left[\overline{PL}(d_0)[dB] + 10n\log(\frac{d}{d_0}) + X_o[dB] \right]$$

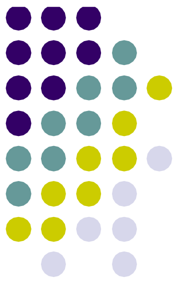
- The antenna gains are included in PL(*d*).

Log-normal Shadowing, *n* and σ



- The log-normal shadowing model indicates the received power at a distance d is normally distributed with a distance dependent mean and with a standard deviation of σ
- In practice the values of n and σ are computed from measured data using linear regression so that the difference between the measured data and estimated path losses are minimized in a mean square error sense.

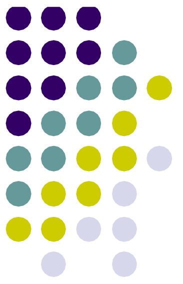
Example of determining n and σ



- Assume $P_r(d0)$ = 0dBm and d_0 is 100m
- Assume the receiver power P_r is measured at distances 100m, 500m, 1000m, and 3000m,
- The table gives the measured values of received power

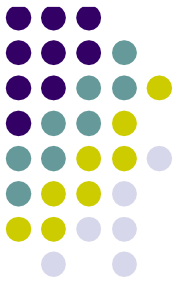
Distance from Transmitter	Received Power
100m	0dBm
500m	-5dBm
1000m	-11dBm
3000m	-16dBm

Example of determining n and σ



- We know the measured values.
- Lets compute the estimates for received power at different distances using long-distance path loss model. (Equation 11)
- $P_r(d_0)$ is given as 0dBm and measured value is also the same.
  - mean_P_r(d) = $P_r(d_0)$ – mean_PL(from_d_0_to_d)
  - Then mean_P_r(d) = 0 – 10log$n$(d/d0)
  - Use this equation to computer power levels at 500m, 1000m, and 3000m.

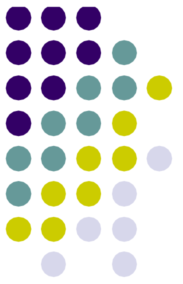
Example of determining n and σ



- $Average_P_r(500m) = 0 - 10\log n(500/100) = -6.99n$
- $Average_P_r(1000m) = 0 - 10\log n(1000/100) = -10n$
- $Average_P_r(3000m) = 0 - 10\log n(3000/100) = -14.77n$

- Now we know the estimates and also measured actual values of the received power at different distances
- In order approximate n, we have to choose a value for n such that the mean square error over the collected statistics is minimized.

Example of determining n and σ: MSE(Mean Square Error)



The mean square error (MSE) is given with the following formula:

$$MSE = \sqrt{\sum_{i=1}^{k} (p_i - \hat{p_i})^2} \quad \text{[Equation 14]}$$

p_i is the actual measured value of power at some distance

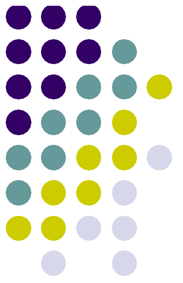
$\hat{p_i}$ is the estimate of power at that distance

k is the number of measurement samples

Since power estimate at some distance depends on *n*, MSE(*n*) is a function of n.

We would like to find a value of n that will minimize this MSE(n) value. We
We will call it MMSE: minimum mean square error.

This can be achieved by writing MSE as a function of n. Then finding the value of n which minimizes this function. This can be done by derivating MSE(n) with respect to n and solving for n which makes the derivative equal to zero.



Example of determining n:

Distance	Measured Value of Pr (dBm)	Estimated Value of Pr (dBm)
100m	0	0
500m	-5	-6.99n
1000m	-11	-10n
3000m	-16	-14.77n

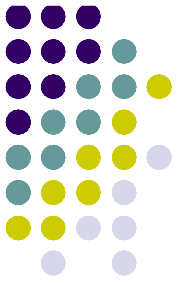
$MSE = (0-0)^2 + (-5-(-6.99n))^2 + (-11-(-10n)^2 + (-16-(-14.77n)^2$

$MSE = 0 + (6.99n - 5)^2 + (10n - 11)^2 + (14.77n - 16)^2$

If we open this, we get MSE as a function of n which as second order polynomial.

We can easily take its derivate and find the value of n which minimizes MSE. (I will not show these steps, since they are trivial).

Example of determining σ:



We are interested in finding the standard deviation about the mean value
For this, we will use the following formula

$$\sigma^2 = \frac{\sum_{i=1}^{k}(p_i - \hat{p}_i)^2}{k}$$

Equation 14.1

p_i is the actual measured value of power at some distance d

$\hat{p}_i$ is the estimate of power at that distance d

k is the number of measuremen t samples

From the above definition of σ, we can derive that :

$$\sigma^2 = MSE(N)/k$$

$$\sigma^2 = MMSE/k$$

$$\sigma = \sqrt{MMSE/k}$$

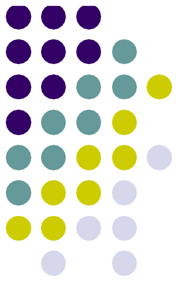
Equation 14.2

where N is the value that minimizes $MSE(n)$

$MMSE$ is minimum mean square error.

$MSE(n)$ formula is given in the previous slides.

Some Statistics Knowledge: Computation of mean (μ), variance (o^2) and standard deviation (o)



Assume we have k samples (k values) $X_1, X_2, \ldots, X_k$:

The mean is denoted by μ.

The variance is denotes by o.

The standard deviation is denotes by o^2.

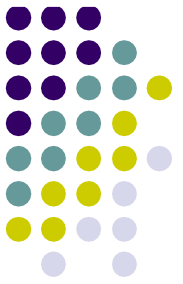
The formulas to computer μ, o, and o^2 is given below:

$$\mu = \frac{\sum_{i=1}^{k} X_i}{k}$$ [Equation 15]

$$o^2 = \frac{\sum_{i=1}^{k} (X_i - \mu)^2}{k}$$ [Equation 16]

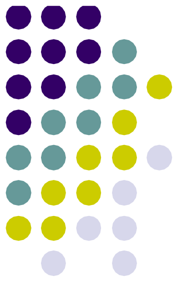
$$o = \sqrt{\frac{\sum_{i=1}^{k} (X_i - \mu)^2}{k}}$$ [Equation 17]

Path loss and Received Power



- In log normal shadowing environment:
  - PL(d) (path loss) and Pr(d) (received power at a distance d) are random variables with a normal distribution in dB about a distance dependent mean.
- Sometime we are interested in answering following kind of questions:
  - What is mean received $P_r(d)$ power (mean_$P_r(d)$)at a distance d from a transmitter
  - What is the probability that the receiver power $P_r(d)$ (expressed in dB power units) at distance d is above (or below) some fixed value μ(again expressed in dB power units such as dBm or dBW).

Received Power and Normal Distribution



- In answering these kind of question, we have to use the properties of normal (gaussian distribution).
- $P_r(d)$ is normally distributed that is characterized by:
  - a mean (μ)
  - a standard deviation (σ)
- We are interested in Probability that $P_r(d) >= \gamma$ or $Pr(d) <= \gamma$

Received Power and Normal Distribution PDF

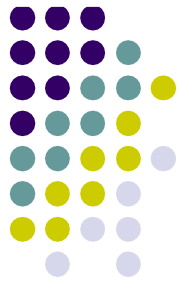
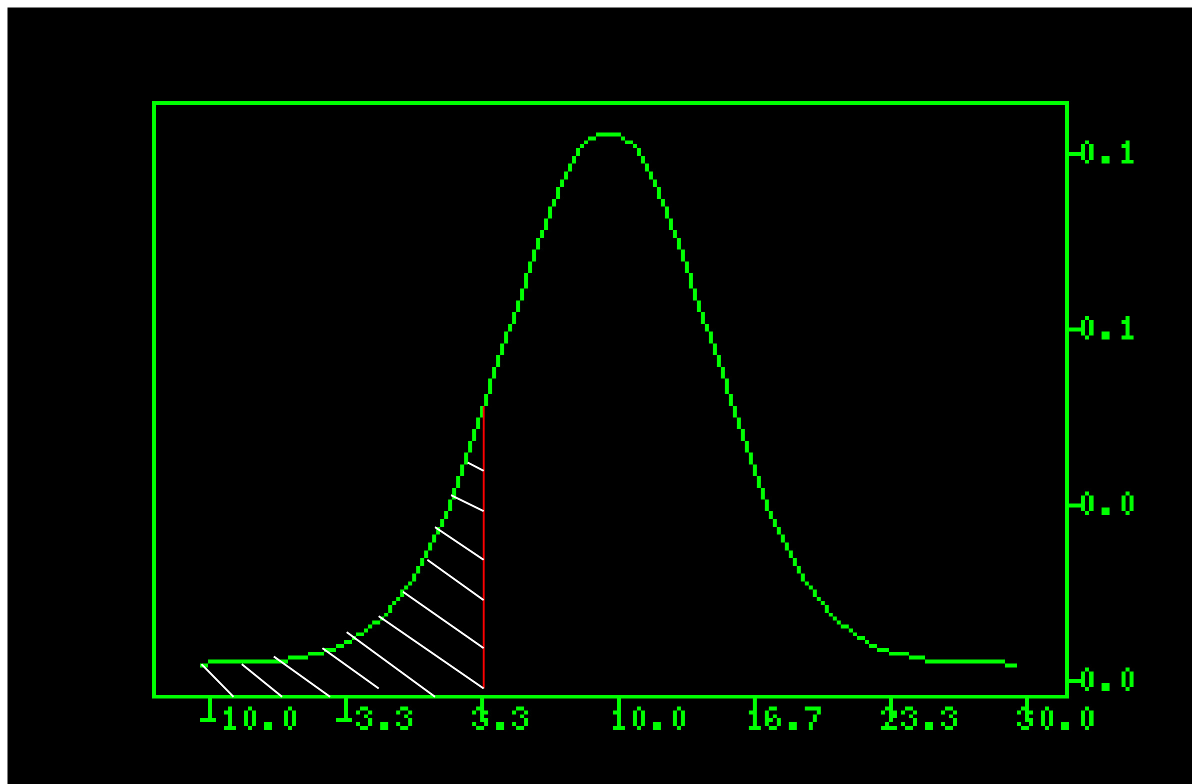


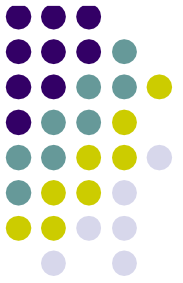
Figure shows the PDF of a normal distribution for the received power P_r at some fixed distance d ($\mu = 10$, $\sigma = 5$)

(x-axis is received power, y-axis probability)



EXAMPLE:

Probability that P_r is smaller than 3.3 (Prob(Pr <= 3.3)) is given with value of the stripped area under the curve.

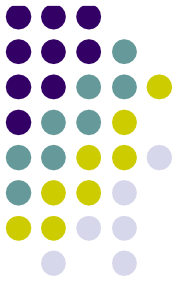


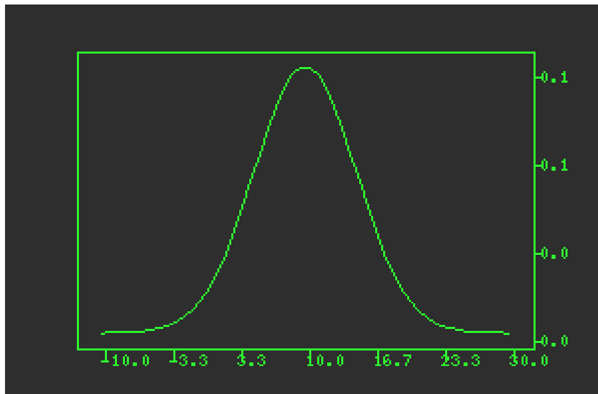
Normal CDF

The figure shows the CDF plot of the normal distribution described previously. Prob(Pr <= 3.3) can be found by finding first the point where vertical line from 3.3 intersects the curve and then by finding the corresponding point on the y-axis. This corresponds to a value of 0.09. Hence Prob(Pr <= 3.3) = 0.09

1.0
0.7
0.3
0.0
-10.0 -3.3 3.3 10.0 16.7 23.3 30.0
0.5
0.090123

Use of Normal Distribution

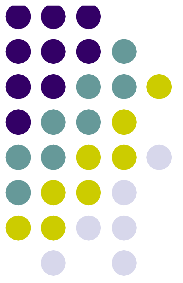




$$p(x)=\frac{1}{o\sqrt{2v}}e^{\frac{-(x-\mu)^2}{2o^2}}$$

[Equation 18]

PDF (probability density function of a normal distribution is characterized by two parameters, μ (mean)and σ (standard deviation), and given with the formula above.



Use of Normal Distribution

To find out the probability that a Gaussian (normal) random variable *X* is above a value *x0,* we have to integrate pdf.

$$\Pr(X > x_0) = \int_{x0}^{\infty} \frac{1}{o\sqrt{2\nu}} e^{\frac{-(x-\mu)^2}{2o^2}} dx$$

Equation 19

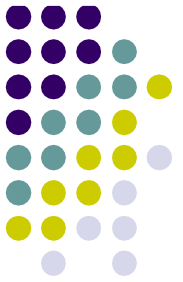
This integration does not have any closed form.

Any Gaussian PDF can be rewritten through substitution of y = x–μ / σ to yield

$$\Pr(y > \frac{x_0 - \mu}{o}) = \int_{(\frac{x0-\mu}{o})}^{\infty} \frac{1}{o\sqrt{2\nu}} e^{\frac{-y^2}{2}} dy$$

Equation 20

Use of Normal Distribution

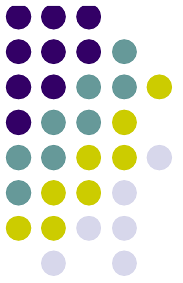


In the above formula, the kernel of the integral is normalized Gaussian PDF function with $\mu = 0$ and $\sigma = 1$.
Evaluation of this function is designed as Q-function and defined as

$$Q(z) = \int_{z}^{\infty} \frac{1}{o\sqrt{2v}} e^{\frac{-y^2}{2}} dy \qquad \text{Equation 21}$$

Hence Equation 19 or 20 can be evaluated as:

$$\Pr(y > \frac{x_0 - \mu}{o}) = Q(\frac{x_0 - \mu}{o}) = Q(z) \qquad \text{Equation 22}$$



Q-Function

Q-Function is bounded by two analytical expressions as follows:

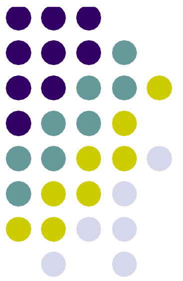
$$(1-\frac{1}{z^2})\frac{1}{z\sqrt{2\nu}}e^{-z^2/2} \leq Q(z) \leq \frac{1}{z\sqrt{2\nu}}e^{-z^2/2}$$ Equation 23

For values greater than 3.0, both of these bounds closely approximate *Q(z).*

Two important properties of *Q(z)* are:

- *Q(-z)* = 1 – *Q(z)* Equation 24
- *Q(0)* = 1/2 Equation 25

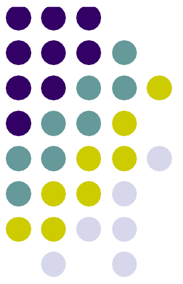
Tabulation of Q-function (0<=z<=3.9)

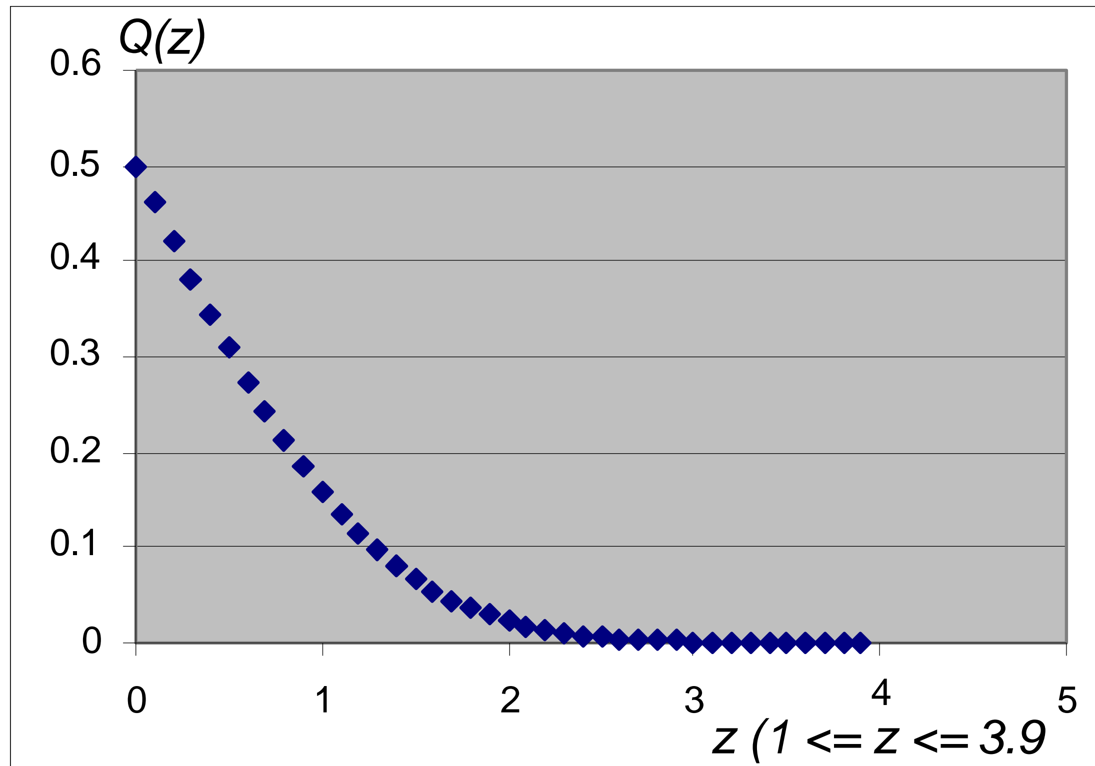


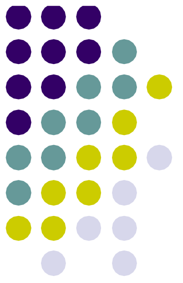
z	Q(z)	z	Q(z)	z	Q(z)	z	Q(z)
0.0	0.5	1.0	0.15866	2.0	0.02275	3.0	0.00135
0.1	0.46017	1.1	0.13567	2.1	0.01786	3.1	0.00097
0.2	0.42074	1.2	0.11507	2.2	0.01390	3.2	0.00069
0.3	0.38209	1.3	0.09680	2.3	0.01072	3.3	0.00048
0.4	0.34458	1.4	0.08076	2.4	0.00820	3.4	0.00034
0.5	0.30854	1.5	0.06681	2.5	0.00621	3.5	0.00023
0.6	0.27425	1.6	0.05480	2.6	0.00466	3.6	0.00016
0.7	0.24196	1.7	0.04457	2.7	0.00347	3.7	0.00011
0.8	0.21118	1.8	0.03593	2.8	0.00256	3.8	0.00007
0.9	0.18406	1.9	0.02872	2.9	0.00187	3.9	0.00005

For values of z higher than 3.9, you should use the equations on the previous slide to compute $Q(z)$.

Q-Function Graph: z versus Q(z)







Erf and Erfc functions

The error function (erf) is defined as:

$$erfc(z) = \frac{2}{\sqrt{\nu}} \int_{z}^{\infty} e^{-x^2} dx \qquad \text{[Equation 26]}$$

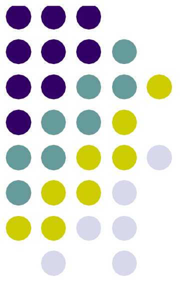
And the complementary error function (erfc) is defined as:

$$erf(z) = \frac{2}{\sqrt{\nu}} \int_{0}^{z} e^{-x^2} dx \qquad \text{[Equation 27]}$$

The *erfc* function is related to *erf* function by:

$$erfc(z) = 1 - erf(z) \qquad \text{[Equation 28]}$$

Erf and Erfc functions



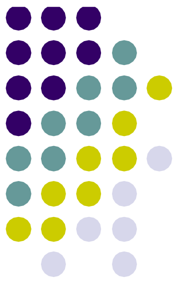
The Q-function is related to erf and erfc functions by:

$$Q(z) = \frac{1}{2}[1 - erf(\frac{z}{\sqrt{2}})] = \frac{1}{2} erfc(\frac{z}{\sqrt{2}})$$ [Equation 29]

$$erfc(z) = 2Q(\sqrt{2}z)$$ [Equation 30]

$$erf(z) = 1 - 2Q(\sqrt{2}z)$$ [Equation 31]

Computation of probability that the received power is below/above a threshold



- We said that $P_r(d)$ is a random variable that is Gaussian distributed with mean μ and std deviation σ. Then:
  - Probability that $P_r(d)$ is above γ is given by:

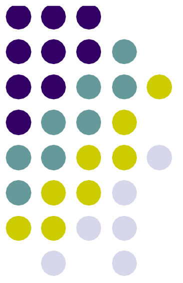
$$\Pr(P_r(d) > \mu) = Q\left(\frac{\mu - \overline{P_r(d)}}{o}\right) \qquad \text{Equation 32}$$

  - Probability that $P_r(d)$ is below γ is given by:

$$\Pr(P_r(d) < \mu) = Q\left(\frac{\overline{P_r(d)} - \mu}{o}\right) \qquad \text{Equation 33}$$

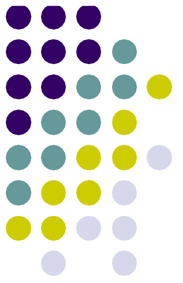
    - P_r(d) bar denotes the average (mean) received power at d.

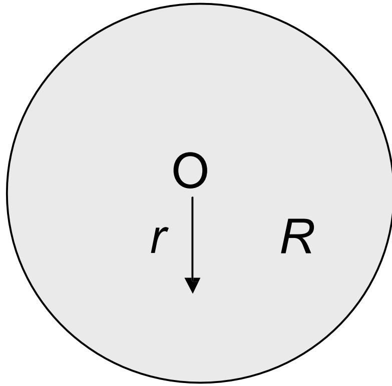
Percentage of Coverage Area



- We are interested in the following problem
  - Given a circular coverage area with radius R from a base station
  - Given a desired threshold power level .
- Find out
  - $U(\gamma)$, the percentage of useful service area
    - i.e the percentage of area with a received signal that is equal or greater than γ, given a known likelihood of coverage at the cell boundary

Percentage of Coverage Area





O is the origin of the cell

r: radial distance *d* from transmitter
0 <= r <= R

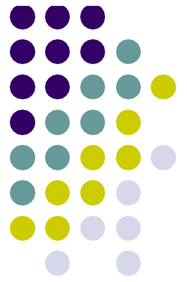
Definition: $P(P_r(r) > \gamma)$ denotes
probability that the random received power at a distance $d = r$ is greater than threshold γ within an incrementally small area *dA*

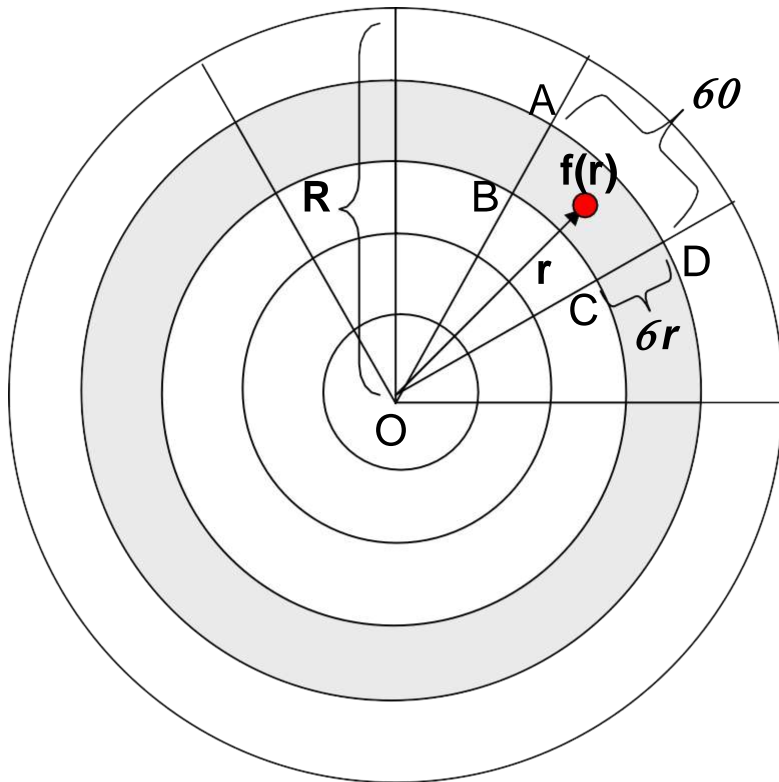
Then $U(\gamma)$ can be found by the following integration over the area of the cell:

$$U(\mu) = \frac{1}{\nu R^2} \int P[P_r(r) > \mu) dA = \frac{1}{\nu R^2} \int_0^{2\nu} \int_0^{R} P[P_r(r) > \mu) r dr d0$$

Equation 34

Integrating f(r) over Circle Area





Lets express the incremental area ΔA between points A, B, C, D.

The area could be approximated as the difference of areas of two sectors:

ΔA = Area(ABCD) ~= Area(OAD) ? Area(OBC)

$$AREA(OAD) = \nu\left(r + \frac{\Delta r}{2}\right)^2 \frac{\Delta 0}{2\nu} = \frac{1}{2}\left(r + \frac{\Delta r}{2}\right)^2 \Delta 0$$

0 is expressed in radians : $0 \le 0 \le 2\nu$

$$AREA(OBC) = \nu\left(r - \frac{\Delta r}{2}\right)^2 \frac{\Delta 0}{2\nu} = \frac{1}{2}\left(r - \frac{\Delta r}{2}\right)^2 \Delta 0$$

$$\Delta A = \frac{1}{2}\left(r + \frac{\Delta r}{2}\right)^2 \Delta 0 - \frac{1}{2}\left(r - \frac{\Delta r}{2}\right)^2 \Delta 0$$

$$\Delta A = \frac{\Delta 0}{2}\left[\left(r + \frac{\Delta r}{2}\right)^2 - \left(r - \frac{\Delta r}{2}\right)^2\right]$$

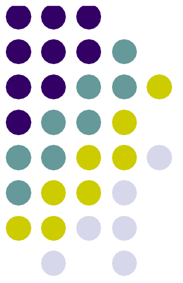
$$\Delta A = \frac{\Delta 0}{2}[2r\Delta r] = r\Delta r\Delta 0$$

Then we can integrate a function f(r) over the surface area of the circle as follows:

$$F = \int f(r)dA$$

$$F = \int_0^{2\nu}\int_0^{R} f(r)rdrd0$$

Percentage of Coverage Area

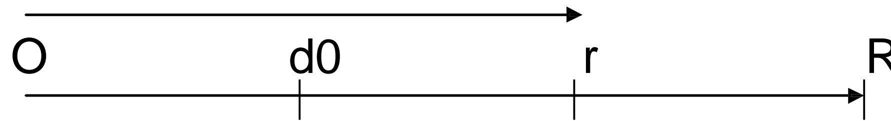


Using equation 32:

$$P(P_r(d) > \mu) = Q\left(\frac{\mu - \overline{P_r(d)}}{\sigma}\right) = Q(\mu - [P_t - \overline{PL}(d_0) + 10n \log(r / d_0)])$$

Equation 33

The path loss at distance r can be expressed as:



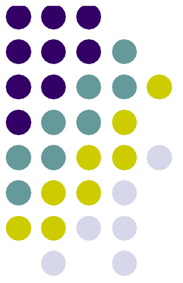
PL(from O to r) = PL(from O to d0) + PL(from d0 to R) - PL(from r to R)
PL(from O to r) = PL(from O to d0) + PL(from d0 to R) + PL(from R to r)
(O is the point where base station is located)

Which can be formally expressed as:

$$\overline{PL}(r) = 10n \log(R / d_0) + 10n \log(r / R) + \overline{PL}(d_0)$$

Equation 34

Percentage of Coverage Area



Equation 33 can be expressed as follows using error function:

$$P(P_r(r) < \mu) = \frac{1}{2} - \frac{1}{2} erf\left(\frac{\mu - [P_t - (\overline{PL}(d) + 10n\log(r/d_0))]}{o\sqrt{2}} \right)$$

Equation 35

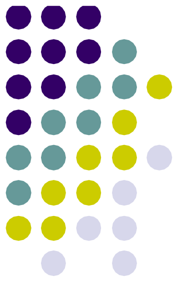
$$P(P_r(d) > \mu) = Q(\frac{\mu - \overline{P_r(d)}}{o}) = \frac{1}{2} - \frac{1}{2} erf(\frac{\mu - \overline{P_r(d)}}{o\sqrt{2}})$$

By combining with Equation 34

$$P(P_r(r) > \mu) = \frac{1}{2} - \frac{1}{2} erf\left(\frac{\mu - [P_t - (\overline{PL}(d_0) + 10n\log(r/d_0) + 10n\log(r/R))]}{o\sqrt{2}} \right)$$

Equation 36

Percentage of Coverage Area



Let the following substitutions happen:

$$a = (\mu - P_t + \overline{PL}(d0) + 10n\log(R/d0))/o\sqrt{2}$$

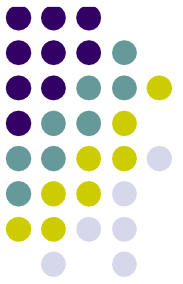
$$b = (10n\log e)/o\sqrt{2}$$

Then

$$U(\mu) = \frac{1}{2} - \frac{1}{R^2}\int_0^R erf\left(a + b\ln\frac{r}{R}\right) r\,dr$$ Equation 37

Substitute $t = a + b\log(r/R)$

$$U(\mu) = \frac{1}{2} - \left(1 - erf(a) + e^{\left(\frac{1-2ab}{b^2}\right)}\left[1 - erf\left(\frac{1-ab}{b}\right)\right]\right)$$ Equation 38



Percentage of Coverage Area

- By choosing a signal level such that $\overline{P_r}(R) = \mu$ (i.e. a =), we obtain:

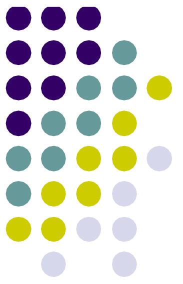
$$U(\mu) = \frac{1}{2} - \left(1 + e^{\left(\frac{1}{b^2}\right)}\left[1 - erf\left(\frac{1}{b}\right)\right]\right)$$

where $b = (10n \log e) / o\sqrt{2}$

Equation 39

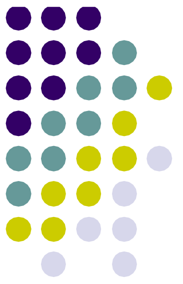
The simplified formula above gives the percentage coverage assuming the mean received power at the cell boundary (r=R) is γ.
In other words, we are assuming: Prob($P_r(R)$ >= γ) = 0.5

Outdoor Propagation



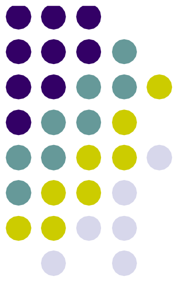
- We will look to the propagation from a transmitter in an outdoor environment
  - The coverage area around a tranmitter is called a cell.
    - Coverage area is defined as the area in which the path loss is at or below a given value.
  - The shape of the cell is modeled as hexagon, but in real life it has much more irregular shapes.
  - By playing with the antenna (tilting and changing the height), the size of the cell can be controlled.
- We will look to the propagation characteristics of the three outdoor environments
  - Propagation in macrocells
  - Propagation in microcells
  - Propagation in street microcells

Macrocells



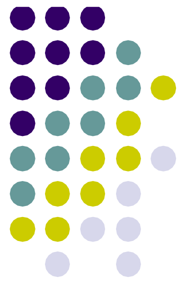
- Base stations at high-points
- Coverage of several kilometers
- The average path loss in dB has normal distribution
  - Avg path loss is result of many forward scattering over a great many of obstacles
    - Each contributing a random multiplicative factor
    - Converted to dB, this gives a sum of random variable
  - Sum is normally distributed because of central limit theorem

Macrocells



- In early days, the models were based on emprical studies
- Okumura did comprehesive measurements in 1968 and came up with a model.
    - Discovered that a good model for path loss was a simple power law where the exponent n is a function of the frequency, antenna heights, etc.
    - Valid for frequencies in: 100MHz – 1920 MHz for distances: 1km – 100km

Okumura Model

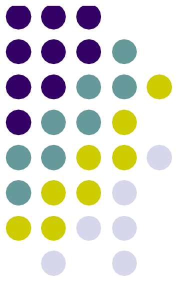


Equation 40

- $L_{50}(d)(dB) = L_F(d) + A_{mu}(f,d) – G(h_{te}) – G(h_{re}) – G_{AREA}$
  - L_{50}: 50th percentile (i.e., median) of path loss
  - $L_F(d)$: free space propagation pathloss.
  - $A_{mu}(f,d)$: median attenuation relative to free space
    - Can be obtained from Okumura's emprical plots shown in the book (Rappaport), page 151.
  - $G(h_{te})$: base station antenna heigh gain factor
  - $G(h_{re})$: mobile antenna height gain factor
  - G_{AREA}: gain due to type of environment

    - $G(h_{te}) = 20\log(h_{te}/200)$ $\quad$ $1000m > h_{te} > 30m$
    - $G(h_{re}) = 10\log(h_{re}/3)$ $\quad$ $h_{re} <= 3m$
    - $G(h_{re}) = 20\log(h_{re}/3)$ $\quad$ $10m > h_{re} > 3m$
      - h_{te}: transmitter antenna height
      - h_{re}: receiver antenna height

Hata Model



- Valid from 150MHz to 1500MHz
- A standard formula
- For urban areas the formula is:
  - $L_{50}(\text{urban,d})(\text{dB}) = 69.55 + 26.16\log f_c - 13.82\log h_{te} - a(h_{re}) + (44.9 - 6.55\log h_{te})\log d$ Equation 41

where

f_c is the ferquency in MHz

h_{te} is effective transmitter antenna height in meters (30-200m)

h_{re} is effective receiver antenna height in meters (1-10m)

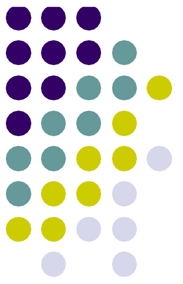
d is T-R separation in km

$a(h_{re})$ is the correction factor for effective mobile antenna height which is a function of coverage area

$a(h_{re}) = (1.1\log f_c - 0.7)h_{re} - (1.56\log f_c - 0.8)$ dB

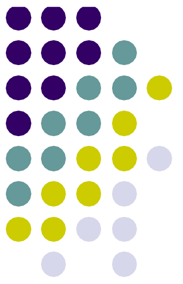
for a small to medium sized city

Microcells



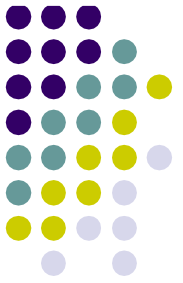
- Propagation differs significantly
  - Milder propagation characteristics
  - Small multipath delay spread and shallow fading imply the feasibility of higher data-rate transmission
  - Mostly used in crowded urban areas
  - If transmitter antenna is lower than the surrounding building than the signals propagate along the streets: Street Microcells

Macrocells versus Microcells



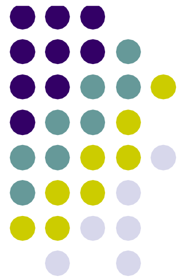
Item	Macrocell	Microcell
Cell Radius	1 to 20km	0.1 to 1km
Tx Power	1 to 10W	0.1 to 1W
Fading	Rayleigh	Nakgami-Rice
RMS Delay Spread	0.1 to 10μs	10 to 100ns
Max. Bit Rate	0.3 Mbps	1 Mbps

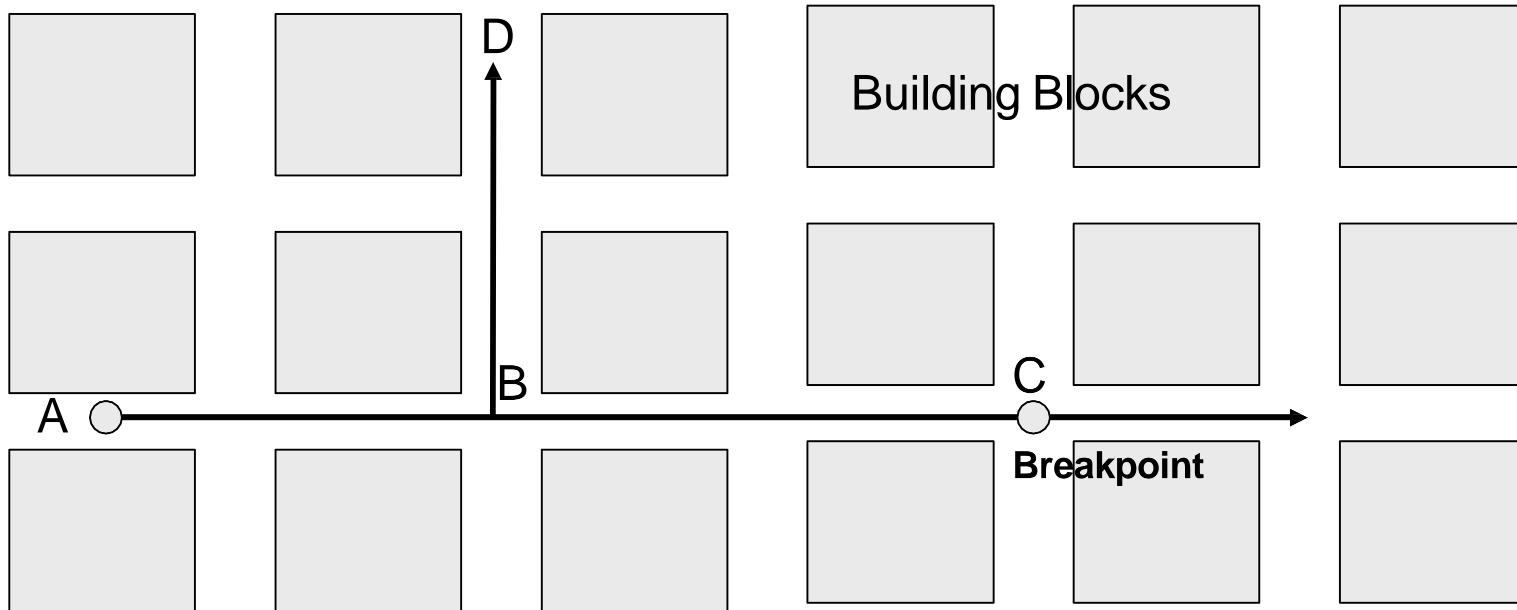
Street Microcells

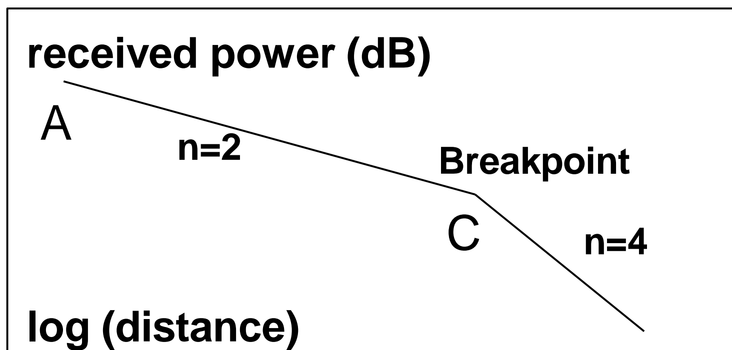


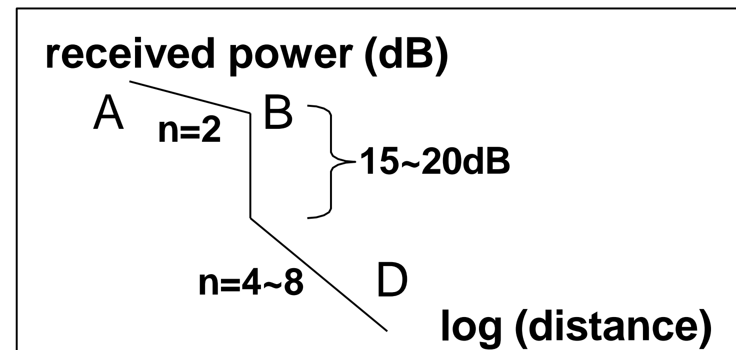
- Most of the signal power propagates along the street.
- The sigals may reach with LOS paths if the receiver is along the same street with the transmitter
- The signals may reach via indirect propagation mechanisms if the receiver turns to another street.

Street Microcells

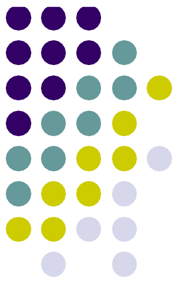






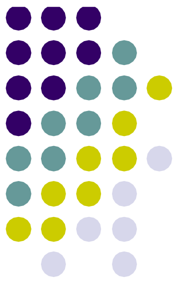


Indoor Propagation



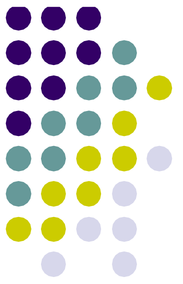
- Indoor channels are different from traditional mobile radio channels in two different ways:
  - The distances covered are much smaller
  - The variablity of the environment is much greater for a much smaller range of T-R separation distances.
- The propagation inside a building is influenced by:
  - Layout of the building
  - Construction materials
  - Building type: sports arena, residential home, factory,...

Indoor Propagation



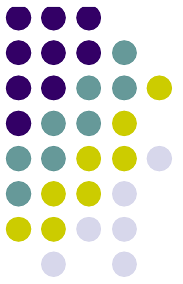
- Indoor propagation is domited by the same mechanisms as outdoor: reflection, scattering, diffraction.
  - However, conditions are much more variable
    - Doors/windows open or not
    - The mounting place of antenna: desk, ceiling, etc.
    - The level of floors
- Indoor channels are classified as
  - Line-of-sight (LOS)
  - Obstructed (OBS) with varying degrees of clutter.

Indoor Propagation



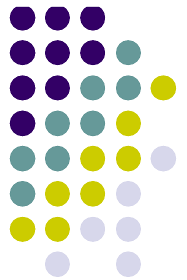
- Buiding types
  - Residential homes in suburban areas
  - Residential homes in urban areas
  - Traditional office buildings with fixed walls (hard partitions)
  - Open plan buildings with movable wall panels (soft partitions)
  - Factory buildings
  - Grocery stores
  - Retail stores
  - Sport arenas

Indoor propagation events and parameters



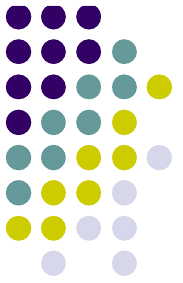
- Temporal fading for fixed and moving terminals
  - Motion of people inside building causes Ricean Fading for the stationary receivers
  - Portable receivers experience in general:
    - Rayleigh fading for OBS propagation paths
    - Ricean fading for LOS paths.
- Multipath Delay Spread
  - Buildings with fewer metals and hard-partitions typically have small rms delay spreads: 30-60ns.
    - Can support data rates excess of several Mbps without equalization
  - Larger buildings with great amount of metal and open aisles may have rms delay spreads as large as 300ns.
    - Can not support data rates more than a few hundred Kbps without equalization.
- Path Loss
  - The following formula that we have seen earlier also describes the indoor path loss:
    - $PL(d)[dBm] = PL(d_0) + 10n\log(d/d_0) + X\sigma$
      - n and σ depend on the type of the building
      - Smaller value for σ indicates the accuracy of the path loss model.

Path Loss Exponent and Standard Deviation Measured for Different Buildings



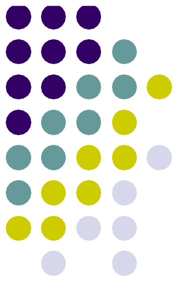
Building	Frequency (MHz)	n	σ (dB)
Retail Stores	914	2.2	8.7
Grocery Store	914	1.8	5.2
Office, hard partition	1500	3.0	7.0
Office, soft partition	900	2.4	9.6
Office, soft partition	1900	2.6	14.1
Factory LOS			
Textile/Chemical	1300	2.0	3.0
Textile/Chemical	4000	2.1	7.0
Paper/Cereals	1300	1.8	6.0
Metalworking	1300	1.6	5.8
Suburban Home			
Indoor Street	900	3.0	7.0
Factory OBS			
Textile/Chemical	4000	2.1	9.7
Metalworking	1300	3.3	6.8

In building path loss factors



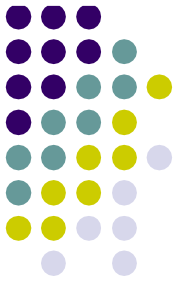
- Partition losses (same floor)
- Partition losses between floors
- Signal Penetration into Buildings

Partition Losses



- There are two kind of partition at the same floor:
  - Hard partions: the walls of the rooms
  - Soft partitions: moveable partitions that does not span to the ceiling

  - The path loss depends on the type of the partitions

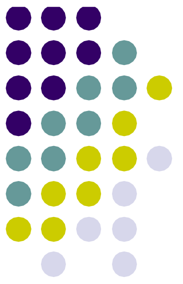


Partition Losses

Average signal loss measurements reported by various researches for radio paths obscructed by some common building material.

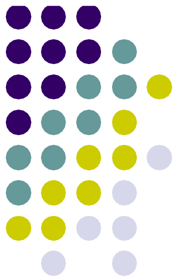
Material Type	Loss (dB)	Frequency (MHz)
All metal	26	815
Aluminim Siding	20.4	815
Concerete Block Wall	3.9	1300
Loss from one Floor	20-30	1300
Turning an Angle in a Corridor	10-15	1300
Concrete Floor	10	1300
Dry Plywood (3/4in) – 1 sheet	1	9600
Wet Plywood (3/4in) – 1 sheet	19	9600
Aluminum (1/8in) – 1 sheet	47	9600

Partition Losses between Floors



- The losses between floors of a building are determined by
  - External dimensions and materials of the building
  - Type of construction used to create floors
  - External surroundings
  - Number of windows
  - Presence of tinting on windows

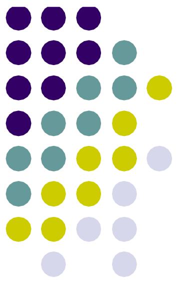
Partition Losses between Floors



Average Floor Attenuation Factor in dB for One, Two, Three and Four Floors in Two Office Buildings

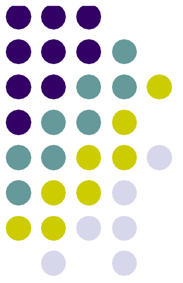
Building	FAF (dB)	σ (dB)
Office Building 1		
Through 1 Floor	12.9	7.0
Through 2 Floors	18.7	2.8
Through 3 Floors	24.4	1.7
Through 4 Floors	27.0	1.5
Office Building 2		
Through 1 Floor	16.2	2.9
Through 2 Floors	27.5	5.4
Through 3 Floors	31.6	7.2

Signal Penetration Into Buildings



- RF signals can penetrate from outside transmitter to the inside of buildings
  - However the siganls are attenuated
- The path loss during penetration has been found to be a function of:
  - Frequency of the signal
  - The height of the building

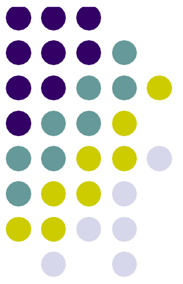
Signal Penetration Into Buildings



- Effect of Frequency
  - Penetration loss decreases with increasing frequency

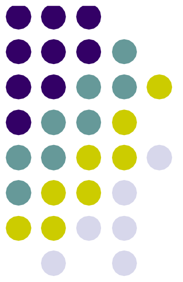
Frequency (MHz)	Loss (dB)
441	16.4
896.5	11.6
1400	7.6

- Effect of Height
  - Penetration loss decreases with the height of the building up-to some certain height
    - At lower heights, the urban clutter induces greater attenuation
  - and then it increases
    - Shadowing affects of adjascent buildings



Conclusion

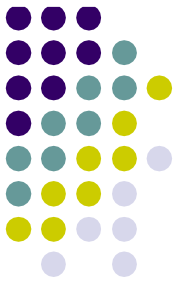
- More work needs to be done to understand the characteristics of wireless channels
- 3D numerical modeling approaches exist
- To achieve PCS, new and novel ways of classifying wireless environments will be needed that are both widely encompassing and reasonably compact.



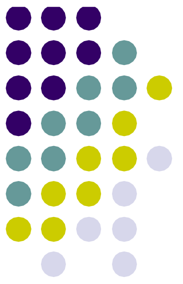
UNIT III

CELLULAR SYSTEM DESIGN FUNDAMENTALS

Outline



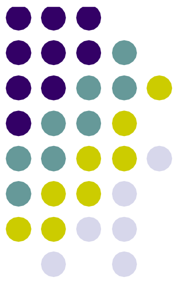
- Small-Scale Multipath Propagation
- Impulse Response Model of a Multipath Channel
- Small-Scale Multipath Measurements
- Parameters of Mobile Multipath Channels
- Types of Small-Scale Fading
- Rayleigh and Ricean Distributions
- Statistical Models for Multipath Fading Channels



Small Scale Fading

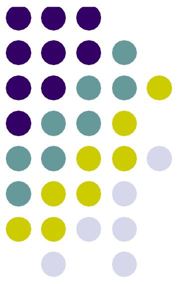
- Describes *rapid fluctuations* of the amplitude, phase of multipath delays of a radio signal over short period of time or travel distance
- Caused by interference between two or more versions of the transmitted signal which arrive at the receiver at slightly different times.
- These waves are called multipath waves and combine at the receiver antenna to give a resultant signal which can vary widely in amplitude and phase.

Small Scale Multipath Propagation



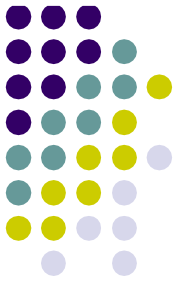
- Effects of multipath
  - Rapid changes in the signal strength
    - Over small travel distances, or
    - Over small time intervals
  - Random frequency modulation due to varying Doppler shifts on different multiples signals
  - Time dispersion (echoes) caused by multipath propagation delays
- Multipath occurs because of
  - Reflections
  - Scattering

Multipath

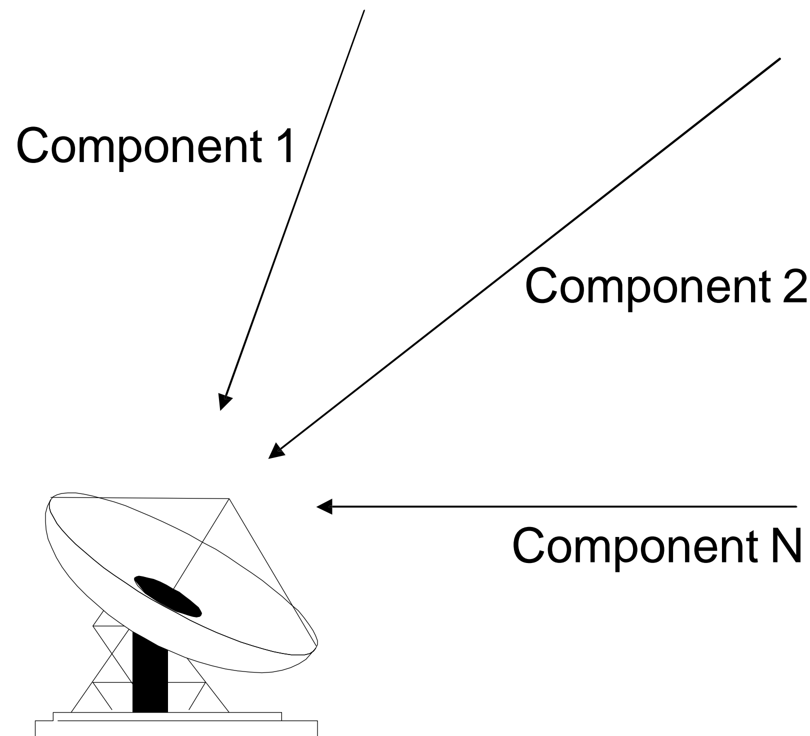


- At a receiver point
  - Radio waves generated from the same transmitted signal may come
    - from different directions
    - with different propagation delays
    - with (possibly) different amplitudes (random)
    - with (possibly) different phases (random)
    - with different angles of arrival (random).
  - These multipath components combine vectorially at the receiver antenna and cause the total signal
    - to fade
    - to distort

Multipath Components

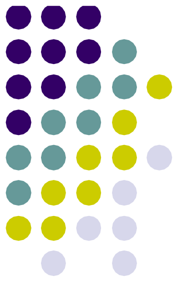


Radio Signals Arriving from different directions to receiver



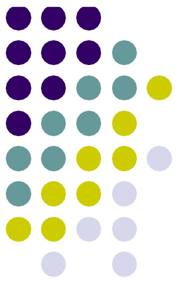
Receiver may be stationary or mobile.

Mobility



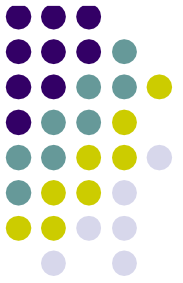
- Other Objects in the radio channels may be mobile or stationary
- If other objects are stationary
  - Motion is only due to mobile
  - Fading is purely a spatial phenomenon (occurs only when the mobile receiver moves)
  - The spatial variations as the mobile moves will be perceived as temporal variations
    - $\Delta t = \Delta d/v$
- Fading may cause disruptions in the communication

Factors Influencing Small Scale Fading



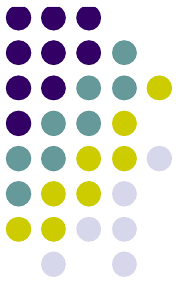
- Multipath propagation
  - Presence of reflecting objects and scatterers cause multiple versions of the signal to arrive at the receiver
    - With different amplitudes and time delays
    - Causes the total signal at receiver to fade or distort
- Speed of mobile
  - Cause Doppler shift at each multipath component
  - Causes random frequency modulation
- Speed of surrounding objects
  - Causes time-varying Doppler shift on the multipath components

Factors Influencing Small Scale Fading



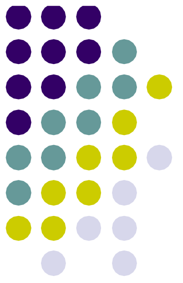
- Transmission bandwidth of the channel
  - The transmitted radio signal bandwidth and bandwidth of the multipath channel affect the received signal properties:
    - If amplitude fluctuates or not
    - If the signal is distorted or not

Doppler Effect



- Whe a transmitter or receiver is moving, the frequency of the received signal changes, i.e. It is different than the frequency of transmissin. This is called Doppler Effect.
- The change in frequency is called Doppler Shift.
  - It depends on
    - The relative velocity of the receiver with respect to transmitter
    - The frequenct (or wavelenth) of transmission
    - The direction of traveling with respect to the direction of the arriving signal.

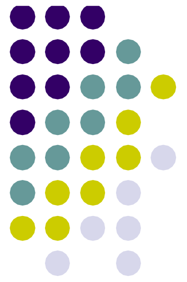
Doppler Shift – Transmitter is moving

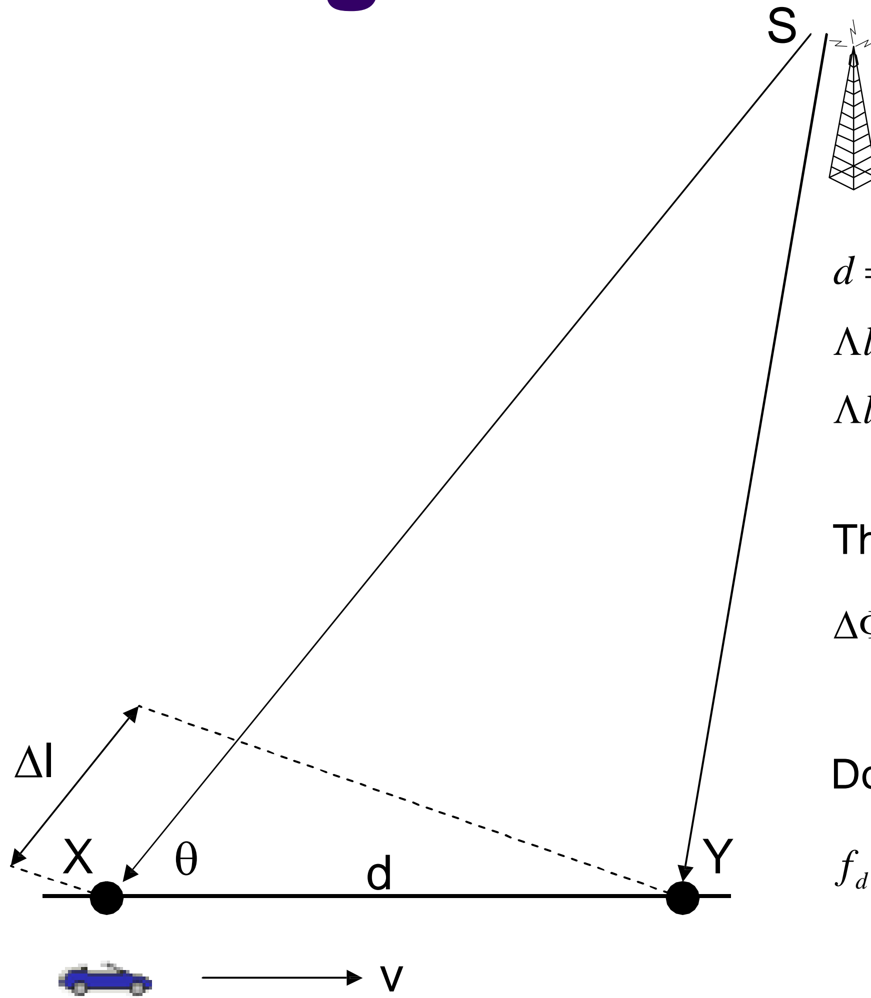


The frequency of the signal that is received behind the transmitter will be smaller

The frequency of the signal that is received in front of the transmitter will be bigger

Doppler Shift –Recever is moving





$$d = |XY|$$

$$\Lambda l = |SX| - |SY| = d \cos 0$$

$$\Lambda l = v \Lambda t \cos 0$$

The phase change in the received signal:

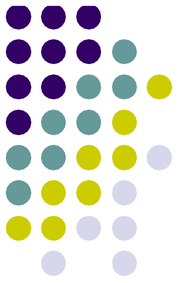
$$\Delta\Phi = \frac{\Lambda l}{Z} 2\nu = \frac{2\nu v \Lambda t}{Z} \cos 0$$

Doppler shift (The apparent change in frequency) :

$$f_d = \frac{1}{2\nu} \frac{\Delta\Phi}{\Delta t} = \frac{v}{Z} \cos 0$$

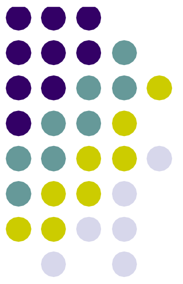
A mobile receiver is traveling from point X to point Y

Doppler Shift

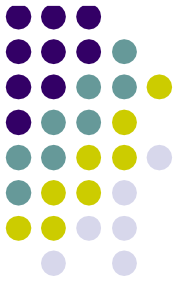


- The Dopper shift is positive
  - If the mobile is moving toward the direction of arrival of the wave.
- The Doppler shift is negative
  - If the mobile is moving away from the direction of arrival of the wave.

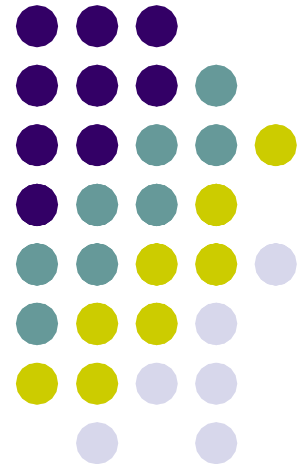
Impulse Response Model of a Multipath Channel



- The wireless channel charcteristics can be expressed by impulse response function
- The channel is time varying channel when the receiver is moving.
- Lets assume first that time variation due strictly to the receiver motion ($t = d/v$)
- Since at any distance $d = vt$, the received power will be combination of different incoming signals, the channel charactesitics or the impulse response funcion depends on the distance *d* between trandmitter and receiver.

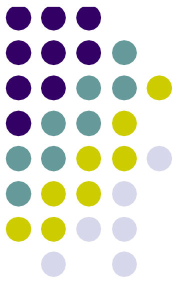


Multipath Channel Modeling



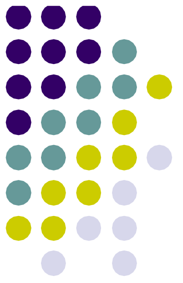
Impulse Response Model of a Multipath Wireless Channel

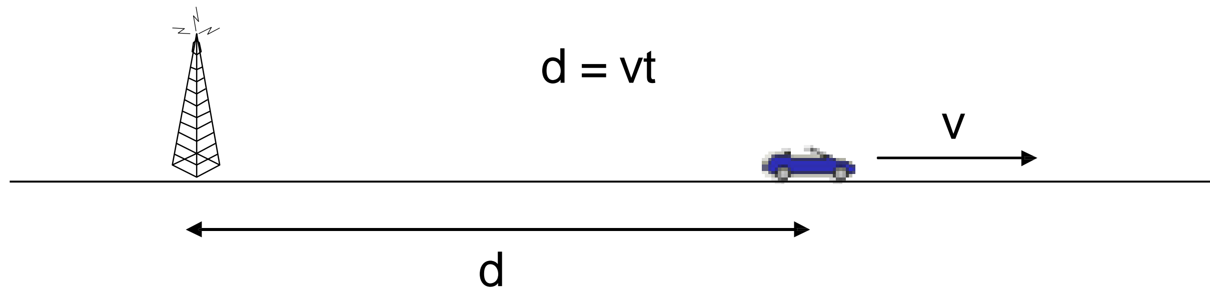
Impulse Response Model of a Multipath Channel



- The wireless channel characteristics can be expressed by impulse response function
- The channel is time varying channel when the receiver is moving.
- Lets assume first that time variation due strictly to the receiver motion (t = *d*/v)
- Since at any distance *d* = vt, the received power will ve combination of different incoming signals, the channel charactesitics or the impulse response funcion depends on the distance *d* between trandmitter and receiver

Impulse Response Model of a Multipath Channel

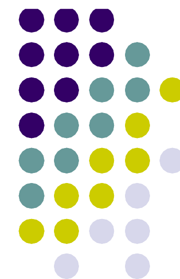




A receiver is moving along the ground at some constant velocity v.
The multipath components that are received at the receiver will have different propagation delays depending on d: distance between transmitter and receiver. Hence the channel impulse response depends on d.

Lets x(t) represents the transmitter signal
y(d,t) represents the received signal at position d.
h(d,t) represents the channel impulse response which is dependent on d (hence time-varying d=vt).

Multipath Channel Model



Building

Multipath
Channel

2nd MC

1st MC

Base
Station

Mobile 2

Building

Building

1st MC

4th MC

Multipath
Channel

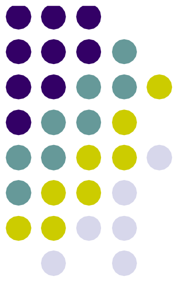
2nd MC

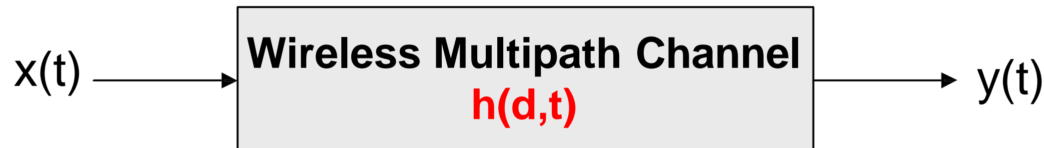
Building

3rd MC
(Multipath Component)

Mobile 1

Impulse Response Model of a Multipath Channel





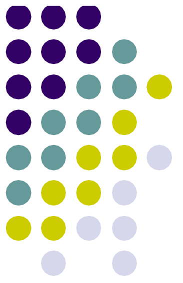
The channel is linear time-varying channel, where the channel characteristics changes with distance (hence time, t = d/v)

$$y(d,t) = x(t) \otimes h(d,t) = \int_{-\infty}^{\infty} x(\tau) h(d, t-\tau) d\tau$$

For a causal system, $h(d,t) = 0$ for $t < 0$; hence

$$y(d,t) = \int_{-\infty}^{t} x(\tau) h(d, t-\tau) d\tau$$

Impulse Response Model



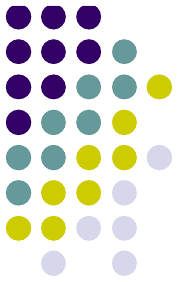
$d = vt$ assume v is constant over time

$$y(vt,t) = \int_{-\infty}^{t} x(\tau)h(vt,t-\tau)d\tau$$

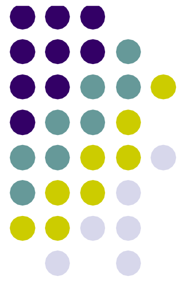
$$y(t) = \int_{-\infty}^{t} x(\tau)h(vt,t-\tau)d\tau = x(t)\otimes h(vt,t) = x(t)\otimes h(d,t)$$

We assume v is constant over short time.
x(t): transmitted waveform
y(t): received waveform
h(t,τ): impulse response of the channel. Depends on d (and therefore t=d/v) and also to the multiple delay for the channel for a fixed value of t. τ is the multipath delay of the channel for a fixed value of t.

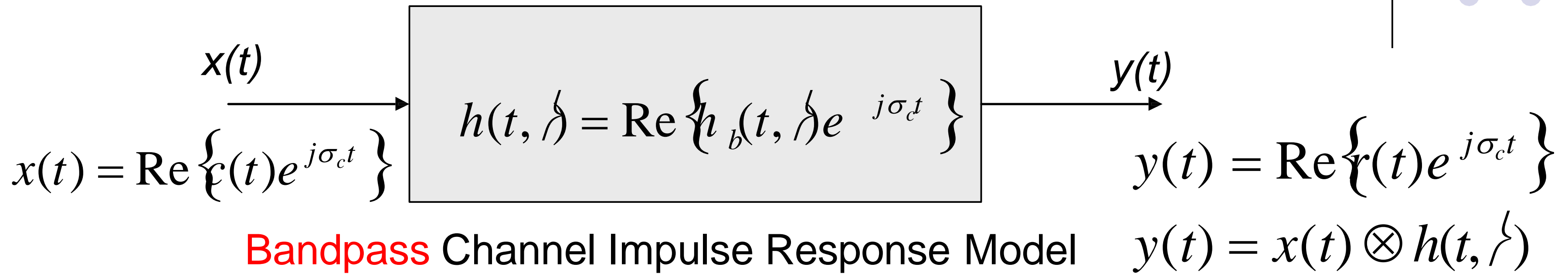
$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t,\tau)d\tau = x(t)\otimes h(t,\tau)$$

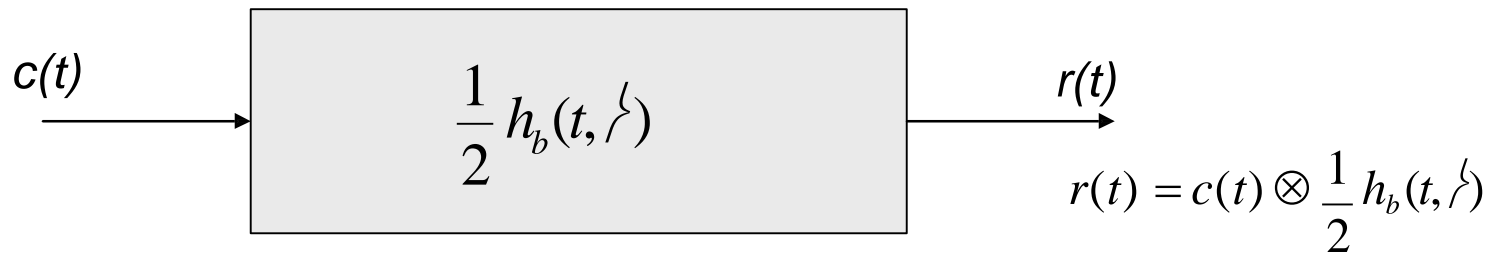


...Continue with Multipath Channel Impulse Response Model



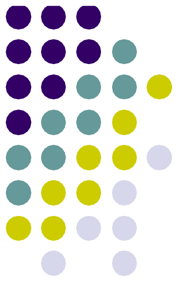
Impulse Response Model





Baseband Equivalent Channel Impulse Response Model

Impulse Response Model



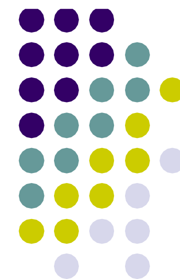
$$r(t) = c(t) \otimes \frac{1}{2} h_b(t, \tau)$$

$$x(t) = \mathrm{Re}\left\{ c(t) e^{j 2 \nu f_c t} \right\} \quad \sigma_c = 2 \nu f_c$$

$$y(t) = \mathrm{Re}\left\{ r(t) e^{j 2 \nu f_c t} \right\}$$

c(t) is the complex envelope representation of the transmitted signal
r(t) is the complex envelope representation of the received signal

$h_b(t,\tau)$ is the complex baseband impulse response



Discrete-time Impulse Response Model of Multipath Channel

Amplitude of Multipath Component

There are *N* multipath components (0..N-1)

Excess Delay Bin

$\tau_o = 0$

$\tau_1 = \Delta\tau$

$\tau_i = (i)\Delta\tau$

$\tau_{N-1} = (N-1)\Delta\tau$

τ (excess delay)

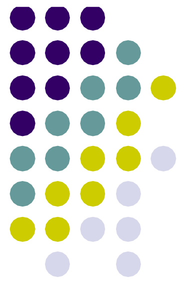
$\Delta\tau$

τ_0 τ_2 τ_i τ_{N-1}

Excess delay: relative delay of the ith multipath componentas compared to the first arriving component

τ_i : *Excesss delay* of ith multipath component, **N6**}: *Maximum excess delay*

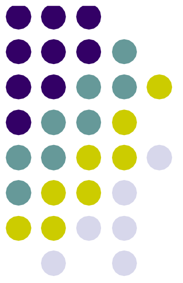
Multipath Components arriving to a Receiver



Ignore the fact that multipath components arrive with different angles, and assume that they arriving with the same angle in 3D.

Each component will have different Amplitude (a_i) and Phase (θ_i)

Baseband impulse response of the Channel



$$h_b(t,\tau) = \sum_{i=0}^{N-1} a_i(t,\tau) e^{j(2\pi f_c \tau_i(t) + \phi_i(t,\tau))} \delta(\tau - \tau_i(t))$$

$a_i(t,\tau)$: the real amplitude of the ith multipath component at time t.

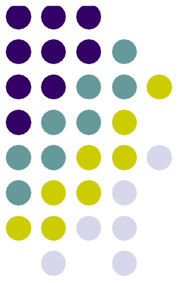
$\tau_i(t)$: excess delay of the ith multipath component at time t.

$2\pi f_c \tau_i(t) + \phi_i(t,\tau)$: Phase term that represents phase shift due to

free space propagatio n of the ith component. Simply represent it with : $\theta_i(t,\tau)$

$\delta(\bullet)$:unit impulse function.

Discrete-Time Impulse Response Model for a Multipath Channel



$h_b(t,\tau)$

t

t_3 t_2 t_1 t_0

$\tau(t_3)$

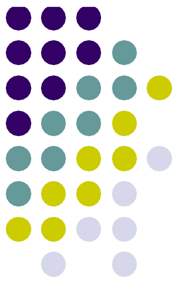
$\tau(t_2)$

$\tau(t_1)$

$\tau(t_0)$

τ_o τ_1 τ_2 τ_3 τ_4 τ_5 τ_6 τ_{N-2} τ_{N-1}

Time-Invariance Assumption



If the channel impulse response is assumed to be time-invariant over small-scale time or distance interval, then the channel impulse response can be simplified as:

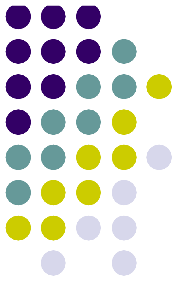
$$h_b(\tau) = \sum_{i=0}^{N-1} a_i e^{j\theta_i} \delta(\tau - \tau_i)$$

When measuring or predicting $h_b(\tau)$, a probing pulse p(t) which approximates the unit impulse function is used at the transmitter. That is:

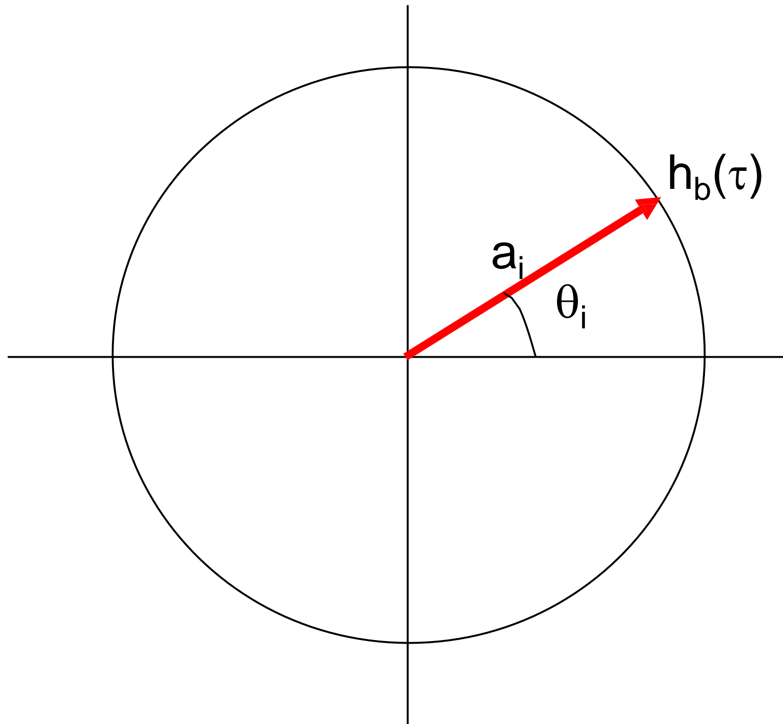
$$p(t) \approx \delta(t - \tau)$$

This is called sounding the channel to determine impulse response.

Complex Baseband Impulse Response



Baseband impulse response $h_b(\tau)$ is a complex number and therefore has a magnitude (amplitude) a_i and a phase θ_i.



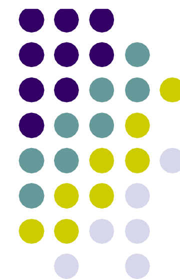
$$h_b(\tau) = a_i e^{j\theta_i}$$

$$h_b(\tau) = a_i(\cos\theta_i + j\sin\theta_i)$$

$$|h_b(\tau)| = a_i$$

you can think of it also as a vector that starts at origin.

Amplitudes and Phases of Multipath Components



1st Arriving Multipath Component
(Say 0th Component)

f_c

$\theta_0 = 0$ (phase)

$2a_0$

Two components emerge from the same source at the same time. They belong to the same transmitter signal. But they travel different paths. They arrive at the same receiver with time difference equal to τ_i.

f_c

$2a_i$

ith Multipath Component

$\theta_i = 2\pi f_c \tau_i$

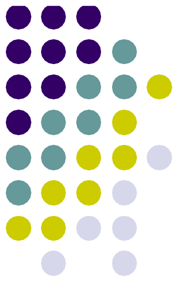
τ

0

τ_i

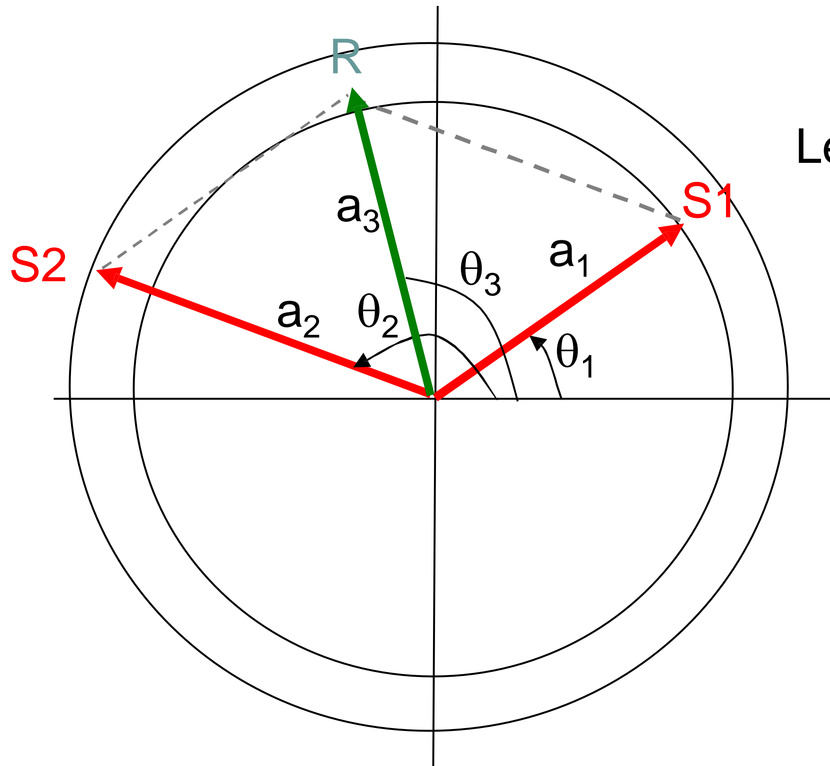
θ_i is expressed in radians

Components arriving at the same time



What happens if two or more multipath components are with the same access delay bin (arrive at the same time)?

Then the received signal is the vectorial addition of two multipath signals.



Example:

Lets assume two signals S1 and S2 arrive at the same time at the receiver:

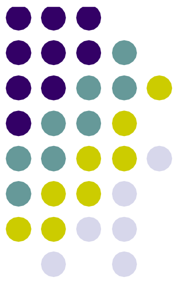
$$S_1 = a_1 e^{j\theta_1} \qquad S_2 = a_2 e^{j\theta_2}$$

$$R = S_1 + S_2$$

$$= a_1 e^{j\theta_1} + a_2 e^{j\theta_2} = a_3 e^{j\theta 3}$$

R is the combined receiver signal.

Components arriving at the same time

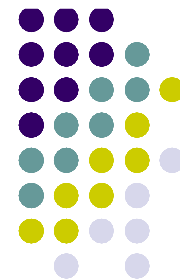


The amplitude and phase of the combined signal (R) depends on the amplitudes and phases of the two components.

Depending on the values of the phases of the components, the combined affect may weaken or strengthen the amplitude of the combined signal.

It is possible that the two signals may totally cancel each other depending on their relative phases on their amplitudes.

Example 1 – Addition of Two Signals



MC: Multipath Component

1st MC

2st MC

Combined Signal

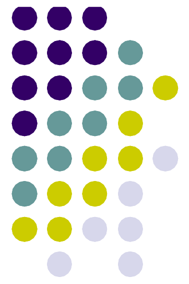
$a_1/a_2=1$

$\theta_1=\pi/16$

$\theta_2=\pi$

cos(x+pi/16)
cos(x+pi)
cos(x+pi/16)+cos(x+pi)

Example 2 – Addition of Two Signals



1st MC

2st MC

Combined Signal

$a_1/a_2=1/3$

$\theta_1=\pi/16$

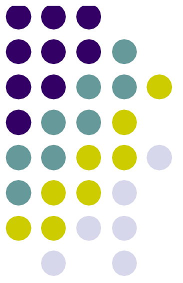
$\theta_2=\pi$

cos(x+pi/16)
3*cos(x+pi)
cos(x+pi/16)+3*cos(x+pi)

3
2
1
0
-1
-2
-3

-10 -5 0 5 10

Power Delay Profile



For small-scale fading, the power delay profile of the channel is found by taking the spatial average of $|h_b(t;\tau)|^2$ over a local area (small-scale area).

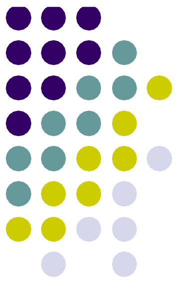
If p(t) has a time duration much smaller than the impulse response of the multipath channel, the received power delay profile in a local area is given by:

$$P(\tau) \approx k\overline{|h_b(t;\tau)|^2}$$

The bar represents the average over the local area of $|h_b(t;\tau)|^2$

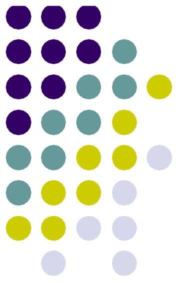
Gain k relates the transmitter power in the probing pulse *p(t)* to the total received power in a multipath delay profile.

Example power delay profile

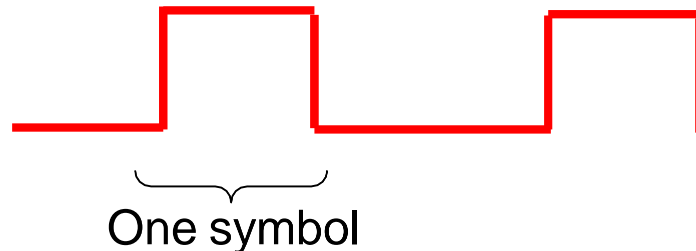


Taken from Dimitrios Mavrakis Homepage:http://www.ee.surrey.ac.uk/Personal/D.Mavrakis/

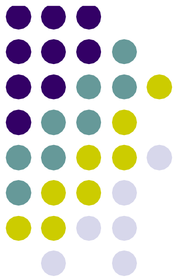
Relationship between Bandwidth and Receiver Power



- What happens when two different signals with different bandwidths are sent through the channel?
    - What is the receiver power characteristics for both signals?
- We mean the bandwith of the baseband signal
    - The bandwidth of the baseband is signal is inversely related with its symbol rate.



Bandwidth of Baseband Signals



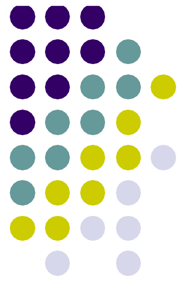
Highbandwidth (Wideband) Signal

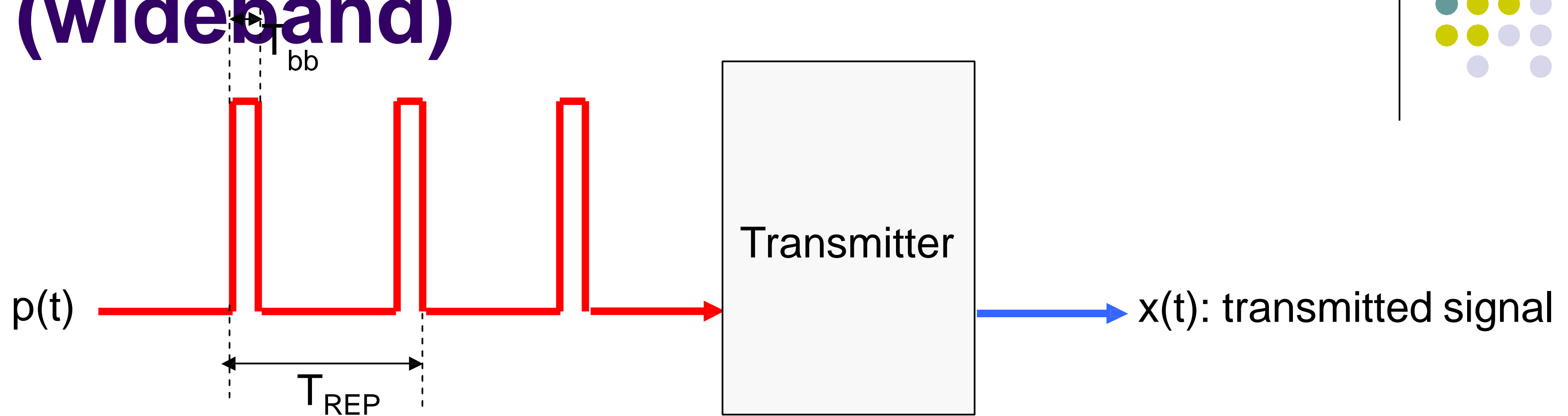
Lowbandwidth (Narrowband) Signal

Continuous Wave (CW) Signal

t

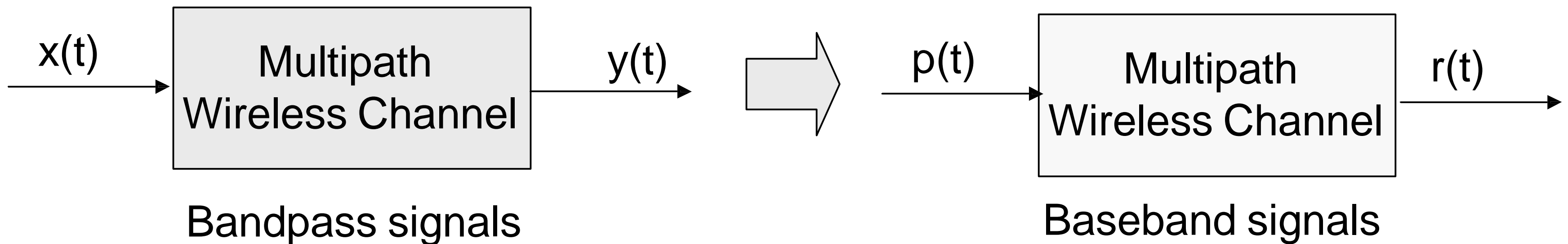
A pulsed probing signal (wideband)



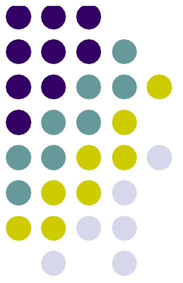


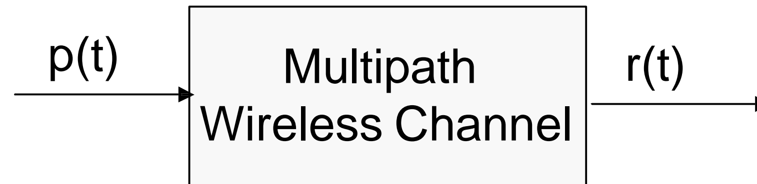
$$T_{REP} >> \tau_{max} \qquad (\tau_{max} : \text{maximum measured excess delay})$$

$$x(t) = \mathrm{Re}\{p(t)e^{j2\pi f_c t}\} = p(t)\cos(2\pi f_c t)$$



Received Power of Wideband Signals





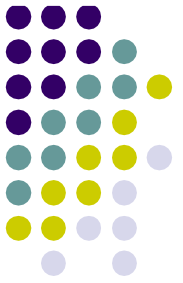
The output r(t) will approximate the channel impulse response since p(t) approximates unit impulses.

$$r(t) = \frac{1}{2} \sum_{i=0}^{N-1} a_i e^{j\theta_i} \cdot p(t - \tau_i)$$

Assume the multipath components have random amplitudes and phases at time t.

$$E_{a,\theta}[P_{WB}] = E_{a,\theta}\left[\left| \sum_{i=0}^{N-1} a_i e^{j\theta_i} \right|^2 \right] = \sum_{i=0}^{N-1} \overline{a_i^{\,2}} = E[P_{WB}]$$

Received Power of Wideband Signals



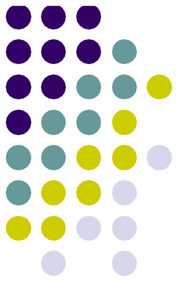
This shows that if all the multipath components of a transmitted signal is resolved at the receiver then:

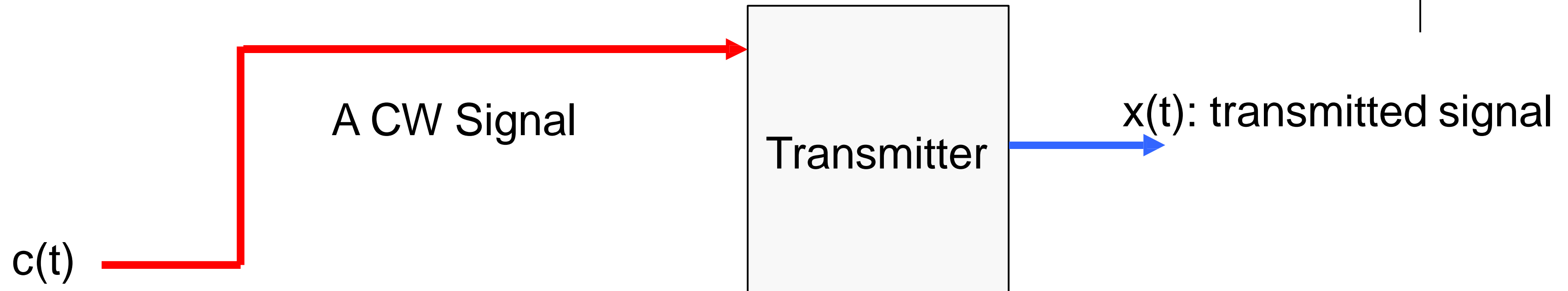
The average small scale received power is simply the sum of received powers in each multipath component.

In practice, the amplitudes of individual multipath components do not fluctuate widely in a local area (for distance in the order of wavelength or fraction of wavelength).

This means the average received power of a wideband signal do not fluctuate significantly when the receiver is moving in a local area.

Received Power of Narrowband Signals





Assume now A CW signal transmitted into the same channel.

Let comlex envelope will be: $c(t) = 2$

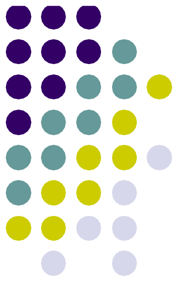
The instantaneous complex envelope of the received signal will be:

$$r(t) = \sum_{i=0}^{N-1} a_i e^{j\theta_i(t,\tau)}$$

The instantaneous power will be:

$$|r(t)|^2 = \left| \sum_{i=0}^{N-1} a_i e^{j\theta_i(t,\tau)} \right|^2$$

Received Power of Narrowband Signals



Over a local area (over small distance – wavelengths), the amplitude a multipath component may not change signicantly, but the phase may change a lot.

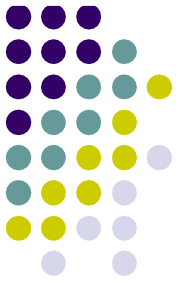
For example:

- if receiver moves λ meters then phase change is 2π.
  In this case the component may add up posively to the total sum Σ.

- if receiver moves $\lambda/4$ meters then phase change is $\pi/2$ (90 degrees) .
  In this case the component may add up negatively to the total sum Σ, hence the instantaneous receiver power.

Therefore for a CW (continues wave, narrowband) signal, the small movements may cause large fluctuations on the instantenous receiver power, which typifies small scale fading for CW signals.

Wideband versus Narrowband Baseband Signals



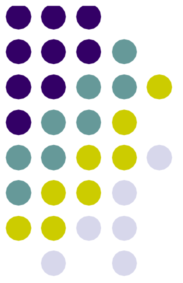
However, the average received power for a CW signal over a local area is equivalent to the average received power for a wideband signal on the local area.

This occurs because the phases of multipath components at different locations over the small-scale region are independently distributed (IID uniform) over $[0,2\pi]$.

In summary:

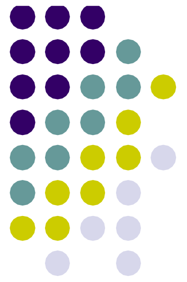
1. Received power for CW signals undergoes rapid fades over small distances
2. Received power for wideband signals changes very little of small distances.
3. However, the local area average of both signals are nearly identical.

Small-Scale Multipath Measurements



- Several Methods
  - Direct RF Pulse System
  - Spread Spectrum Sliding Correlator Channel Sounding
  - Frequency Domain Channel Sounding
- These techniques are also called channel sounding techniques

Direct RF Pulse System



Tx

f_c

Pulse Generator

RF Link

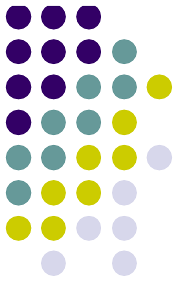
Rx

BPF

Detector

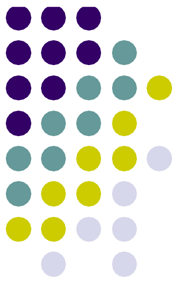
Digital Oscilloscope

Parameters of Mobile Multipath Channels

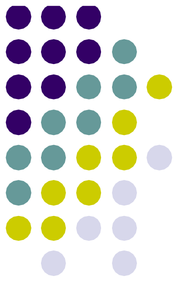


- Time Dispersion Parameters
  - Grossly quantifies the multipath channel
  - Determined from Power Delay Profile
  - Parameters include
    - Mean Access Delay
    - RMS Delay Spread
    - Excess Delay Spread (X dB)
- Coherence Bandwidth
- Doppler Spread and Coherence Time

Measuring PDPs



- Power Delay Profiles
  - Are measured by channel sounding techniques
  - Plots of relative received power as a function of excess delay
  - They are found by averaging *intantenous* power delay measurements over a local area
    - Local area: no greater than 6m outdoor
    - Local area: no greater than 2m indoor
      - Samples taken at $\lambda/4$ meters approximately
      - For 450MHz – 6 GHz frequency range.



Timer Dispersion Parameters

Determined from a power delay profile.

Mean excess delay($\bar{\tau}$):

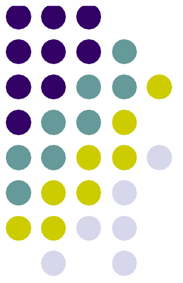
$$\bar{\tau} = \frac{\sum_k a_k^2 \tau_k}{\sum_k a_k^2} = \frac{\sum_k P(\tau_k)(\tau_k)}{\sum_k P(\tau_k)}$$

Rms delay spread (σ_τ):

$$\sigma_\tau = \sqrt{\overline{\tau^2} - (\bar{\tau})^2}$$

$$\overline{\tau^2} = \frac{\sum_k a_k^2 \tau_k^2}{\sum_k a_k^2} = \frac{\sum_k P(\tau_k)(\tau_k^2)}{\sum_k P(\tau_k)}$$

Timer Dispersion Parameters

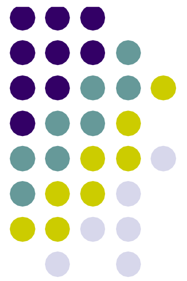


Maximum Excess Delay (X dB):

Defined as the time delay value after which the multipath energy falls to X dB below the maximum multipath energy (not necesarily belonging to the first arriving component).

It is also called *excess delay spread*.

RMS Delay Spread



RMS Delay Spread = 46.40 ns

Maximum Excess Delay < 10 dB = 84 ns

Threshold Level = -20 dB

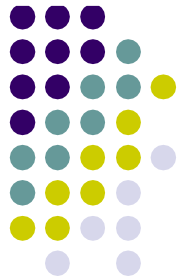
Mean Excess Delay = 45.05 ns

Normalized Received Power (dB Scale)

10
0
−10
−20
−30

−50 0 50 100 150 200 250 300 350 400 450

Excess Delay (ns)



PDP Outdoor

Display Threshold = -111.5 dBm per 40 ns
RMS Delay Spread = 22.85 microseconds
Worst Case San Francisco measurement
Mensell St.

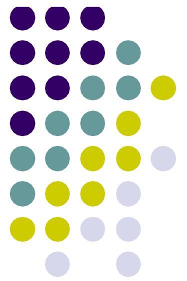
Received Signal Level, [dBm per 40 ns]

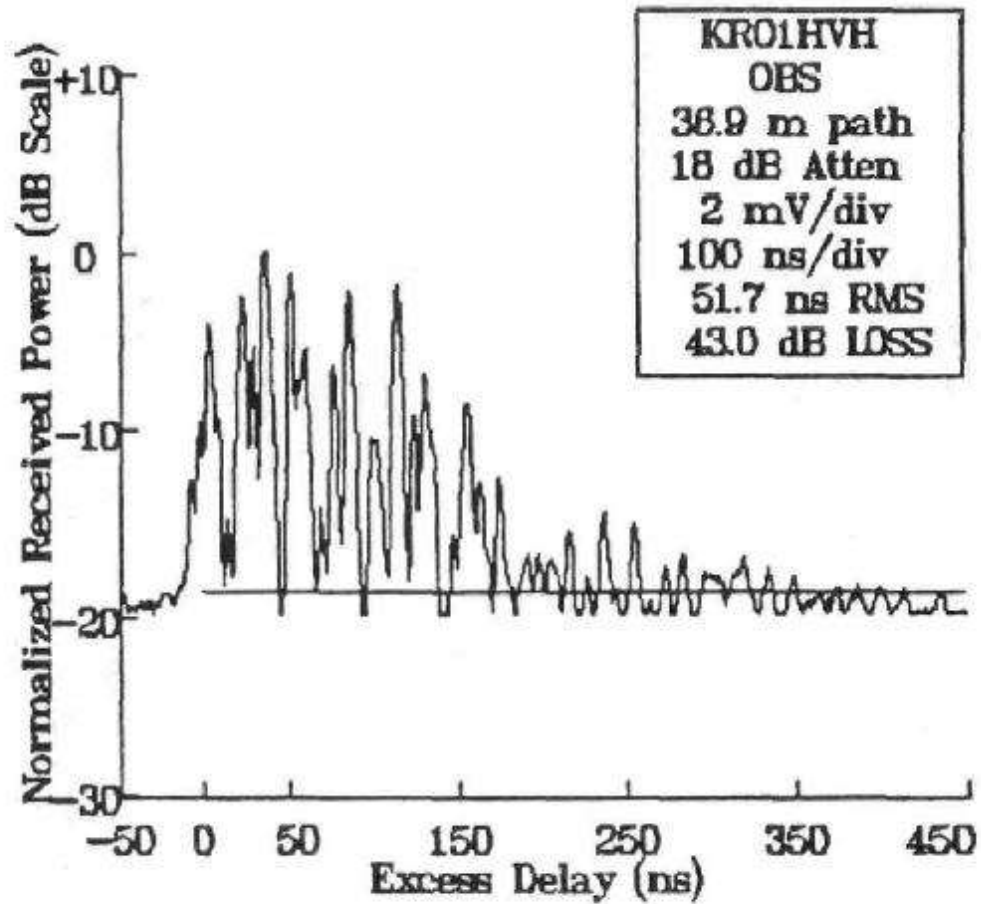
-85
-90
-95
-100
-105
-110
-115

0
20
40
60
80
100

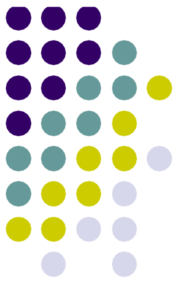
Excess Delay Time, [microseconds]

PDP Indoor



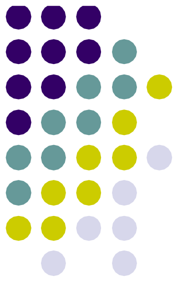


Noise Threshold

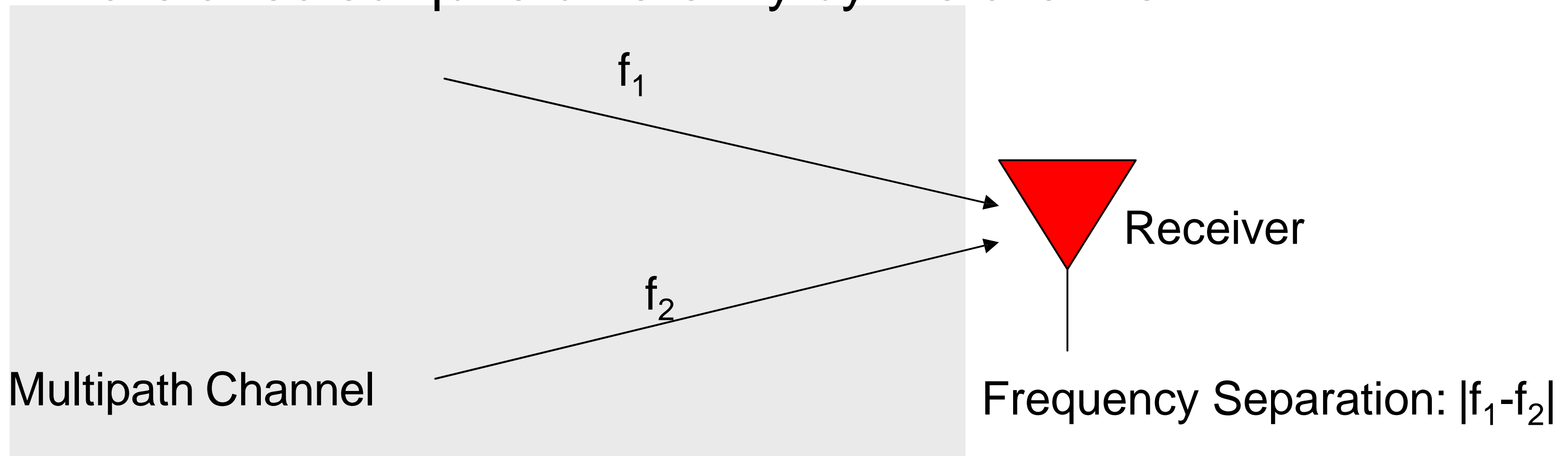


- The values of time dispersion parameters also depend on the noise threshold (the level of power below which the signal is considered as noise).
- If noise threshold is set too low, then the noise will be processed as multipath and thus causing the parameters to be higher.

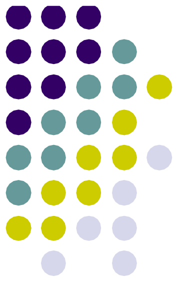
Coherence Bandwidth (B_C)



- Range of frequencies over which the channel can be considered flat (i.e. channel passes all spectral components with equal gain and linear phase).
  - It is a definition that depends on RMS Delay Spread.
- Two sinusoids with frequency separation greater than B_c are affected quite differently by the channel.



Coherence Bandwidth



Frequency correlation between two sinusoids: $0 <= C_{r1, r2} <= 1$.

If we define Coherence Bandwidth (B_C) as the range of frequencies over which the frequency correlation is above 0.9, then

$$B_C = \frac{1}{50\sigma}$$

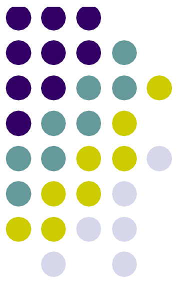
σ is rms delay spread.

If we define Coherence Bandwidth as the range of frequencies over which the frequency correlation is above 0.5, then

$$B_C = \frac{1}{5\sigma}$$

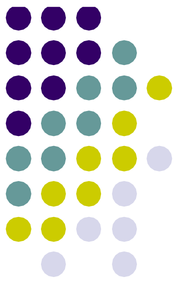
This is called 50% coherence bandwidth.

Coherence Bandwidth



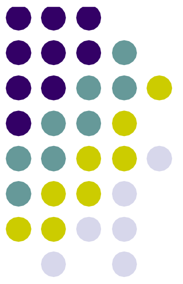
- Example:
  - For a multipath channel, σ is given as 1.37μs.
  - The 50% coherence bandwidth is given as: $1/5\sigma$ = 146kHz.
    - This means that, for a good transmission from a transmitter to a receiver, the range of transmission frequency (channel bandwidth) should not exceed 146kHz, so that all frequencies in this band experience the same channel characteristics.
    - Equalizers are needed in order to use transmission frequencies that are separated larger than this value.
    - This coherence bandwidth is enough for an AMPS channel (30kHz band needed for a channel), but is not enough for a GSM channel (200kHz needed per channel).

Coherence Time



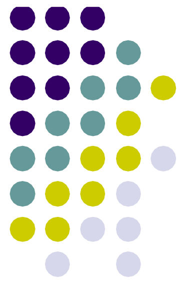
- Delay spread and Coherence bandwidth describe the time dispersive nature of the channel in a local area.
  - They don’t offer information about the time varying nature of the channel caused by relative motion of transmitter and receiver.
- Doppler Spread and Coherence time are parameters which describe the time varying nature of the channel in a small-scale region.

Doppler Spread



- Measure of spectral broadening caused by motion
- We know how to compute Doppler shift: f_d
- Doppler spread, B_D, is defined as the maximum Doppler shift: $f_m = v/\lambda$
- If the baseband signal bandwidth is much greater than B_D then effect of Doppler spread is negligible at the receiver.

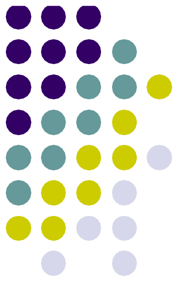
Coherence Time



Coherence time is the time duration over which the channel impulse response is essentially invariant.

If the symbol period of the baseband signal (reciprocal of the baseband signal bandwidth) is greater the coherence time, than the signal will distort, since channel will change during the transmission of the signal

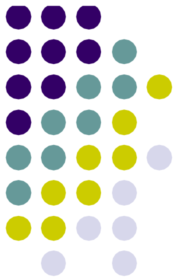
Coherence Time



Coherence time is also defined as: $$T_C \approx \sqrt{\frac{9}{16 \pi f_m^2}} = \frac{0.423}{f_m}$$

Coherence time definition implies that two signals arriving with a time separation greater than T_C are affected differently by the channel.

Types of Small-scale Fading



Small-scale Fading
(Based on Multipath Time Delay Spread)

Flat Fading

1. BW Signal < BW of Channel
2. Delay Spread < Symbol Period

Frequency Selective Fading

1. BW Signal > Bw of Channel
2. Delay Spread > Symbol Period

Small-scale Fading
(Based on Doppler Spread)

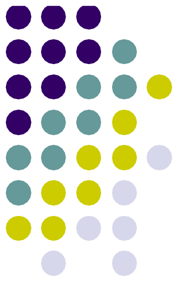
Fast Fading

1. High Doppler Spread
2. Coherence Time < Symbol Period
3. Channel variations faster than baseband signal variations

Slow Fading

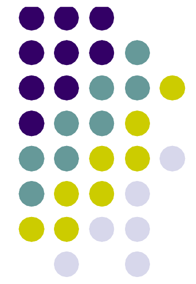
1. Low Doppler Spread
2. Coherence Time > Symbol Period
3. Channel variations smaller than baseband signal variations

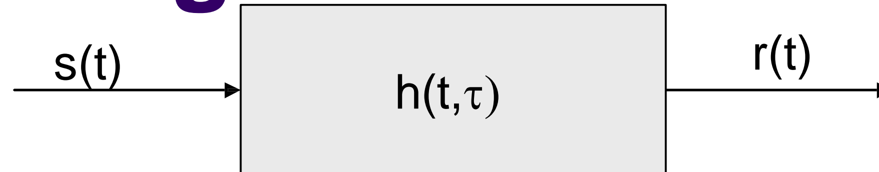
Flat Fading



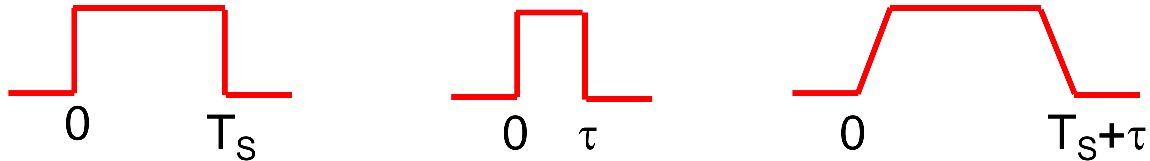
- Occurs when the amplitude of the received signal changes with time
  - For example according to Rayleigh Distribution
- Occurs when symbol period of the transmitted signal is much larger than the Delay Spread of the channel
  - Bandwidth of the applied signal is narrow.
- May cause deep fades.
  - Increase the transmit power to combat this situation.

Flat Fading





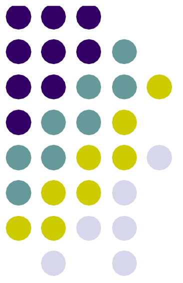
$\tau << T_S$



Occurs when:
$B_S << B_C$
and
$T_S >> \sigma_\tau$

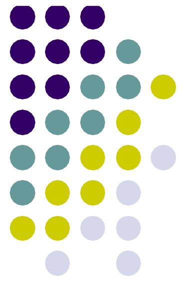
B_C: Coherence bandwidth
B_S: Signal bandwidth
T_S: Symbol period
σ_τ: Delay Spread

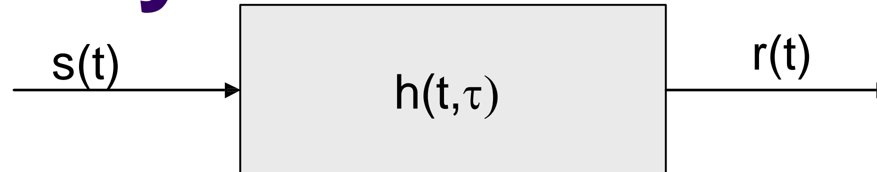
Frequency Selective Fading



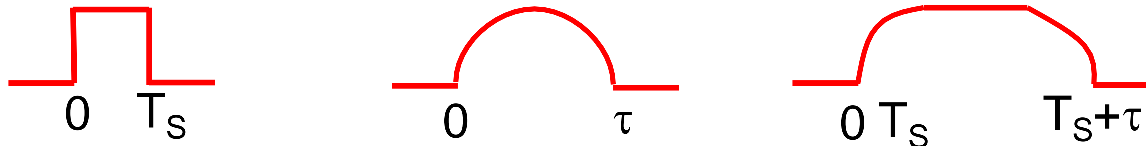
- Occurs when channel multipath delay spread is greater than the symbol period.
  - Symbols face time dispersion
  - Channel induces Intersymbol Interference (ISI)
- Bandwidth of the signal s(t) is wider than the channel impulse response.

Frequency Selective Fading





$\tau >> T_S$



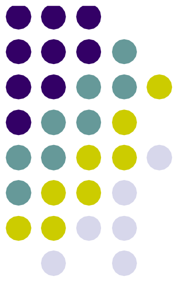
Causes distortion of the received baseband signal

Causes Inter-Symbol Interference (ISI)

Occurs when:
$B_S > B_C$
and
$T_S < \sigma_\tau$

As a rule of thumb: $T_S < \sigma_\tau$

Fast Fading

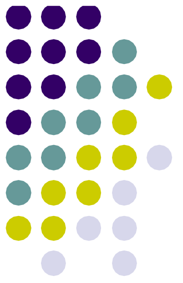


- Due to Doppler Spread
  - Rate of change of the channel characteristics is **larger** than the Rate of change of the transmitted signal
  - The channel changes during a symbol period.
  - The channel changes because of receiver motion.
  - Coherence time of the channel is smaller than the symbol period of the transmitter signal

Occurs when:
$B_S < B_D$
and
$T_S > T_C$

B_S: Bandwidth of the signal
B_D: Doppler Spread
T_S: Symbol Period
T_C: Coherence Bandwidth

Slow Fading

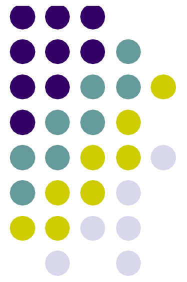


- Due to Doppler Spread
  - Rate of change of the channel characteristics is **much smaller** than the

Rate of change of the transmitted signal

Occurs when:
$B_S >> B_D$
and
$T_S << T_C$

B_S: Bandwidth of the signal
B_D: Doppler Spread
T_S: Symbol Period
T_C: Coherence Bandwidth



Different Types of Fading

T_S

Symbol Period of Transmitting Signal

Flat Slow Fading

Flat Fast Fading

σ_τ

Frequency Selective Slow Fading

Frequency Selective Fast Fading

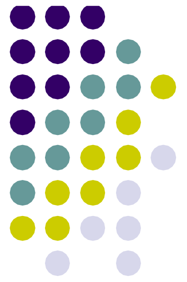
T_C

T_S

Transmitted Symbol Period

With Respect To SYMBOL PERIOD

Different Types of Fading



B_S

Frequency Selective Fast Fading

Frequency Selective Slow Fading

Transmitted Baseband Signal Bandwidth

B_C

Flat Fast Fading

Flat Slow Fading

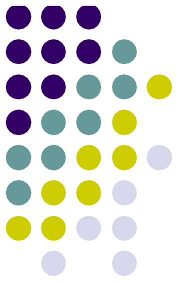
B_D

B_S

Transmitted Baseband Signal Bandwidth

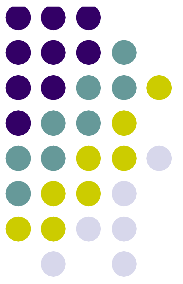
With Respect To BASEBAND SIGNAL BANDWIDTH

Fading Distributions

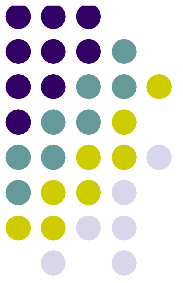


- Describes how the received signal amplitude changes with time.
  - Remember that the received signal is combination of multiple signals arriving from different directions, phases and amplitudes.
  - With the received signal we mean the baseband signal, namely the envelope of the received signal (i.e. r(t)).
- Its is a statistical characterization of the multipath fading.
- Two distributions
  - Rayleigh Fading
  - Ricean Fading

Rayleigh and Ricean Distributions



- Describes the received signal envelope distribution for channels, where all the components are non-LOS:
  - i.e. there is no line-of–sight (LOS) component.
- Describes the received signal envelope distribution for channels where one of the multipath components is LOS component.
  - i.e. there is one LOS component.

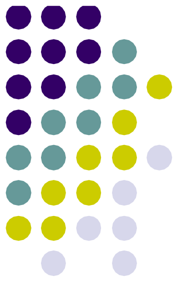


Rayleigh Fading

Amplitude in dB

Time

Rayleigh



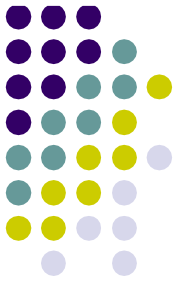
Rayleigh distribution has the probability density function (PDF) given by:

$$p(r) = \begin{cases} \dfrac{r}{\sigma^2} e^{\left(-\frac{r^2}{2\sigma^2}\right)} & (0 \le r \le \infty) \\ 0 & (r < 0) \end{cases}$$

σ^2 is the time average power of the received signal before envelope detection.
σ is the rms value of the received voltage signal before envelope detection

Remember: $\overline{P}$ (average power) $\propto V_{rms}^2$ (see end of slides 5)

Rayleigh



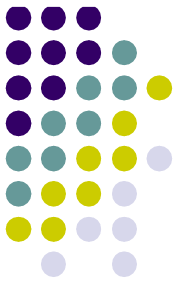
The probability that the envelope of the received signal does not exceed a specified value of R is given by the CDF:

$$P(R) = P_r(r \le R) = \int_0^R p(r)dr = 1 - e^{-\frac{R^2}{2o^2}}$$

$$r_{mean} = E[r] = \int_0^\infty rp(r)dr = o\sqrt{\frac{\nu}{2}} = 1.2533\,o$$

$r_{median} = 1.177\,o$ found by solving $\frac{1}{2} = \int_0^{r_{median}} p(r)dr$

$$r_{rms} = \sqrt{2}o$$



Rayleigh PDF

0.6065/σ

mean = 1.2533σ

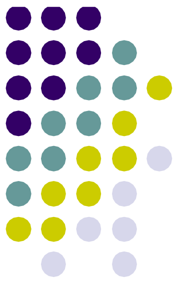
median = 1.177σ

variance = 0.4292σ^2

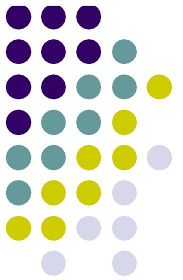
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

0
σ
2σ
3σ
4σ
5σ

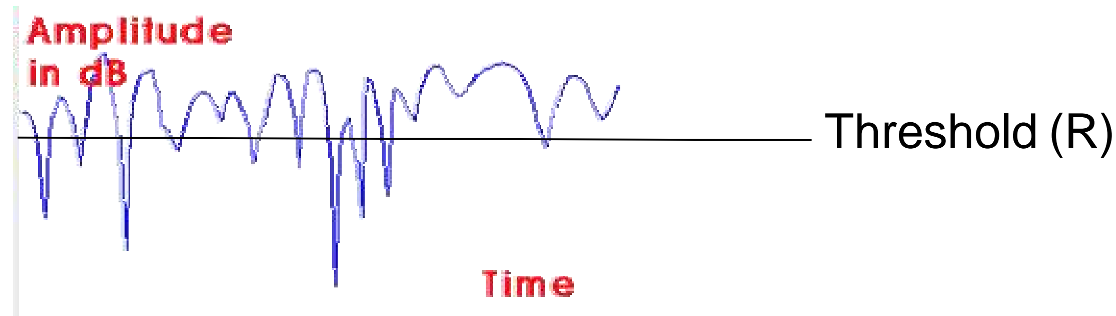
Ricean Distribution



- When there is a stationary (non-fading) LOS signal present, then the envelope distribution is Ricean.
- The Ricean distribution degenerates to Rayleigh when the dominant component fades away.



Level Crossing Rate (LCR)



LCR is defined as the expected rate at which the Rayleigh fading envelope, normalized to the local rms signal level, crosses a specified threshold level R in a positive going direction. It is given by:

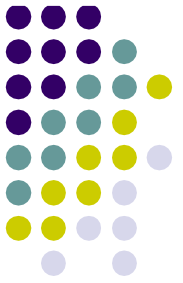
$$N_R = \sqrt{2\nu} f_m \, \theta e^{-\theta^2}$$

where

$\theta = R / r_{rms}$ (specfied envelope value normalized to rms)

N_R : crossings per second

Average Fade Duration



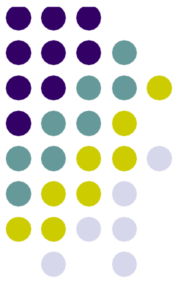
Defined as the average period of time for which the received signal is below a specified level R.

For Rayleigh distributed fading signal, it is given by:

$$\bar{\tau} = \frac{1}{N_R} \Pr[r \le R] = \frac{1}{N_R}\left(1 - e^{-\theta^2}\right)$$

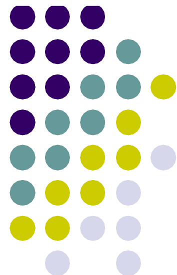
$$\bar{\tau} = \frac{e^{\theta^2} - 1}{\theta f_m \sqrt{2\nu}}, \qquad \theta = \frac{R}{r_{rms}}$$

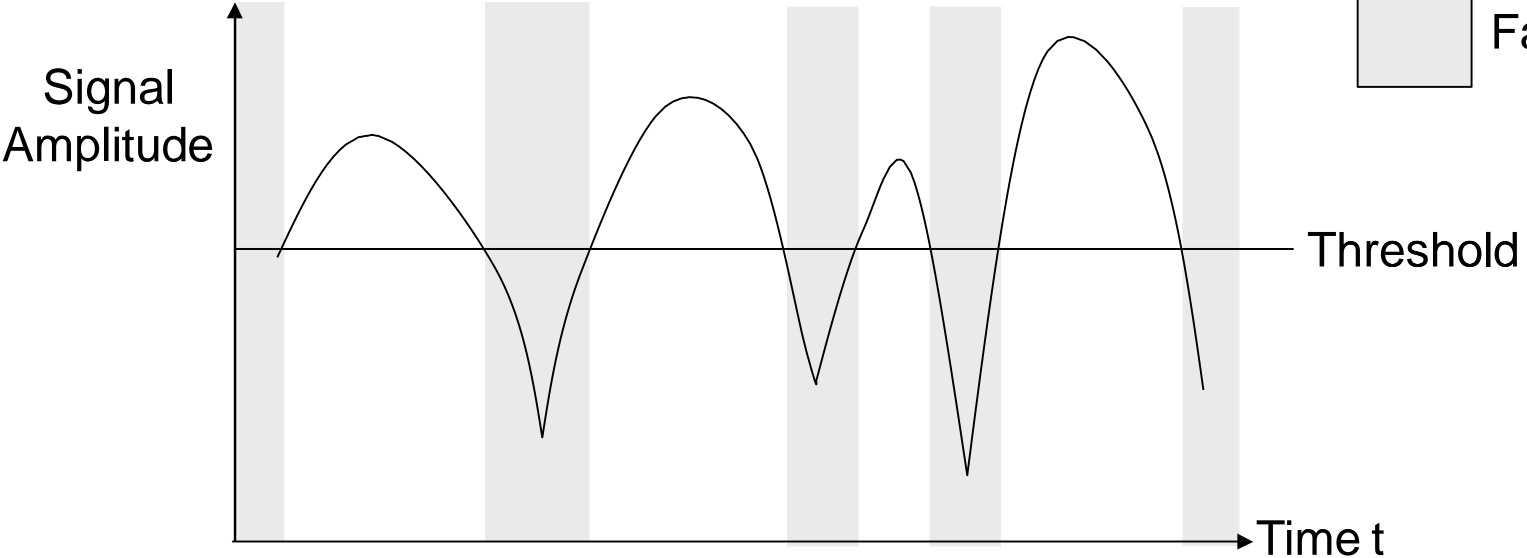
Statistical Models for Multipath Fading Channels

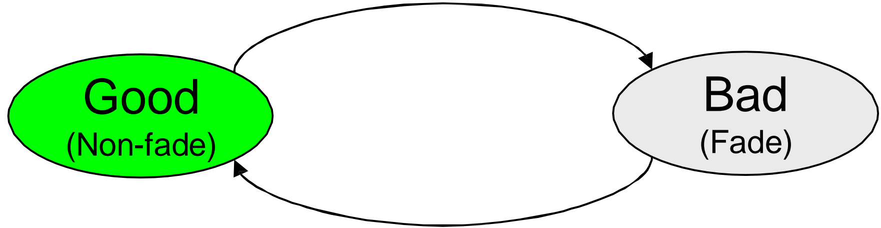


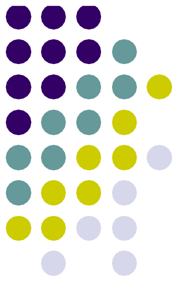
- Clarke's Model for Flat Fading
- Two-ray Rayleigh Fading Model
- Saleh and Valenzuela Indoor Statistical Model
- SIRCIM and SMRCIM Indoor and Outdoor Statistical Models

Fading Model – Gilbert-Elliot Model

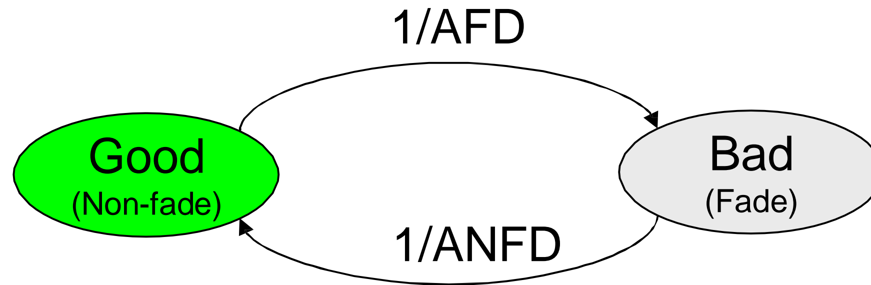






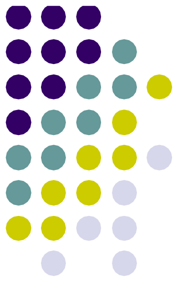


Gilbert-Elliot Model



The channel is modeled as a Two-State Markov Chain.
Each state duration is memory-less and exponentially distributed.

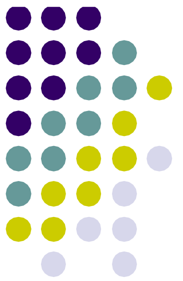
The rate going from Good to Bad state is: 1/AFD (AFD: Avg Fade Duration)
The rate going from Bad to Good state is: 1/ANFD (ANFD: Avg Non-Fade Duration)



UNIT-IV

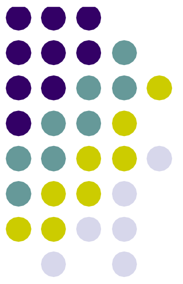
Equalization and Diversity

Outline



- Introduction
- Fundamentals of Equalization
- Survey of Equalization Techniques
  - Linear Equalizers
  - Nonlinear Equalization
  - Algorithms for Adaptive Equalization
- Fundamentals of diversity
- Survey of Diversity Techniques
  - Frequency/Time/Space/Polarization Diversity
  - Selection/MRC/EGC Combining
  - RAKE Receiver
  - Interleaving

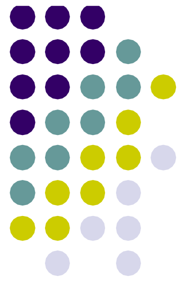
4.1 Introduction



- The properties of mobile radio channels:
  - Multipath fading -> time dispersion, ISI
  - Doppler spread -> dynamical fluctuation

  These effects have a strong negative impact on the bit error rate of any modulation.
- Mobile communication systems require signal processing techniques that improve the link performance in hostile mobile radio environments.
- Three popular techniques:
  - Equalization: compensates for ISI
  - Diversity: compensates for channel fading
  - Channel coding: detects or corrects errors

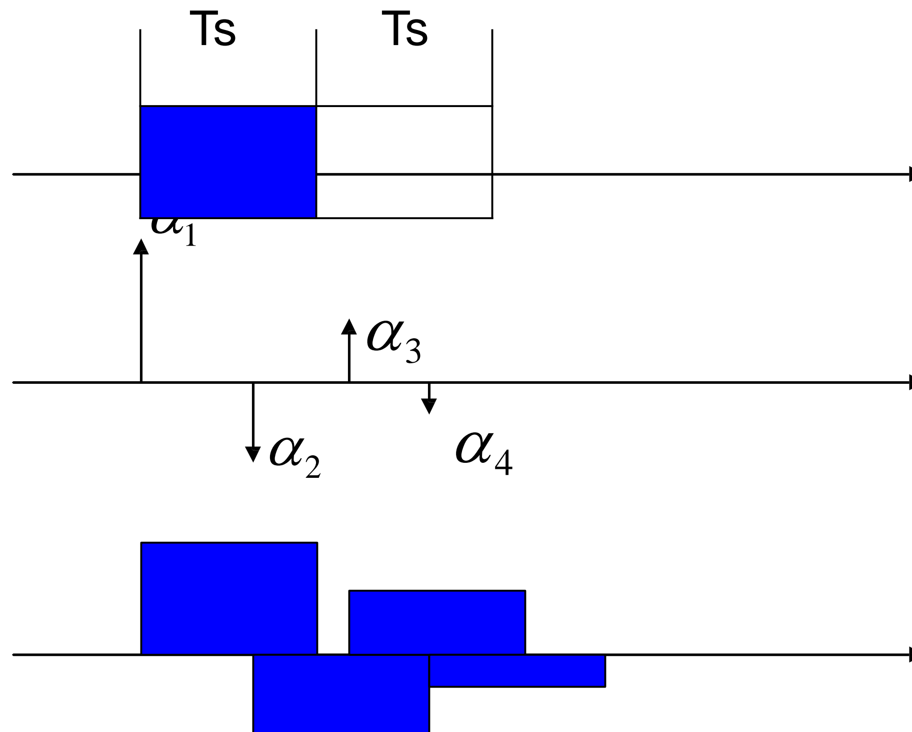
  These techniques can be deployed independently or jointly.

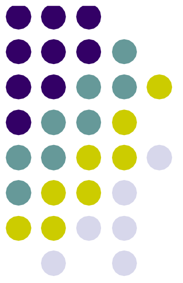


Transmitted signal: $s(t)$

Channel model: $h(t) = \sum_k \alpha_k \delta(t - \tau_k)$

Received signal: $r(t) = s(t) * h(t) = \sum_k \alpha_k s(t - \tau_k)$





(1) Equalization

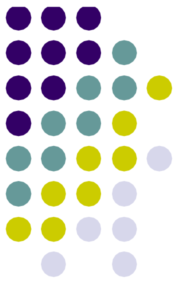
- **If the modulation bandwidth exceeds the coherence bandwidth of the radio channel, ISI occurs and modulation pulses are spread in time.**

- **Equalization compensates for intersymbol interference (ISI) created by multipath within time dispersive channels.**

  An equalizer within a receiver compensates for the average range of expected channel amplitude and delay characteristics.

- **Equalizers must be adaptive**

  since the channel is generally unknown and time varying.



(2) Diversity

- **Usually employed to reduce the depth and duration of the fades experienced by a receiver in a flat fading (narrowband) channel.**

  Without increasing the transmitted power or bandwidth.

- **Can be employed at both base station and mobile receivers.**
- **Types of diversity:**

  .antenna polarization diversity

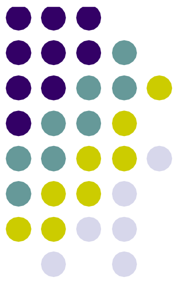
  .frequency diversity

  .time diversity.

  For example, CDMA systems often use a RAKE receiver, which provides link improvement through time diversity

- **Spatial diversity is the most common one.**

  While one antenna sees a signal null, one of the other antennas may see a signal peak.



(3) Channel Coding

- **Used to Improve mobile communication link performance by adding redundant data bits in the transmitted message.**

  At the baseband portion of the transmitter, a channel coder maps a digital message sequence into another specific code sequence containing a greater number of bits than originally contained in the message.

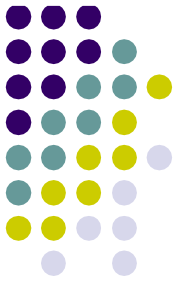
- **The coded message is then modulated for transmission in the wireless channel.**

- **coding can be considered to be a post detection technique.**

  Because decoding is performed after the demodulation portion

- **two general types of channel codes:**

  block codes
  convolutional codes.



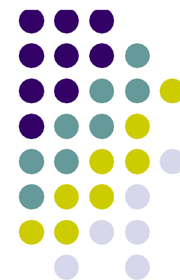
- **Channel coding is generally treated independently from the type of modulation used**

  but this has changed recently with the use of trellis coded modulation schemes that combine coding and modulation to achieve large coding gains without any bandwidth expansion.

Notes

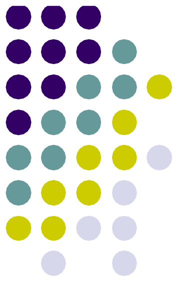
- **The three techniques of equalization, diversity, and channel coding are used to improve radio link performance (i.e. to minimize the instantaneous bit error rate)**
- **but the approach, cost, complexity, and effectiveness of each technique varies widely in practical wireless communication systems.**

4.2 Fundamentals of Equalization



- **Intersymbol interference (ISI)**
  - **caused by multipath propagation (time dispersion) ;**
  - **cause bit errors at the receiver;**
  - **the major obstacle to high speed data transmission over mobile radio channels.**

- **Equalization**
  - **a technique used to combat ISI;**
  - **can be any signal processing operation that minimizes ISI;**
  - **usually track the varying channel adaptively.**



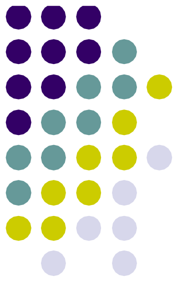
Operating modes of an adaptive equalizer

- **Training (first stage)**
  - **A known fixed-length training sequence is sent by the transmitter so that the receiver's equalizer may average to a proper setting.**
  - **The training sequence is designed to permit an equalizer at the receiver to acquire the proper filter coefficients in the worst possible channel conditions**

    The training sequence is typically a pseudorandom binary signal or a fixed, prescribed bit pattern.

    Immediately following the training sequence, the user data is sent.
  - **The time span over which an equalizer converges is a function of**
    1. **the equalizer algorithm**
    2. **the equalizer structure**
    3. **the time rate of change of the multipath radio channel.**

  Equalizers require periodic retraining in order to maintain effective ISI cancellation.



Operating modes of an adaptive equalizer

- **Tracking (second stage)**

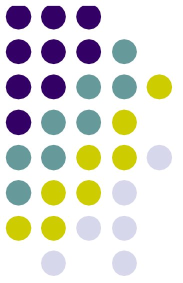
  Immediately following the training sequence, the user data is sent.

  - **As user data are received, the adaptive algorithm of the equalizer tracks the changing channel and adjusts its filter characteristics over time.**
  - **commonly used in digital communication systems**

    where user data is segmented into short time blocks.
  - **TDMA wireless systems are particularly well suited for equalizers.**

    data in fixed-length time blocks,

    training sequence usually sent at the beginning of a block

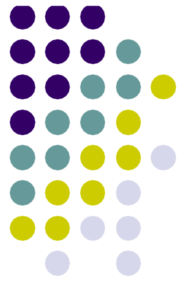


Communication system with an adaptive equalizer

- **Equalizer can be implemented at baseband or at IF in a receiver.**
- **Since the baseband complex envelope expression can be used to represent bandpass waveforms and, thus, the channel response, demodulated signal, and adaptive equalizer algorithms are usually simulated and implemented at baseband**

Block diagram of a simplified communications system using an adaptive equalizer at the receiver is shown in next page

Communication system with an adaptive equalizer

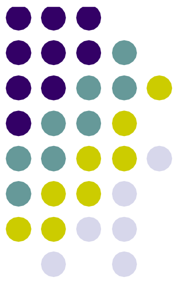


$x(t)$ → Modulator → Transmitter → Radio Channel → RF Front End → IF Stage → Detector Matched Filter → + ← $n_b(t)$

$f(t)$

$y(t)$ → Adaptive Equalizer $h_{eq}(t)$ → $\hat{d}(t)$ → Decision Maker → $d(t)$

Σ → $e(t)$



Relevant equations

$$y(t) = x(t) * f(t) + n_b(t)$$

$$\hat{d}(t) = x(t) * f(t) * h_{eq}(t) + n_b(t) * h_{eq}(t)$$

$$h_{eq}(t) = \sum_k c_k \delta(t - nT_s)$$

To eliminate ISI, we must have

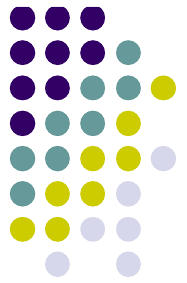
$$h_{eq}(t) * f(t) = \delta(t) \qquad H_{eq}(f) = \frac{1}{F(f)}$$

an equalizer is an inverse filter of the channel.

In frequency selective channel, enhances the frequency components with small amplitudes, attenuates the strong frequencies

therefore provide a flat, composite, received frequency response and linear phase response.

4.3 A Generic Adaptive Equalizer



y_k Z^{-1} y_{k-1} Z^{-1} y_{k-2} Z^{-1} ------ Z^{-1}

w_0 w_1 w_2 ------ w_N

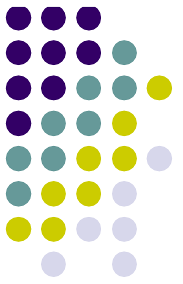
Σ $\hat{d}_k$

Adaptive algorithm that updates the weights

e_k Σ

Prior knowledge: d_k

4.3 A Generic Adaptive Equalizer

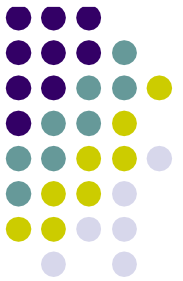


- **A transversal filter with**
  - **N delay elements**
  - **N+1 taps**
  - **N+1 tunable complex multipliers**
  - **N+1 weights:**
- **These weights are updated continuously by the adaptive algorithm**

  either on a sample by sample basis or on a block by block basis.
- **The adaptive algorithm is controlled by the error signal e_k.**

  e_k *is derived by comparing the output of the equalizer with some signal which is either an exact scaled replica of the transmitted signal x_k or which represents a known property of the transmitted signal.*

4.3 A Generic Adaptive Equalizer



- **A cost function is used**

  the cost function is minimized by using e_kThe, and the weights are updated iteratively.

- **For example, The least mean squares (LMS) algorithm can serve as a cost function.**
- **Iterative operation based on LMS**

  New weights = Previous weights + (constant) x (Previous error) x (Current input vector)

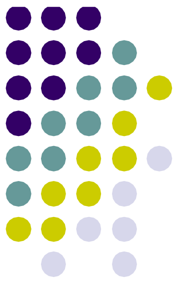
  Where

  Previous error = Previous desired output — Previous actual output

  This process is repeated rapidly in a programming loop while the equalizer attempts to converge

  Upon reaching convergence, the adaptive algorithm freezes the filter weights until the error signal exceeds an acceptable level or until a new training sequence is sent.

4.3 A Generic Adaptive Equalizer



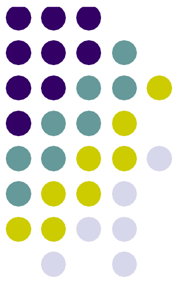
- **Techniques used to minimize the error**
  - **gradient**
  - **steepest decent algorithms**

- **Based on classical equalization theory, the most common cost function is MSE**

MSE----mean square error (MSE) between the desired signal and the output of the equalizer

Denoted by $E[e(k) \cdot e^*(k)]$

4.3 A Generic Adaptive Equalizer



Blind algorithms

- more recent class of adaptive algorithms
- able to exploit characteristics of the transmitted signal and do not require training sequences.

  provide equalizer convergence without burdening the transmitter with training overhead

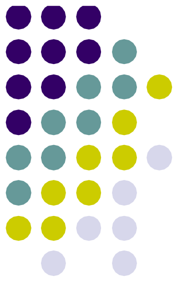
  able to acquire equalization through property restoral techniques of the transmitted signal,
- Two techniques:
  - the constant modulus algorithm (CMA)

    used for constant envelope modulation

    forces the equalizer weights to maintain a constant envelope on the received signal
  - spectral coherence restoral algorithm (SCORE).

    exploits spectral redundancy or cyclostationarity in the transmitted signal

4.4 Equalizers in a Communications Receiver



- **Because noise is present, an equalizer is unable to achieve perfect performance.**
- **Therefore, the instantaneous combined frequency response will not always be flat, resulting in some finite prediction error.**
- **The mean squared error (MSE) $E[e_k^2]$ is one of the most important measures of how well an equalizer works.**

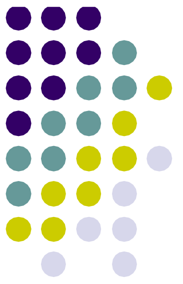
    Minimizing MSE $E[e_k^2]$ tends to reduce the bit error rate.

- **For wireless communication links, it would be best to minimize the instantaneous probability of error instead of MSE**

    generally results in nonlinear equations

    much more difficult to solve in real-time

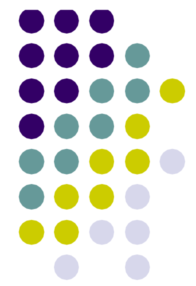
4.5 Survey of Equalization Techniques



- **Equalization techniques can be subdivided into two general categories:**
  - **linear equalization**
    - *The output of the decision maker is not used in the feedback path to adapt the equalizer.*
  - **nonlinear equalization**
    - *The output of the decision maker is used in the feedback path to adapt the equalizer.*

- **Many filter structures are used to implement linear and nonlinear equalizers**

- **For each structure, there are numerous algorithms used to adapt the equalizer.**



Classification of equalizers

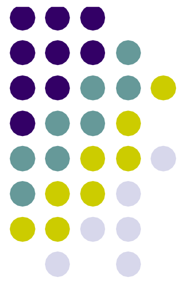
Equalizer

Types:
- Linear
- Nonlinear
  - DFE
  - ML Symbol Detector
  - MLSE

Structures:

Linear: Transversal	Linear: Lattice	DFE: Transversal	DFE: Lattice	ML Symbol Detector / MLSE: Transversal Channel Est.
Zero forcing LMS RLS Fast RLS Sq. root RLS	Gradient RLS	LMS RLS Fast RLS Sq. root RLS	Gradient RLS	LMS RLS Fast RLS Sq. root RLS

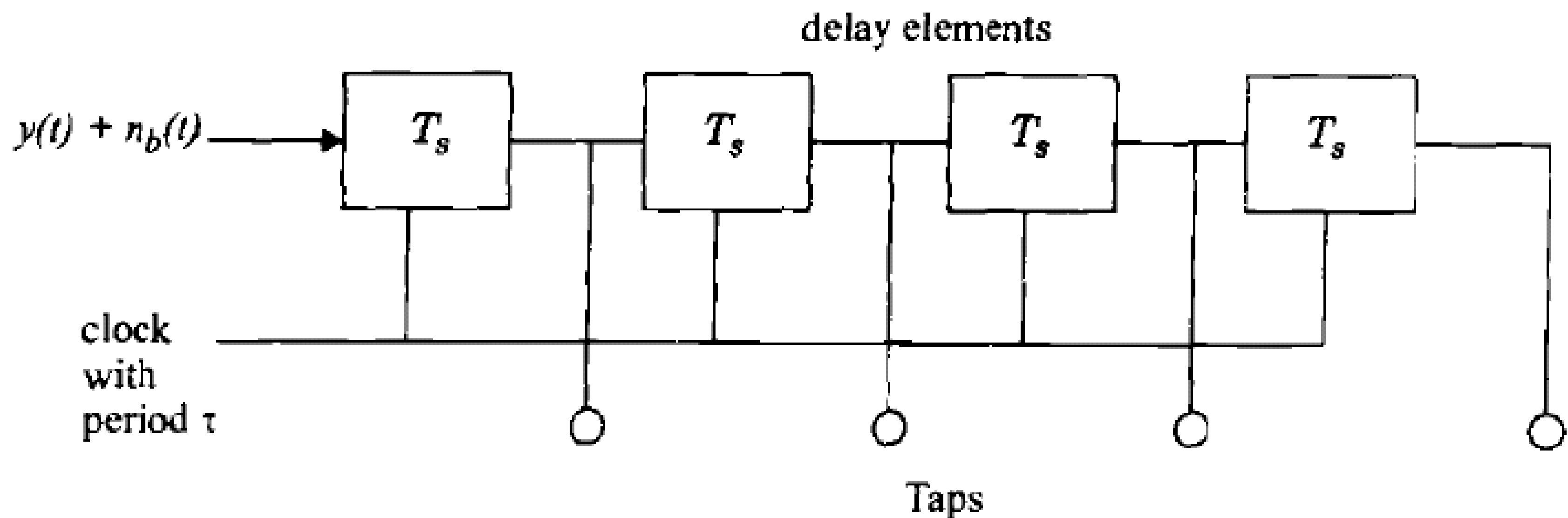
Algorithms



Most common structure:
---- Linear transversal equalizer (LTE)

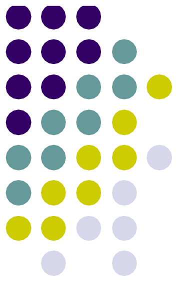
- **made up of tapped delay lines, with the tappings spaced a symbol period (Ts) apart**
- **the transfer function can be written as a function of the delay operator** $-j\sigma T_s$ **or** Z^{-1}

Assuming that the delay elements have unity gain and delay Ts, of a linear



Basic linear transversal equalizer structure

Most common structure:
---- Linear transversal equalizer (LTE)



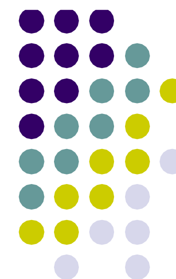
Two types of LTE

- **finite impulse response (FIR) filter**
  - **The simplest LTE uses only feedforward taps**
  - **Transfer function is a polynomial in Z^{-1}**
  - **has many zeroes but poles only at z = 0**

    Usually simply called a transversal filter

- **Infinite impulse response (IIR) filter**
  - **has both feedforward and feedback taps**
  - **transfer function is a rational function of Z^{-1} with poles and zeros.**
  - **tend to be unstable when used in channels where the strongest pulse arrives after an echo pulse (i.e., leading echoes)**

    rarely used.



Feedback taps

b_1 b_2 b_3 b_4

$y(t)$ Σ T_s T_s T_s T_s

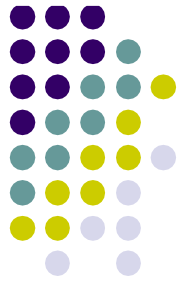
a_1 a_2 a_3 a_4

Feedforward taps

Σ $\hat{d}(t)$

Tapped delay line filter with both feedforward and feedback taps (IIR)

4.6 Linear Equalizers



Transversal filter implementation (LTE)

Input y_{k+N1} z^{-1} y_{k+N1-1} z^{-1} y_k y_{k-N2+1} z^{-1} z^{-1} y_{k-N2}

c^*_{-N1} c^*_{-N1+1} c^*_0 c^*_{N2-1} c^*_{N2}

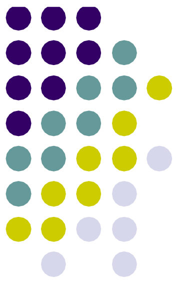
$\hat{d}_k$ 输出 d_k Output

门限检测器

Threshold Detector

This type of equalizer is the simplest.

4.6 Linear Equalizers



- **current and past values of the received signal are linearly weighted by the filter coefficient and summed to produce the output,**

  If the delays and the tap gains are analog, the continuous output of the equalizer is sampled at the symbol rate and the samples are applied to the decision device.

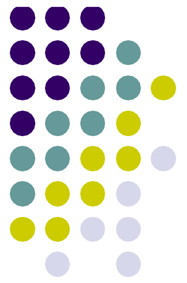
  Implementation is usually carried out in the digital domain where the samples of the received signal are stored in a shift register.

- **The output before decision making (threshold detection)**

$$\hat{d}_k = \sum_{n=-N_1}^{N_2} (c_n^*) y_{k-n}$$

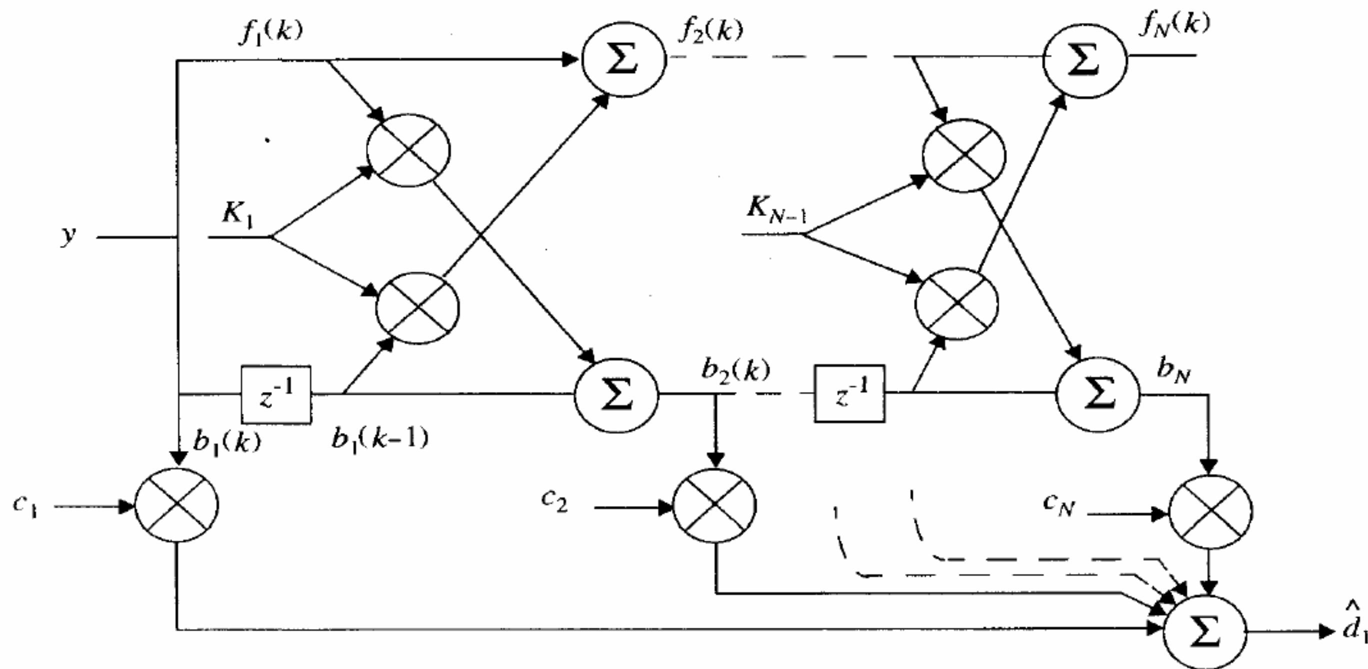
- **The minimum MSE it can achieve**

$$E[|e(n)|^2] = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{N_0}{|F(e^{j\omega T})|^2 + N_0} d\omega$$



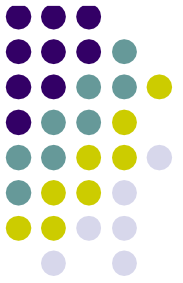
4.6 Linear Equalizers

Lattice filter implementation



Numerical stable, faster convergence, Complicated

4.6 Linear Equalizers

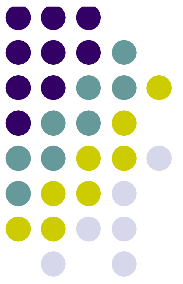


- **Two main advantages of the lattice equalizer**
  - **numerical stability**
  - **faster convergence**

- **Unique structure allows dynamic assignment of the most effective length**
  - **When channel is not very time dispersive Only a fraction of the stages are used.**
  - **channel becomes more time dispersive Length can be increased without stopping the operation**

- **Drawback: more complicated than LTE**

4.7 Nonlinear Equalization



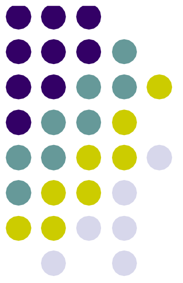
- **Linear equalizers do not perform well on channels which have deep spectral nulls in the passband.**

  In an attempt to compensate for the distortion, the linear equalizer places too much gain in the vicinity of the spectral null, thereby enhancing the noise present in those frequencies.

- **Nonlinear equalizers are used in applications where the channel distortion is too severe for a linear equalizer to handle.**

- **Three very effective nonlinear equalizer**
  - **Decision Feedback Equalization (DFE)**
  - **Maximum Likelihood Symbol Detection**
  - **Maximum Likelihood Sequence Estimation (MLSE)**

4.7.1 Decision Feedback Equalization (DFE)



Basic idea:

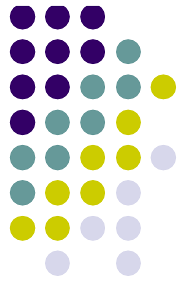
once an information symbol has been detected, the ISI that it induces on future symbols can be estimated and subtracted out before detection of subsequent symbols.

- **DFE Can be realized in either the direct transversal form or as a lattice filter.**
- **The LTE form consists of a feedforward filter (FFF) and a feedback filter (FBF).**

  The FBF is driven by decisions on the output of the detector, and its coefficients can be adjusted to cancel the ISI on the current symbol from past detected symbols.

- **The equalizer has N1 + N2 + I taps in FFF and N3 taps in FBF**

4.7.1 Decision Feedback Equalization (DFE)



Input

y_{k+N1} z^{-1} y_{k+N1-1} z^{-1} y_k y_{k-N2+1} z^{-1} z^{-1} y_{k-N2}

c^*_{-N1} c^*_{-N1+1} c^*_0 c^*_{N2-1} c^*_{N2}

$\hat{d}_k$ Output d_k

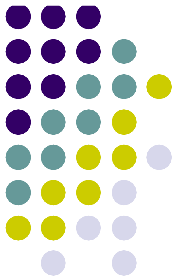
Feedforward Filter

d_{k-N3} z^{-1} d_{k-1} z^{-1} d_k

F^*_{N3} F^*_{N3-1} F^*_1

Feedback Filter

4.7.1 Decision Feedback Equalization (DFE)



The output of DFE $\hat{d}_k = \sum_{n=-N_1}^{N_2} c_n^* y_{k-n} + \sum_{i=1}^{N_3} F_i d_{k-i}$

The minimum mean square error of DFE

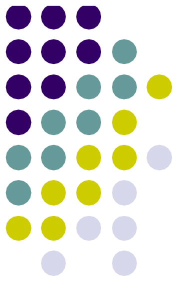
$$E[|e(n)|^2]_{min} = \exp\left\{\frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} \ln\left[\frac{N_0}{|F(e^{j\omega T})|^2 + N_0}\right] d\omega\right\}$$

- **It can be seen that the minimum MSE for a DFE is always smaller than that of an LTE**

  Unless $|F(e^{j\sigma T})|$ *is a constant, where adaptive equalization is not needed*

- **If there are nulls in the** $|F(e^{j\sigma T})|$ **, a DFE has significantly smaller minimum MSE than an LTE.**

4.7.1 Decision Feedback Equalization (DFE)

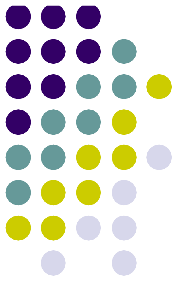


Conclusion

- **an LTE is well behaved when the channel spectrum is comparatively flat**
- **a DFE is more appropriate for severely distorted wireless channels.**

- **If the channel is severely distorted or exhibits nulls in the spectrum**
  - **the performance of an LTE deteriorates and the mean squared error of a DFE is much better than a LTE.**
  - **Also, an LTE has difficulty equalizing a nonminimum phase channel**

    where the strongest energy arrives after the first arriving signal component.

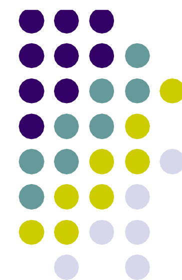


Another form of DFE----predictive DFE

- **also consists of a feed forward filter (FFF) as in the conventional DFE.**
- **Difference: the feedback filter (FBF) is driven by an input sequence formed by the difference of the output of the detector and the output of the feed forward filter.**

  the FBF here is called a noise predictor because it predicts the noise and the residual ISI contained in the signal at the FFF output and subtracts from it
- **The predictive DFE performs as well as the conventional DFE as the limit in the number of taps in the FFF and the FBF approach infinity.**
- **The FEF in the predictive DFE can also be realized as a lattice structure**

Another form of DFE----predictive DFE



Received Signal Samples

Feedforward Filter

+

Σ

-

Decision Device

Output Decision

-

Σ

+

Output Prediction

Error Signal

Error Signal

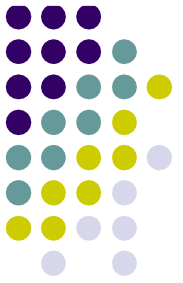
Feedback Filter

-

Σ

+

4.7.2 Maximum Likelihood Sequence Estimation (MLSE) equalizer



The MSE-based linear equalizers are optimum with respect to the criterion of minimum probability of symbol error when the channel does not introduce any amplitude distortion.

Yet this is precisely the condition in which an equalizer is needed for a mobile communications link.

- **MLSE uses various forms of the classical maximum likelihood receiver structure.**

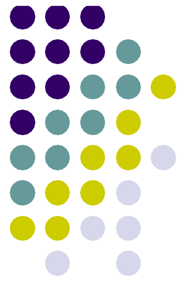
- **the MLSE tests all possible data sequences (rather than decoding each received symbol by itself), and chooses the data sequence with the maximum probability as the output.**

  A channel impulse response simulator is used within the algorithm,

- **Drawback: An MLSE usually has a large computational requirement**

  especially when the delay spread of the channel is large.

4.7.2 Maximum Likelihood Sequence Estimation (MLSE) equalizer



$y(t)$

Matched Filter

$z(t)$

$\{z_n\}$

MLSE

$\{a_n\}$

$\{s_i\}$

Delay

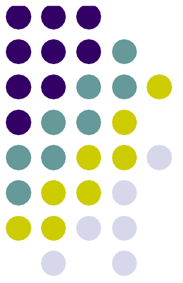
+

−

e

Channel Estimator

4.7.2 Maximum Likelihood Sequence Estimation (MLSE) equalizer



- **The MLSE can be viewed as a problem in estimating the state of a discrete time finite state machine**

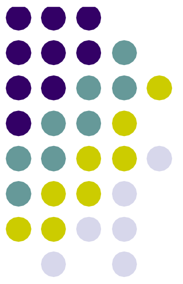
  The channel has M^L states, where M is the size of the symbol alphabet of the modulation.

- **An M^L trellis is used by the receiver to model the channel over time.**

  The Viterbi algorithm then tracks the state of the channel by the paths through the trellis.

- **The MLSE is optimal in the sense that it minimizes the probability of a sequence error.**

4.7.2 Maximum Likelihood Sequence Estimation (MLSE) equalizer



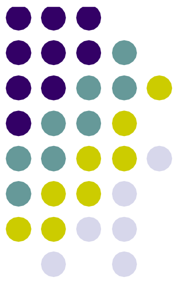
NOTES:

- **The MLSE requires knowledge of the channel characteristics**
  in order to compute the metrics for making decisions.

- **The MLSE also requires knowledge of the statistical distribution of the noise corrupting the signal**
  the probability distribution of the noise determines the form of the metric for optimum demodulation of the received signal.

- **The matched filter operates on the continuous time signal, whereas the MLSE and channel estimator rely on discretized (nonlinear) samples.**

4.8 Algorithms for Adaptive Equalization



- **Equalizer requires a specific algorithm to update the coefficients and track the channel variations.**

  Since it compensates for an unknown and time-varying channel

- **This section outlines three of the basic algorithms for adaptive equalization.**

  Though the algorithms detailed in this section are derived for the linear, transversal equalizer, they can be extended to other equalizer structures, including nonlinear equalizers.

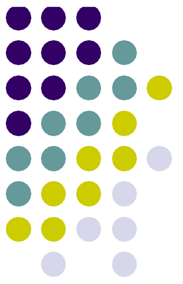
4.8 Algorithms for Adaptive Equalization

Factors determining the performance of an algorithm:

- **Rate of convergence (fast or slow?)**
  - **Defined as the number of iterations required for the algorithm, in response to stationary inputs, to converge close enough to the optimum solution.**
  - **A fast rate of convergence allows the algorithm to adapt rapidly to a stationary environment of unknown statistics.**
  - **Furthermore, it enables the algorithm to track statistical variations when operating in a nonstationary environment.**

- **Misadjustment (precise or not?)**
  - **Provides a quantitative measure of the amount by which the final value of the mean square error, averaged over an ensemble of adaptive filters, deviates from the optimal minimum mean square error.**

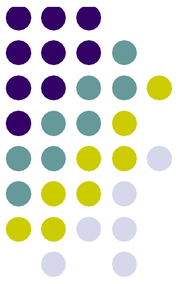
4.8 Algorithms for Adaptive Equalization



Factors determining the performance of an algorithm:

- **Computational complexity (simple or complex?)**
  - **Number of operations required to make one complete iteration of the algorithm.**
- **Numerical properties (stable or not?)**
  - **When an algorithm is implemented numerically, inaccuracies are produced due to round-off noise and representation errors in the computer.**
  - **These kinds of errors influence the stability of the algorithm.**

4.8 Algorithms for Adaptive Equalization



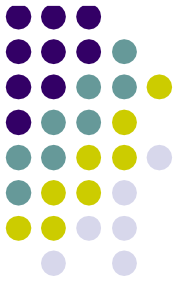
Practical considerations for choice of an equalizer structure and its algorithm

- **The cost of the computing platform (affordable or not?)**
  especially when used in user equipments

- **The power budget (power limited applications or else?)**
  In portable radio applications, battery drain at the subscriber unit is a paramount consideration

- **The radio propagation characteristics (fast fading & time delay spread?)**
  The speed of the mobile unit determines the channel fading rate and the Doppler spread, which is directly related to the coherence time of the channel

4.8 Algorithms for Adaptive Equalization

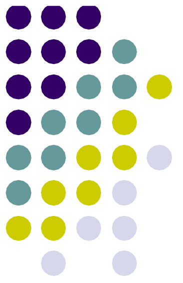


Three classic equalizer algorithms

- **Zero Forcing Algorithm (ZF)**
- **Least Mean Square Algorithm (LMS)**
- **Recursive Least Squares Algorithm (RLS)**

Please read references for detailed information on a specific algorithm.

4.8.1 Zero Forcing (ZF) Algorithm



Criterion:

to force the samples of the combined channel and equalizer impulse response to zero at all but one of sample points in the tapped delay line filter.

Disadvantage:

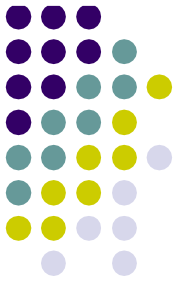
may excessively amplify noise at frequencies where the folded channel spectrum has high attenuation.

Suitability:

Wireline communications

$$H_{eq}(f) = \frac{1}{H_{ch}(f)}, |f| < \frac{1}{2T}$$

4.8.2 Least Mean Square (LMS) Algorithm



Criterion:

to minimize the **mean square error** (MSE) between the desired equalizer output and the actual equalizer output.

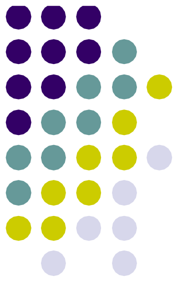
Minimize $\frown = E[e_k^{*} \cdot e_k]$

Must be solved iteratively

Simplest algorithm, requires only 2N + l operations per iteration.

- **The LMS equalizer maximizes the signal to distortion ratio at its output within the constraints of the equalizer filter length.**
- **a step size** α **is used to control the convergence rate and the stability**

4.8.2 Least Mean Square (LMS) Algorithm



Disadvantage: **low convergence rate.**

Because of the only one parameter α

Especially when the eigenvalues of the input covariance matrix RNN have a very large spread, i.e, $Z_{max} / Z_{min} >> 1$

- **If an input signal has a time dispersion characteristic that is greater than the propagation delay through the equalizer, then the equalizer will be unable to reduce distortion.**

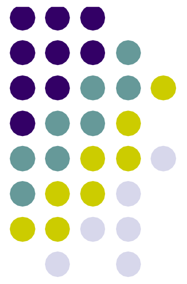
- **To prevent the adaptation from becoming unstable, the value of** α **is chosen from** $0 < \alpha < 2 / \sum_{i=1}^{N} \lambda_i$

  where Z_i *is the ith eigenvalue of the covariance matrix* R_{NN}.

- **The step size** α **can be controlled by the total input power in order to avoid instability in the equalizer [Hay86].**

  since $\sum^{N} \lambda_i = y_N^T(n) y_N(n)$

4.8.3 Recursive Least Squares (RLS) Algorithm



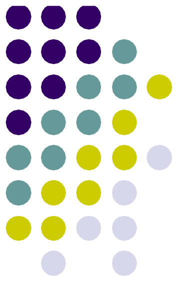
- RLS is Proposed to improve the convergence rate of LMS algorithm.

- Error measures expressed in terms of a time average of the actual received signal instead of a statistical average.

$$J(n) = \sum_{i=1}^{n} \lambda^{n-i} e^{*}(i,n)\, e(i,n)$$

  - Z is the weighting coefficient that can change the performance of the equalizer.
  - If a channel is time-invariant, Z can be set to 1. Usually 0.8-1.
  - The value of Z has no influence on the rate of convergence, but does determine the tracking ability.
  - The smaller the Z, the better the tracking ability of the equalizer.
  - However, if Z is too small, the equalizer will be unstable

4.8.3 Recursive Least Squares (RLS) Algorithm

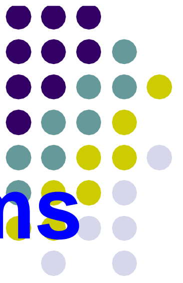


Advantage: high convergence rate

Disadvantage: sometimes unstable

The RLS algorithm described above, called the Kalman RLS algorithm

Uses 2.5N2 + 4.5N arithmetic operations per iteration.



4.8.4 Summary of equalization algorithms

- **There are number of variations of the LMS and RLS algorithms**
- **RLS algorithms have similar convergence and tracking performances, which are much better than the LMS algorithm.**

Usually have high computational requirement and complex program structures.

Some RLS algorithms tend to be unstable.

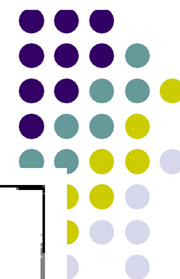
About FTF

- **Among the RLS algorithms, fast transversal filter (FTF) algorithm requires the least computation**
- **a rescue variable can be used to avoid instability.**

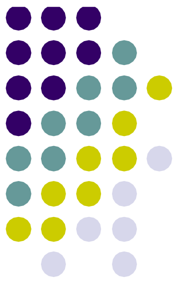
However, rescue techniques tend to be a bit tricky for widely varying mobile radio channels.

FTF is not widely used.

Comparison of Various Algorithms for Adaptive Equalization [Pro9l]



Algorithm	Number of Multiply Operations	Advantages	Disadvantages
LMS Gradient DFE	$2N + 1$	Low computational complexity, simple program	Slow convergence, poor tracking
Kalman RLS	$2.5N^2 + 4.5N$	Fast convergence, good tracking ability	High computational complexity
FTF	$7N + 14$	Fast convergence, good tracking, low computational complexity	Complex programming, unstable (but can use rescue method)
Gradient Lattice	$13N - 8$	Stable, low computational complexity, flexible structure	Performance not as good as other RLS, complex programming
Gradient Lattice DFE	$13N_1 + 33N_2 - 36$	Low computational complexity	Complex programming
Fast Kalman DFE	$20N + 5$	Can be used for DFE, fast convergence and good tracking	Complex programming, computation not low, unstable
Square Root RLS DFE	$1.5N^2 + 6.5N$	Better numerical properties	High computational complexity



4.9 Fractionally Spaced Equalizers(FSE)

- **In the presence of channel distortion, the matched filter prior to the equalizer must be matched to the channel and the corrupted signal.**

  Usually get the suboptimal result because the channel response is unknown.

  This results in a significant degradation in performance.

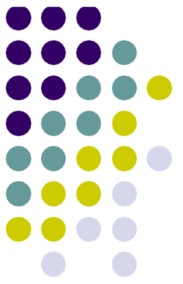
- **FSE is based on sampling the incoming signal at least as fast as the Nyquist rate.**

- **The FSE compensates for the channel distortion before aliasing effects occur due to the symbol rate sampling.**

  FSE incorporates the functions of a matched filter and equalizer into a single filter structure.

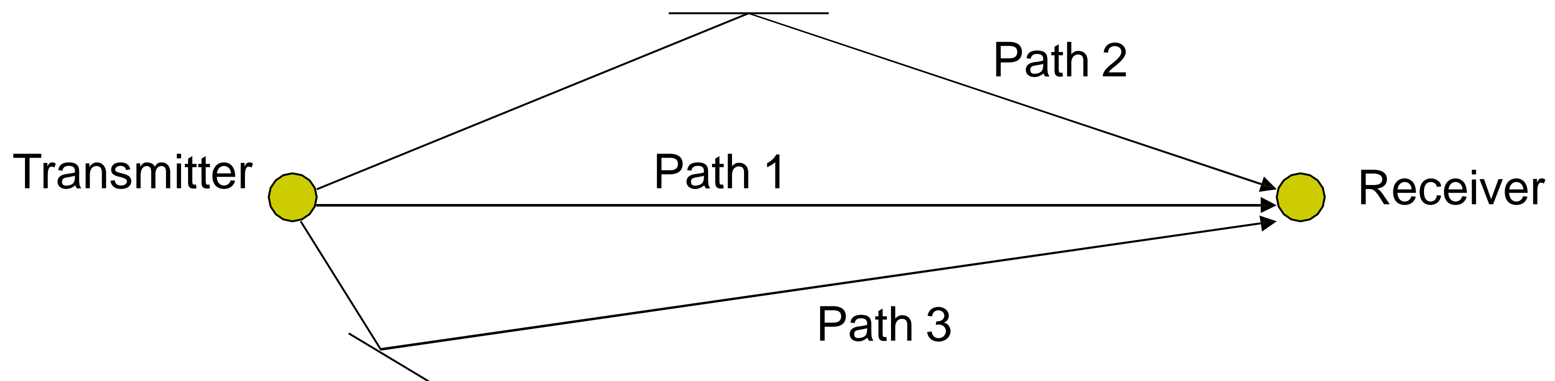
- **Simulation results demonstrate the effectiveness of the FSE over a symbol rate equalizer. (see [Ste94])**

4.10 Fundamentals of Diversity Techniques

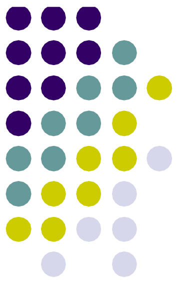


- **Random nature of radio propagation:**
  - **Multipath propagation**
  - **Independent fading of each Multipath component**
  - **If one radio path undergoes a deep fade, another independent path may have a strong signal**

- **Diversity exploits the random nature of radio propagation by finding independent signal paths for communication, so as to boost the instantaneous SNR at the receiver.**



4.10 Fundamentals of Diversity Techniques

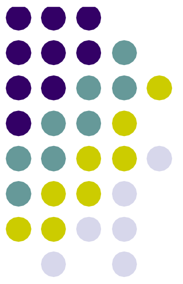


- **Diversity is a powerful communication receiver technique that provides wireless link improvement at relatively low cost.**
- **Requires no training**
- **In virtually all applications, diversity decisions are made by the receiver, and are unknown to the transmitter.**

Two types of diversity

- **Microscopic diversity ⟷ small scale fading**
- **Macroscopic diversity ⟷ large scale fading**

4.10 Fundamentals of Diversity Techniques



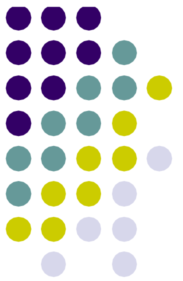
Microscopic diversity

- **Small-scale fades: deep and rapid amplitude fluctuations over distances of just a few wavelengths.**
  caused by multiple reflections from the surroundings in the vicinity of the mobile.
  results in a Rayleigh fading distribution of signal strength over small distances.

- **Microscopic diversity techniques can exploit the rapidly changing signal.**
  For example, use two antennas at the receiver (separated by a fraction of a meter), one may receive a null while the other receives a strong signal.
  By selecting the best signal at all times, a receiver can mitigate small-scale fading effects
  Called antenna diversity or space diversity

- **Samples: Rake receiver, MIMO transmission**

4.10 Fundamentals of Diversity Techniques



Macroscopic diversity

- **Large-scale fading: caused by shadowing due to variations in both the terrain profile and the nature of the surroundings.**

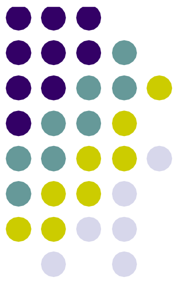
  In deeply shadowed conditions, the received signal strength at a mobile can drop well below that of free space.

  log-normally distributed with a standard deviation of about 10 dB in urban environments.

- **Macro-scope diversity: By selecting a base station which is not shadowed when others are, the mobile can improve substantially the average ratio on the forward link.**

  It is the mobile that takes advantage of large separations between the serving base stations.

4.10 Fundamentals of Diversity Techniques



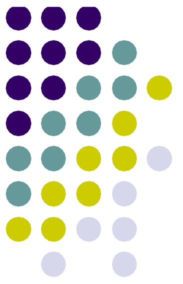
Macroscopic diversity

- **Macroscopic diversity is also useful at the base station receiver.**
  By using base station antennas that are sufficiently separated in space, the base station is able to improve the reverse link by selecting the antenna with the strongest signal from the mobile.

- **Used to combat slow fading (shadowing)**

- **Samples: Base-station handoff in cellular networks**

4.10 Fundamentals of Diversity Techniques



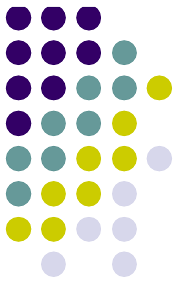
Macro-scope diversity

Base station

Base station

Mobile

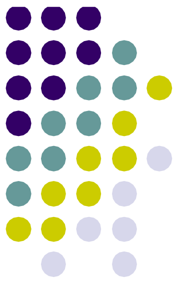
4.10 Fundamentals of Diversity Techniques



- **Strategies used in diversity techniques**
  - **Selection diversity**
  - **Maximal ratio combining diversity**
  - **Equal-gain combining diversity**
  - **Hybrid schemes**

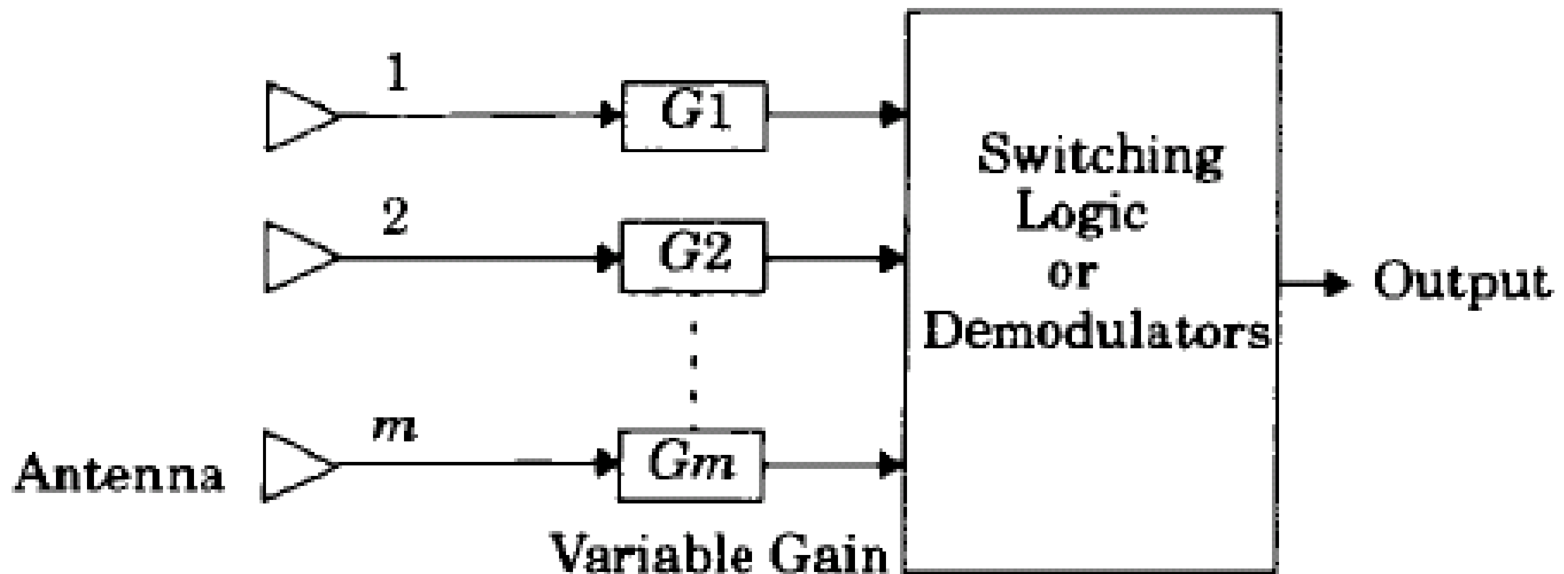
- **Practical considerations**
  - **effectiveness, complexity, cost, and etc.**

4.10.1 Derivation of Selection Diversity improvement

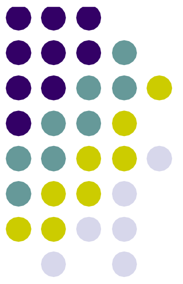


- **Consider M** independent **Rayleigh fading channels available areceiver.**

  Each channel is called a diversity branch.



4.10.1 Derivation of Selection Diversity improvement



- Further assume that each branch has the same average SNR given by

$$SNR = \Gamma = \frac{E_b}{N_0}\overline{\alpha^2}$$

Where we assume $\overline{\alpha^2} = 1$.

- If each branch has an instantaneous $SNR = \mu_i$, then the pdf of μ_i is

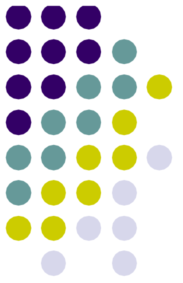
$$p(\gamma_i) = \frac{1}{\Gamma}e^{\frac{-\gamma_i}{\Gamma}} \quad \gamma_i \geq 0$$

where Γ is the mean SNR of each branch.

- The probability that a single branch has SNR less than some threshold y μ is

$$Pr[\gamma_i \leq \gamma] = \int_0^{\gamma} p(\gamma_i)d\gamma_i = \int_0^{\gamma} \frac{1}{\Gamma}e^{\frac{-\gamma_i}{\Gamma}}d\gamma_i$$

4.10.1 Derivation of Selection Diversity improvement



- **Now, the probability that all M independent diversity branches receive signals which are simultaneously less than some specific SNR threshold μ is**

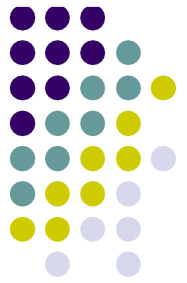
$$Pr[\gamma_1, \ldots, \gamma_M \le \gamma] = (1 - e^{-\gamma/\Gamma})^M = P_M(\gamma)$$

This is the probability of all branches failing to achieve $SNR = \mu_i$.

- **If a single branch achieves $SNR > \mu$, then the probability that $SNR > \mu$ for one or more branches is given by**

$$Pr[\gamma_i > \gamma] = 1 - P_M(\gamma) = 1 - (1 - e^{-\gamma/\Gamma})^M$$

This is the probability of exceeding a threshold when selection diversity is used.



PERCENT PROBABILITY THAT AMPLITUDE > ABSCISSA

0.01
0.1
0.5
2.0
6.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
95.0
98.0
99.0
99.5
99.8
99.9
99.95
99.98
99.99

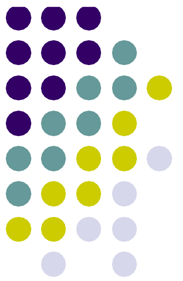
M=1
M=2
M=3
M=4
M=6

-40 -30 -20 -10 0 10

10 LOG (γ/Γ), dB

Graph of probability distributions of $SNR = \gamma$ threshold for M branch selection diversity. The

4.10.1 Derivation of Selection Diversity improvement



How to determine the average signal-to-noise ratio of the received signal when diversity is used?

- **First of all, find the pdf of μ (the instantaneous SNR when M branches are used). Thus we compute the derivation of CDF $P_M(\mu)$,**

$$p_M(\gamma) = \frac{d}{d\gamma}P_M(\gamma) = \frac{M}{\Gamma}(1-e^{-\gamma/\Gamma})^{M-1}e^{-\gamma/\Gamma}$$

- **Then, we can compute the average SNR, $\bar{\mu}$,**

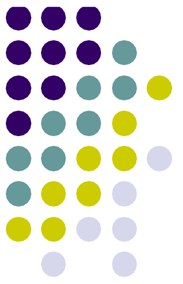
$$\bar{\gamma} = \int_0^\infty \gamma p_M(\gamma)d\gamma = \Gamma\int_0^\infty Mx(1-e^{-x})^{M-1}e^{-x}dx$$

where $x = \mu/\Gamma$.

The above equation can be evaluated to yield the average SNR improvement offered by selection diversity.

$$\frac{\bar{\mu}}{\Gamma} = \sum_{k=1}^{M}\frac{1}{k}$$

4.10.1 Derivation of Selection Diversity improvement



- **Selection diversity offers an average improvement in the link margin without requiring additional transmitter power or sophisticated receiver circuitry.**

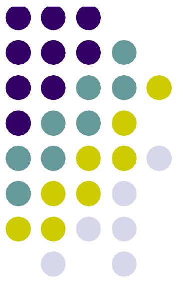
  The diversity improvement can be directly related to the average bit error rate for various modulations.

- **Selection diversity is easy to implement because all that is needed is a side monitoring station and an antenna switch at the receiver.**

- **However, it is not an optimal diversity technique because it does not use all of the possible branchessimultaneously.**

  Maximal ratio combining uses each of the M branches in a co-phased and weighted manner such that the highest achievable SNR is available at the receiver at all times.

4.10.1 Derivation of Selection Diversity improvement



Example

Assume four branch diversity is used, where each branch receives an independent Rayleigh fading signal. If the average SNR is 20 dB, determine the probability that the SNR will drop below 10 dB. Compare this with the case of a single receiver without diversity.

Solution

For this example the specified threshold $\gamma = 10$ dB, $\Gamma = 20$ dB, and there are four branches. Thus $\gamma/\Gamma = 0.1$ and using equation (6.58),

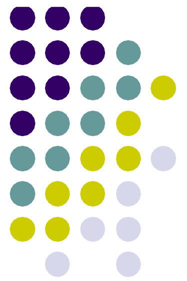
$$P_4(10\text{ dB}) = (1 - e^{-0.1})^4 = 0.000082$$

When diversity is not used, equation (6.58) may be evaluated using $M = 1$.

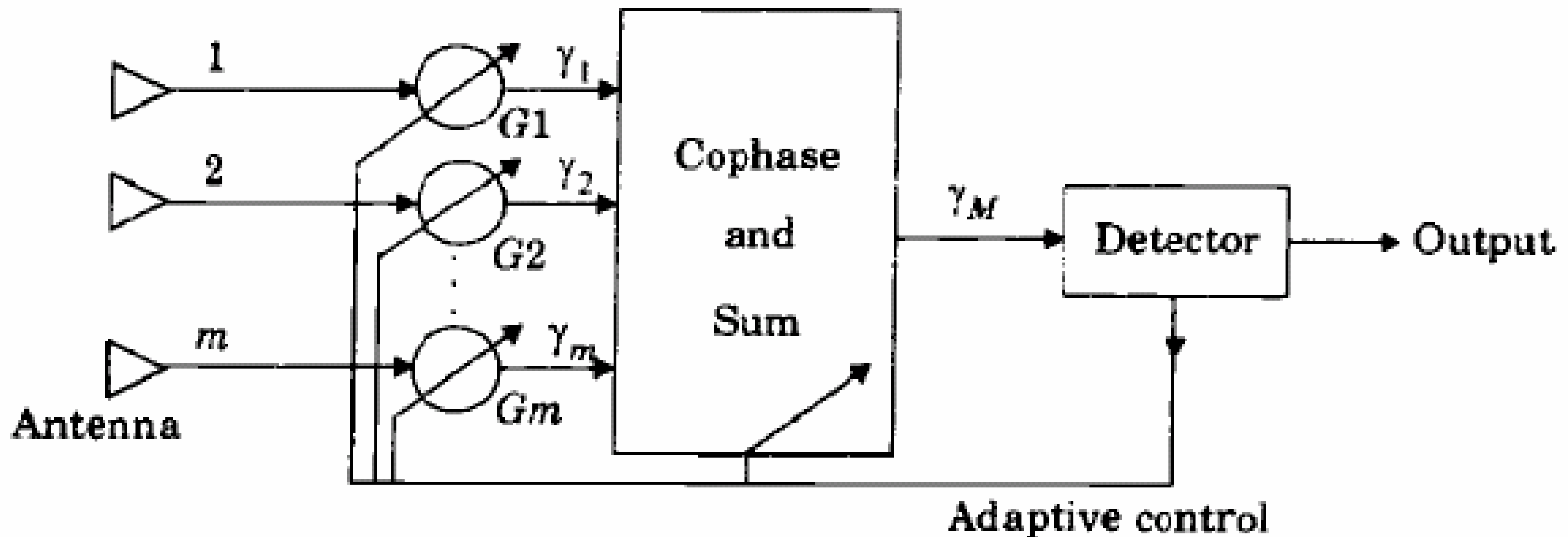
$$P_1(10\text{ dB}) = (1 - e^{-0.1})^1 = 0.095$$

Notice that without diversity the *SNR* drops below the specified threshold with a probability that is three orders of magnitude greater than if four branch diversity is used!

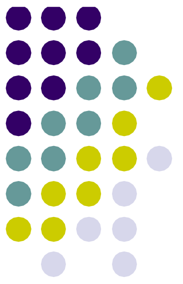
7.10.2 Derivation of Maximal Ratio Combining Improvement



In maximal ratio combining, the voltage signals r_i *from each of the M diversity branches are co-phased to provide coherent voltage addition and are individually weighted to provide optimal SNR.*



4.10.2 Derivation of Maximal Ratio Combining Improvement



1) The SNR out of the diversity combiner:

- **If each branch has gain G_i, then the resulting signal envelope applied to the detector is**

$$r_M = \sum_{i=1}^{M} G_i r_i$$

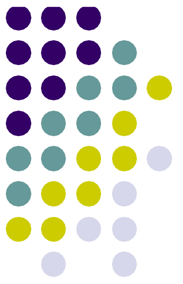
- **Assuming that each branch has the same average noise power *N*, the total noise power N_T applied to the detector is simply the weighted sum of the noise in each branch. Thus**

$$N_T = N \sum_{i=1}^{M} G_i^2$$

which results in an SNR applied to the detector, μ_M, given by

$$\gamma_M = \frac{r_M^2}{2N_T}$$

4.10.2 Derivation of Maximal Ratio Combining Improvement



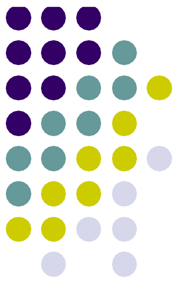
- **Using Chebychev's inequality, μ_M is maximized when $G_i = r_i / N$, which leads to**

$$\gamma_M = \frac{1}{2}\frac{\sum (r_i^2/N)^2}{N\sum (r_i^2/N^2)} = \frac{1}{2}\sum_{i=1}^{M}\frac{r_i^2}{N} = \sum_{i=1}^{M}\gamma_i \qquad \text{(7-66)}$$

- **Conclusion:**

  The SNR out of the diversity combiner is simply the sum of the SNRs in each branch.

4.10.2 Derivation of Maximal Ratio Combining Improvement



2) The pdf of μ_M

- **According to Chapter 3, μ_M is a Chi-square distribution of 2M Gaussian random variables. Thus, the pdf for μ_M is**

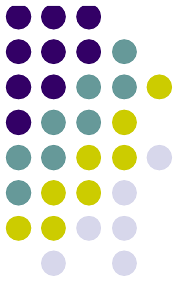
$$p(\gamma_M) = \frac{\gamma_M^{M-1} e^{-\gamma_M/\Gamma}}{\Gamma^M (M-1)!} \quad \text{for } \gamma_M \geq 0 \qquad \text{(7-68)}$$

3) The CDF of μ_M

- **According to the abovementioned pdf, The probability that μ_M is less than some SNR threshold μ is**

$$Pr\{\gamma_M \leq \gamma\} = \int_0^\gamma p(\gamma_M) d\gamma_M = 1 - e^{-\gamma/\Gamma} \sum_{k=1}^{M} \frac{(\gamma/\Gamma)^{k-1}}{(k-1)!}$$

4.10.2 Derivation of Maximal Ratio Combining Improvement



4) The average SNR out of the diversity combiner, $\bar{\mu}_M$

- $\bar{\mu}_M$ **can be calculated by using the pdf of** μ_M **(Eq. (7.68)). But the direct way is to calculate it from Eq. (7-66).**

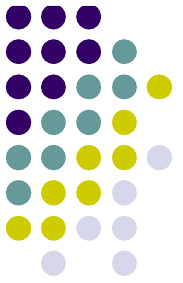
$$\overline{\gamma_M} = \sum_{i=1}^{M} \overline{\gamma_i} = \sum_{i=1}^{M} \Gamma = M\Gamma$$

- **That is to say, the average SNR,** $\bar{\mu}_M$ **, is simply the sum of the individual** μ_i **from each branch.**

The control algorithms for setting the gains and phases for maximal ratio combining receivers are similar to those required in equalizers and RAKE receivers.

Maximal ratio combining can be applied to virtually any diversity application, although often at much greater cost and complexity than other diversity techniques.

4.10.3 Practical Space Diversity Considerations



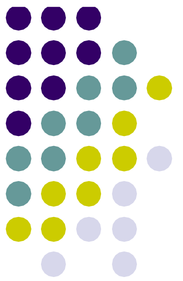
- **Space diversity (also known as antenna diversity), is one of the most popular forms of diversity used in wireless systems.**

- **The signals received from spatially separated antennas on the mobile would have essentially uncorrelated envelopes for antenna separations of one half wavelength or more.**

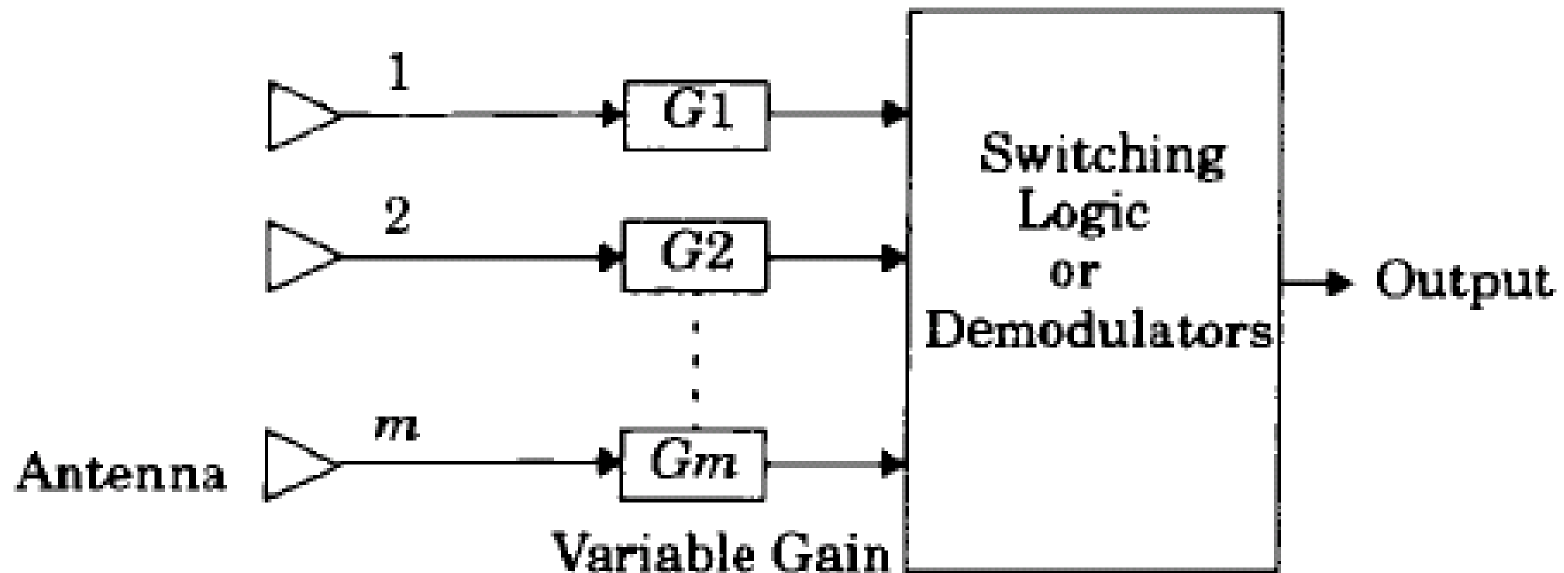
- **Space diversity can be used at either the mobile or base station, or both.**

Since the important scatterers are generally on the ground in the vicinity of the mobile, when base station diversity is used, the antennas must be spaced considerably far apart to achieve decorrelation (several tens of wavelengths).

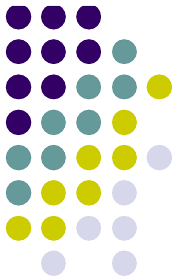
4.10.3 Practical Space Diversity Considerations



general block diagram of a space diversity scheme



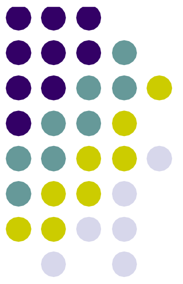
4.10.3 Practical Space Diversity Considerations



Space diversity reception methods can be classified into four categories

- 1. Selection diversity
- 2. Feedback diversity
- 3. Maximal ratio combining
- 4. Equal gain diversity

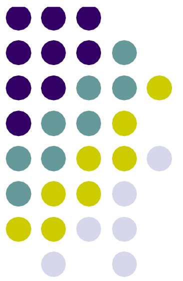
4.10.3 Practical Space Diversity Considerations



(1) Selection Diversity

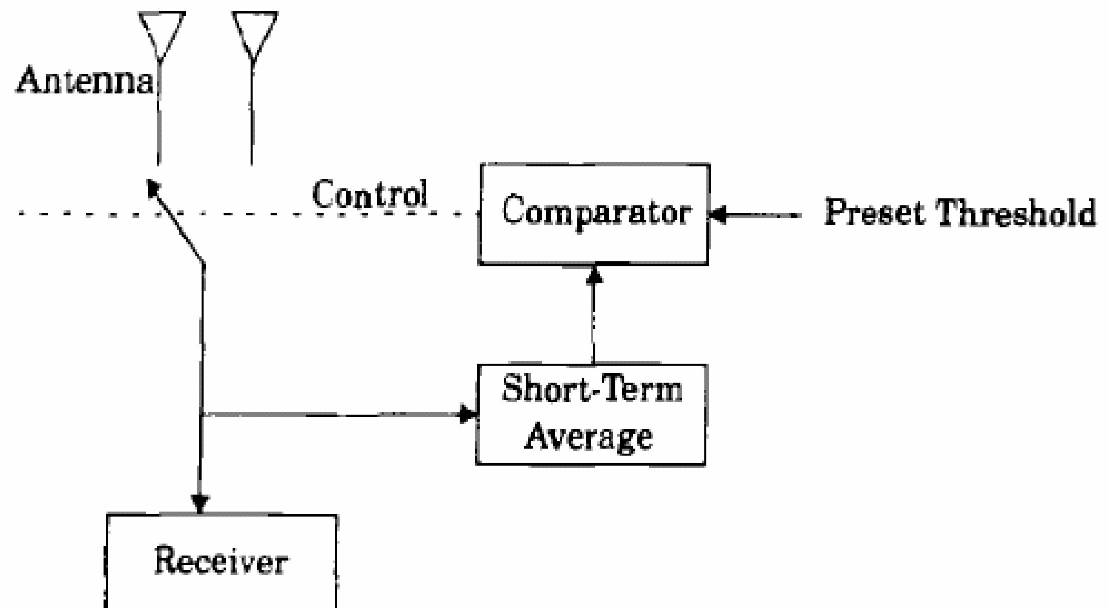
- The simplest diversity technique.
- The receiver branch having the highest instantaneous SNR is connected to the demodulator.
- The antenna signals themselves could be sampled and the best one sent to a single demodulator.
- In practice, the branch with the largest (S + N) /N is used, since it is difficult to measure SNR.
- A practical selection diversity system cannot function on a truly instantaneous basis, but must be designed so that the internal time constants of the selection circuitry are shorter than the reciprocal of the signal fading rate.

4.10.3 Practical Space Diversity Considerations

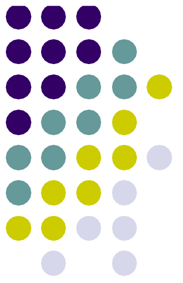


(2) Feedback or Scanning Diversity

- Very **similar** to selection diversity
- The M signals are **scanned** in a fixed sequence until one is found to be above a predetermined threshold.
- This signal is then received until it falls **below threshold** and the scanning process is again initiated.
- The resulting fading statistics are somewhat **inferior** to those obtained by the other methods.
- **Advantage**: very simple to implement (only one receiver is required).



4.10.3 Practical Space Diversity Considerations

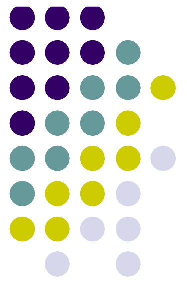


(3) Maximal Ratio Combining

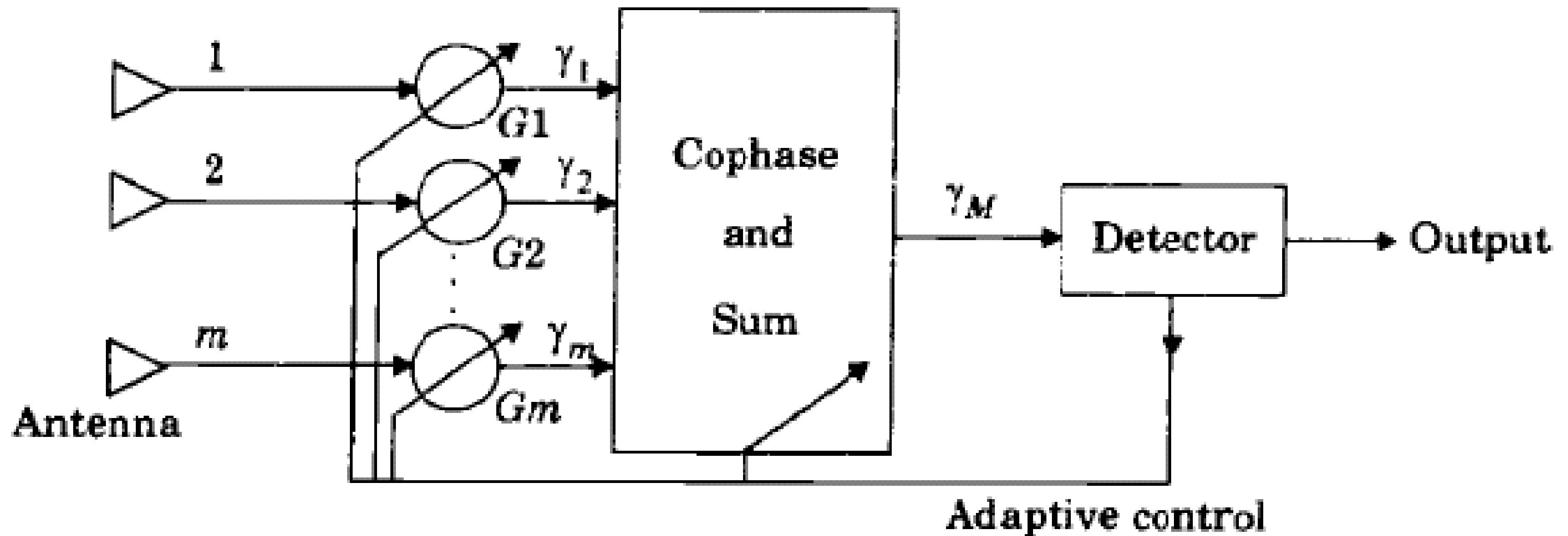
- **The signals from all of the M branches are weighted and then summed.**
- **The individual signals must be co-phased before being summed.**

  requires an individual receiver and phasing circuit for each antenna element.
- **Output SNR equal to the sum of the individual SNRs.**
- **Advantage: produces an output with an acceptable SNR even when none of the individual signals are themselves acceptable.**
- **Gives the best statistical reduction of fading of any known linear diversity combiner.**

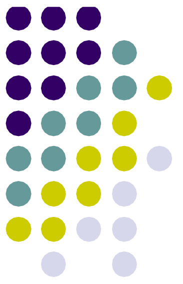
4.10.3 Practical Space Diversity Considerations



Maximal Ratio Combiner



4.10.3 Practical Space Diversity Considerations

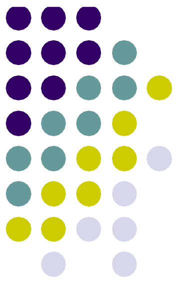


(4) Equal Gain Combining

In certain cases, it is not convenient to provide for the variable weighting capability required for true maximal ratio combining. In such cases, the branch

- **Equal gain combining diversity sets all weights to unity but the signals from each branch are co-phased.**
- **The possibility of producing an acceptable signal from a number of unacceptable inputs is still retained,**
- **The performance is only marginally inferior to maximal ratio combining and superior to selection diversity.**

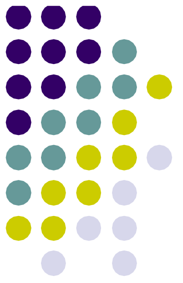
4.10. 4 PolarIzation Diversity



At the base station, space diversity is considerably less practical .

- polarization diversity only provides two diversity branches, but allows the antenna elements to be co-located.
- Measured horizontal and vertical polarization paths between a mobile and a base station are reported to be uncorrelated.
- Decorrelation for the signals in each polarization is caused by multiple reflections.
- The reflection coefficient for each polarization is different, which results in different amplitudes and phases for each, or at least some, of the reflections.
- After sufficient random reflections, the polarization state of the signal will be independent of the transmitted polarization.

In practice, however, there is some dependence of the received polarization on the transmitted polarization.



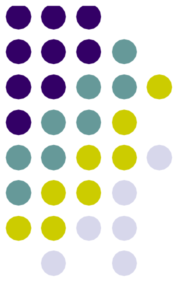
4.10.5 Frequency Diversity

- **Transmits information on more than one carrier frequency.**
  frequencies separated by more than the coherence bandwidth of the channel will not experience the same fades.

- **Frequency diversity is often employed in microwave LOS links.**

- In practice, 1:N protection switching is provided by a radio licensee,
  When diversity is needed, the appropriate traffic is simply switched to the backup frequency.

- **Disadvantage: not only requires spare bandwidth but also requires that there be as many receivers as there are channels used for the frequency diversity.**
  for critical traffic, the expense may be justified.

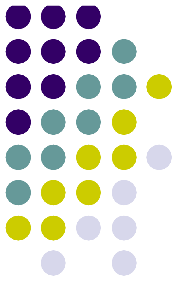


4.10.5 Frequency Diversity

- **New OFDM modulation and access techniques exploit frequency diversity by providing simultaneous modulation signals with error control coding across a large bandwidth.**

- **If a particular frequency undergoes a fade, the composite signal will still be demodulated.**

4.10.6 Time Diversity

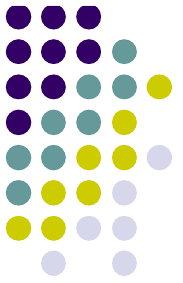


- **Time diversity repeatedly transmits information at time spacings that exceed the coherence time of the channel**

  Multiple repetitions of the signal will be received with independent fading conditions.

- **One modem implementation of time diversity involves the use of the RAKE receiver for spread spectrum CDMA, where the multipath channel provides redundancy in the transmitted message.**

4.11 RAKE Receiver



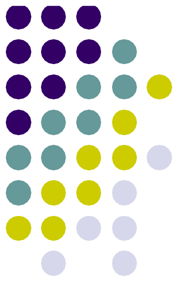
- In CDMA spread spectrum systems, the spreading codes are designed to provide very low correlation between successive chips.

- If the multipath components are delayed in time by more than a chip duration, they appear like uncorrelated noise at a CDMA receiver, and equalization is not required.

- However, since there is useful information in the multipath components, CDMA receivers may combine the time delayed versions of the original signal transmission in order to improve the signal to noise ratio at the receiver

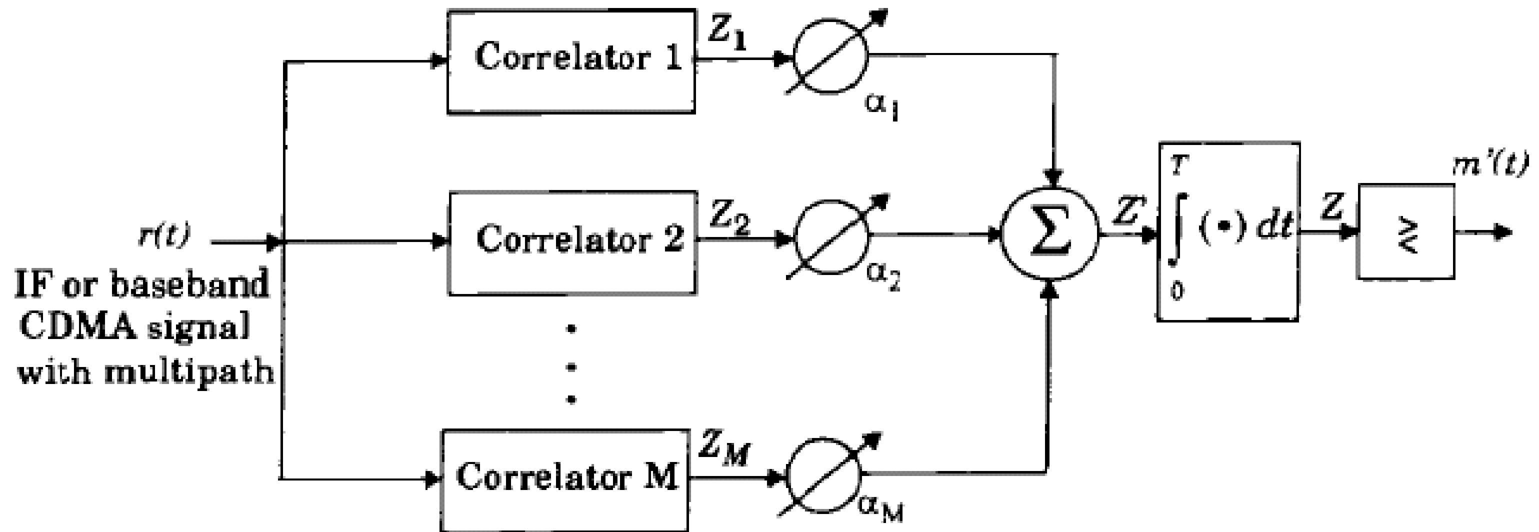
- A RAKE is employed to do this:

  It attempts to collect the time-shifted versions of the original signal by providing a separate correlation receiver for each of the multipath signals.

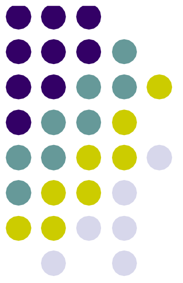


4.11 RAKE Receiver

The RAKE receiver is essentially a diversity receiver designed specifically for CDMA, where the diversity is provided by the fact that the multipath components are practically uncorrelated from one another when their relative propagation delays exceed a chip period.



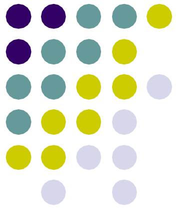
An M branch (M-finger) RAKE receiver implementation. Each correlator detects a time shifted version of the original CDMA transmission, and each finger of the RAKE correlates to a portion of the signal which is delayed by at least one chip in time from the other fingers.



UNIT V

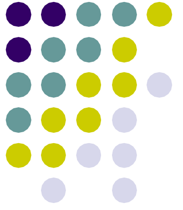
WIRELESS NETWOKS

Wireless?



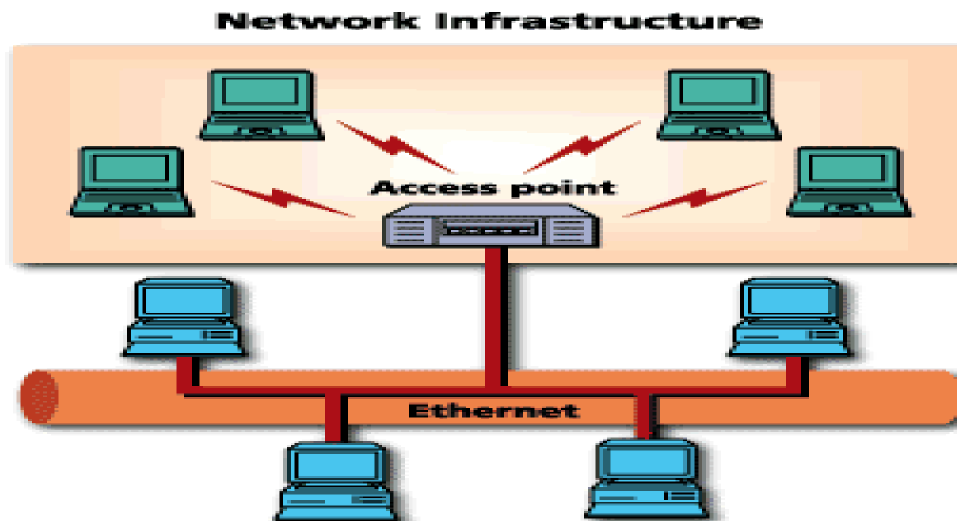
- A wireless LAN or WLAN is a wireless local area
network that uses radio waves as its carrier.
- The last link with the users is wireless, to give a network connection to all users in a building or campus.
- The backbone network usually uses cables

Common Topologies

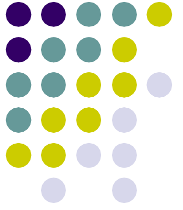


The wireless LAN connects to a wired LAN

- There is a need of an access point that bridges wireless LAN traffic into the wired LAN.
- The access point (AP) can also act as a repeater for wireless nodes, effectively doubling the maximum possible distance between nodes.



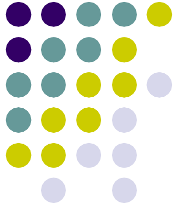
Common Topologies



Complete Wireless Networks

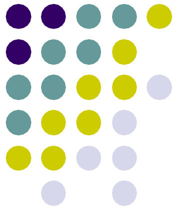
- The physical size of the network is determined by the maximum reliable
  propagation range of the radio signals.
- Referred to as **ad hoc** networks
- Are self-organizing networks without any centralized control
- Suited for temporary situations such as meetings and conferences.

How do wireless LANs work?



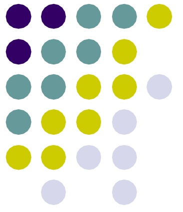
Wireless LANs operate in almost the same way as wired LANs, using the same networking protocols and supporting the most of the same applications.

How are WLANs Different?



- They use specialized **physical and data link** protocols
- They integrate into existing networks through **access points** which provide a bridging function
- They let you stay connected as you **roam** from one coverage area to another
- They have unique **security** considerations
- They have specific **interoperability** requirements
- They require **different hardware**
- They offer **performance** that differs from wired LANs.

Physical and Data Link Layers



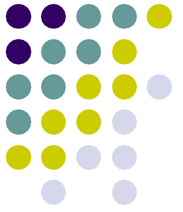
Physical Layer:

- The wireless **NIC** takes **frames** of data from the link layer, scrambles the data in a predetermined way, then uses the modified data stream to modulate a **radio carrier signal**.

Data Link Layer:

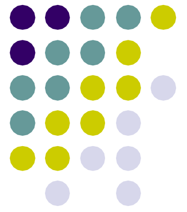
- Uses **C**arriers-**S**ense-**M**ultiple-**A**ccess with **C**ollision **A**voidance (CSMA/CA).

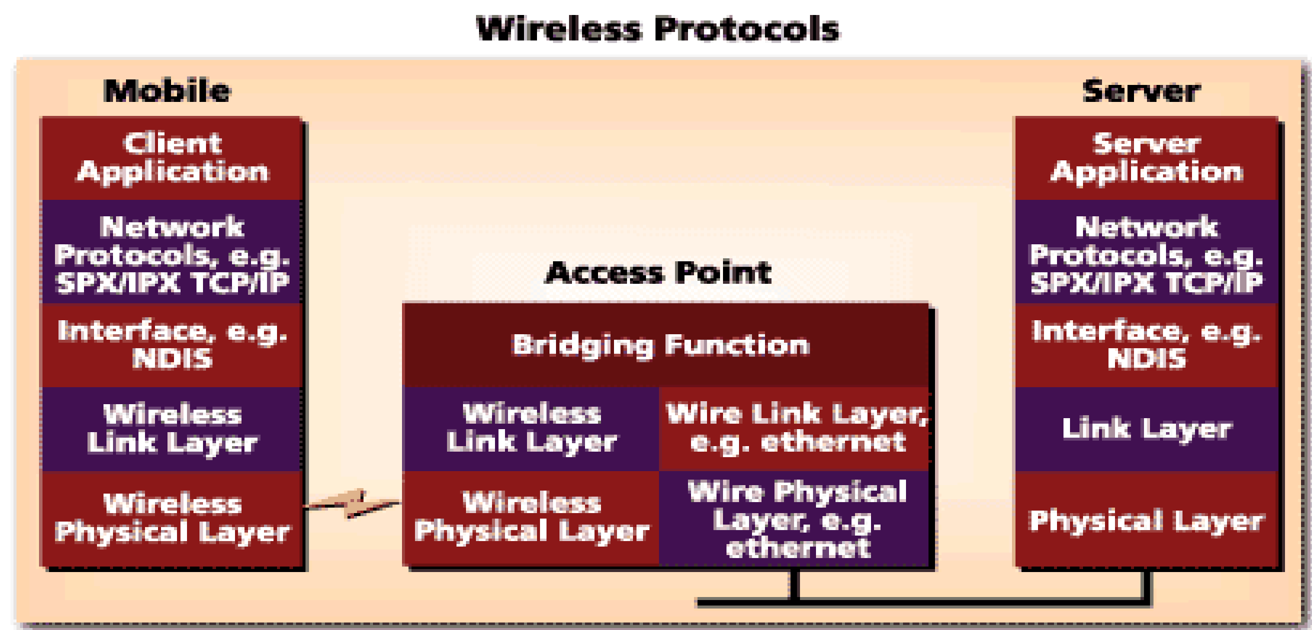
Integration With Existing Networks



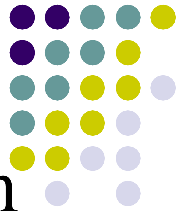
- Wireless Access Points (APs) - a small device that bridges wireless traffic to your network.
- Most access points bridge wireless LANs into Ethernet networks, but Token-Ring options are available as well.

Integration With Existing Networks

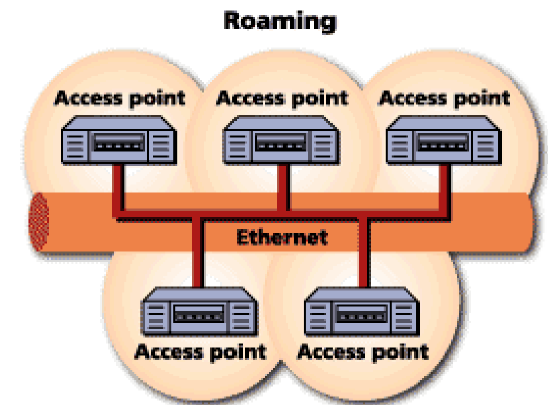




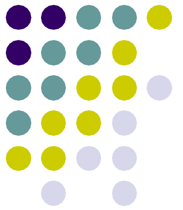
Roaming



- Users maintain a continuous connection as they roam from one physical area to another
- Mobile nodes automatically register with the new access point.
- Methods: DHCP, Mobile IP
- IEEE 802.11 standard does not address roaming, you may need to purchase equipment from one vendor if your users need to roam from one access point to another.



Security

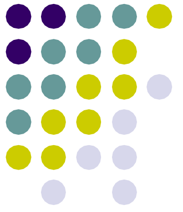


- In theory, spread spectrum radio signals are inherently difficult to decipher without knowing the exact hopping sequences or direct sequence codes used
- The IEEE 802.11 standard specifies optional security called "**W**ired **E**quivalent **P**rivacy" whose goal is that a wireless LAN offer privacy equivalent to that offered by a wired LAN. The standard also specifies optional authentication measures.

Interoperability

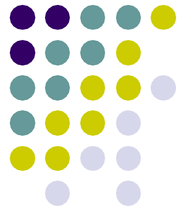
- Before the IEEE 802.11 interoperability was based on cooperation between vendors.
- IEEE 802.11 only standardizes the physical and medium access control layers.
- Vendors must still work with each other to ensure their IEEE 802.11 implementations interoperate
- Wireless Ethernet Compatibility Alliance (WECA) introduces the Wi-Fi Certification to ensure cross-vendor interoperability of 802.11b solutions

Hardware



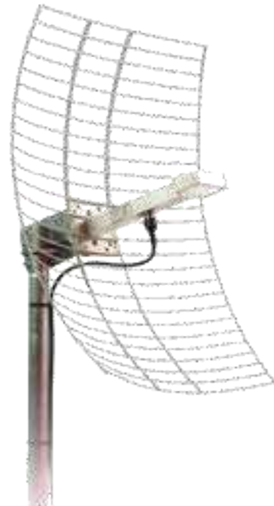
- PC Card, either with integral antenna or with external antenna/RF module.
- ISA Card with external antenna connected by cable.
- Handheld terminals
- Access points

Hardware



CISCO Aironet 350 series

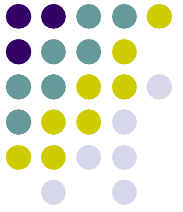
Wireless Handheld Terminal



Semi Parabolic Antenna

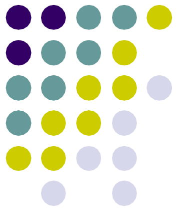
BreezeCOM AP

Performance



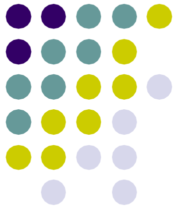
- **802.11a** offers speeds with a theoretically maximum rate of 54Mbps in the 5 GHz band
- **802.11b** offers speeds with a theoretically maximum rate of 11Mbps at in the 2.4 GHz spectrum band
- **802.11g** is a new standard for data rates of up to a theoretical maximum of 54 Mbps at 2.4 GHz.

What is 802.11?



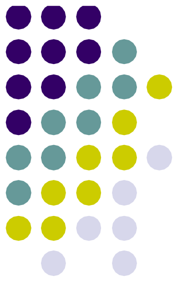
- A family of wireless LAN (WLAN) specifications developed by a working group at the Institute of Electrical and Electronic Engineers (IEEE)
- Defines standard for WLANs using the following four technologies
  - Frequency Hopping Spread Spectrum (FHSS)
  - Direct Sequence Spread Spectrum (DSSS)
  - Infrared (IR)
  - Orthogonal Frequency Division Multiplexing (OFDM)
- Versions: 802.11a, 802.11b, 802.11g, 802.11e, 802.11f, 802.11i

802.11 - Transmission



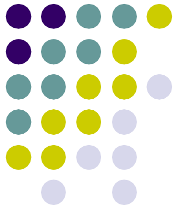
- Most wireless LAN products operate in unlicensed radio bands
  - 2.4 GHz is most popular
  - Available in most parts of the world
  - No need for user licensing
- Most wireless LANs use spread-spectrum radio
  - Resistant to interference, secure
  - Two popular methods
    - Frequency Hopping (FH)
    - Direct Sequence (DS)

Frequency Hopping Vs. Direct Sequence



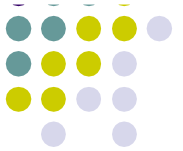
- FH systems use a radio carrier that "hops" from frequency to frequency in a pattern known to both transmitter and receiver
  - Easy to implement
  - Resistance to noise
  - Limited throughput (2-3 Mbps @ 2.4 GHz)
- DS systems use a carrier that remains fixed to a specific frequency band. The data signal is spread onto a much larger range of frequencies (at a much lower power level) using a specific encoding scheme.
  - Much higher throughput than FH (11 Mbps)
  - Better range
  - Less resistant to noise (made up for by redundancy – it transmits at least 10 fully redundant copies of the original signal at the same time)

802.11a



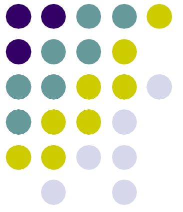
- Employs Orthogonal Frequency Division Multiplexing (OFDM)
  - Offers higher bandwidth than that of 802.11b, DSSS (Direct Sequence Spread Spectrum)
  - 802.11a MAC (Media Access Control) is same as 802.11b
- Operates in the 5 GHz range

802.11a Advantages



- Ultra-high spectrum efficiency
  - 5 GHz band is 300 MHz (vs. 83.5 MHz @ 2.4 GHz)
  - More data can travel over a smaller amount of bandwidth
- High speed
  - Up to 54 Mbps
- Less interference
  - Fewer products using the frequency
    - 2.4 GHz band shared by cordless phones, microwave ovens, Bluetooth, and WLANs

802.11a Disadvantages

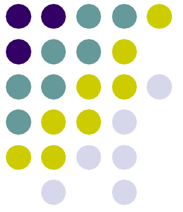


- Standards and Interoperability
  - Standard not accepted worldwide
  - No interoperability certification available for 802.11a products
  - Not compatible or interoperable with 802.11b
- Legal issues
  - License-free spectrum in 5 GHz band not available worldwide
- Market
  - Beyond LAN-LAN bridging, there is limited interest for 5 GHz adoption

802.11a Disadvantages

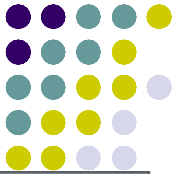
- Cost
  - 2.4 GHz will still has >40% cost advantage
- Range
  - At equivalent power, 5 GHz range will be ~50% of 2.4 GHz
- Power consumption
  - Higher data rates and increased signal require more power
    - OFDM is less power-efficient then DSSS

802.11a Applications



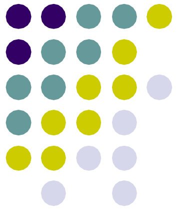
- Building-to-building connections
- Video, audio conferencing/streaming video, and audio
- Large file transfers, such as engineering CAD drawings
- Faster Web access and browsing
- High worker density or high throughput scenarios
  - Numerous PCs running graphics-intensive applications

802.11a Vs. 802.11b



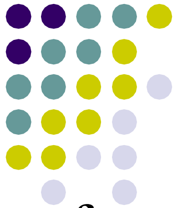
802.11a vs. 802.11b	802.11a	802.11b
Raw data rates	Up to 54 Mbps (54, 48, 36, 24,18, 12 and 6 Mbps)	Up to 11 Mbps (11, 5.5, 2, and 1 Mbps)
Range	50 Meters	100 Meters
Bandwidth	UNII and ISM (5 GHz range)	ISM (2.4000—2.4835 GHz range)
Modulation	OFDM technology	DSSS technology

802.11g



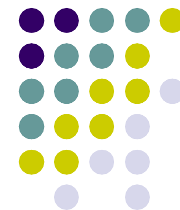
- 802.11g is a high-speed extension to 802.11b
  - Compatible with 802.11b
  - High speed up to 54 Mbps
  - 2.4 GHz (vs. 802.11a, 5 GHz)
  - Using ODFM for backward compatibility
  - Adaptive Rate Shifting

802.11g Advantages



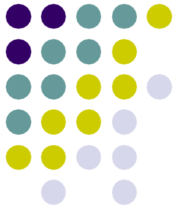
- Provides higher speeds and higher capacity requirements for applications
  - Wireless Public Access
- Compatible with existing 802.11b standard
- Leverages Worldwide spectrum availability in 2.4 GHz
- Likely to be less costly than 5 GHz alternatives
- Provides easy migration for current users of 802.11b WLANs
  - Delivers backward support for existing 802.11b products
- Provides path to even higher speeds in the future

802.11e Introduces Quality of Service



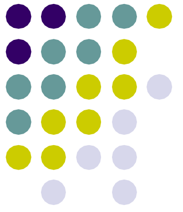
- Also know as P802.11 TGe
- Purpose:
  - To enhance the 802.11 Medium Access Control (MAC) to improve and manage Quality of Service (QoS)
- Cannot be supported in current chip design
- Requires new radio chips
  - Can do basic QoS in MAC layer

802.11f – Inter Access Point Protocol

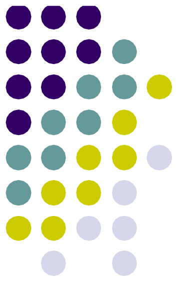


- Also know as P802.11 TGf
- Purpose:
  - To develop a set of requirements for Inter-Access Point Protocol (IAPP), including operational and management aspects

802.11b Security Features



- Wired Equivalent Privacy (**WEP**) – A protocol to protect link-level data during wireless transmission between clients and access points.
- Services:
  - **Authentication**: provides access control to the network by denying access to client stations that fail to authenticate properly.
  - **Confidentiality**: intends to prevent information compromise from casual eavesdropping
  - **Integrity**: prevents messages from being modified while in transit between the wireless client and the access point.



END

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

EC3551- Transmission Lines and RF Systems

Semester - 05

Notes

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

Vision

To excel in providing value based education in the field of Electronics and Communication Engineering, keeping in pace with the latest technical developments through commendable research, to raise the intellectual competence to match global standards and to make significant contributions to the society upholding the ethical standards.

Mission

- ✓ To deliver Quality Technical Education, with an equal emphasis on theoretical and practical aspects.
- ✓ To provide state of the art infrastructure for the students and faculty to upgrade their skills and knowledge.
- ✓ To create an open and conducive environment for faculty and students to carry out research and excel in their field of specialization.
- ✓ To focus especially on innovation and development of technologies that is sustainable and inclusive, and thus benefits all sections of the society.
- ✓ To establish a strong Industry Academic Collaboration for teaching and research, that could foster entrepreneurship and innovation in knowledge exchange.
- ✓ To produce quality Engineers who uphold and advance the integrity, honour and dignity of the engineering.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

1. To provide the students with a strong foundation in the required sciences in order to pursue studies in Electronics and Communication Engineering.
2. To gain adequate knowledge to become good professional in electronic and communication engineering associated industries, higher education and research.
3. To develop attitude in lifelong learning, applying and adapting new ideas and technologies as their field evolves.
4. To prepare students to critically analyze existing literature in an area of specialization and ethically develop innovative and research oriented methodologies to solve the problems identified.
5. To inculcate in the students a professional and ethical attitude and an ability to visualize the engineering issues in a broader social context.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO1: Design, develop and analyze electronic systems through application of relevant electronics, mathematics and engineering principles.

PSO2: Design, develop and analyze communication systems through application of fundamentals from communication principles, signal processing, and RF System Design & Electromagnetics.

PSO3: Adapt to emerging electronics and communication technologies and develop innovative solutions for existing and newer problems.

1. Transmission Line Theory

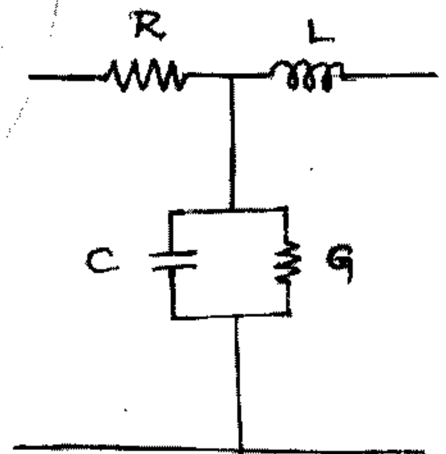
Transmission Line :

Transmission Line is a conductive method of guiding electrical signal from one end to another end.

Types :

1. Parallel lines → Open wire line; → Two wire parallel line
2. Twisted Pair cable → shielded; → unshielded
3. Flat Ribbon cable
4. coaxial cable
5. Strip lines

Equivalent circuit of Txn. Line :



R → Resistance (ohm/unit length)

L → Inductance (Henry/unit Length)

G → Conductance (mho/unit Length)

C → capacitance (Farad/unit Length)

R, L, G and C are called as primary constants of Txn. Lines.

Uniform

Uniform Txn. Line :

when R, L, G and C are uniformly distributed through out the line then the line is called as Uniform Txn. Line.

Secondary constants of Txn. line

1. characteristic Impedance, 'Z_0'

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

2. Propagation Constant, 'γ'

$$\gamma = P = \sqrt{ZY} = \sqrt{(R + j\omega L)(G + j\omega C)}$$

$$\gamma = \alpha + j\beta$$

3. Wavelength, 'λ'

$$\lambda = \frac{c}{f} \quad \text{(or)} \quad \lambda = \frac{2\pi}{\beta}$$

4. Phase Velocity, 'V_p'

$$V_p = \lambda f = \frac{\omega}{\beta}$$

To find 1° constants from 2° constants

$$R + j\omega L = \gamma \cdot Z_0$$

$$G + j\omega C = \gamma / Z_0$$

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R + j\omega L}{G + j\omega C} \times \frac{R + j\omega L}{R + j\omega L}}$$

$$= \sqrt{\frac{(R + j\omega L)^2}{(R + j\omega L)(G + j\omega C)}}$$

$$= \frac{R + j\omega L}{\sqrt{(R + j\omega L)(G + j\omega C)}}$$

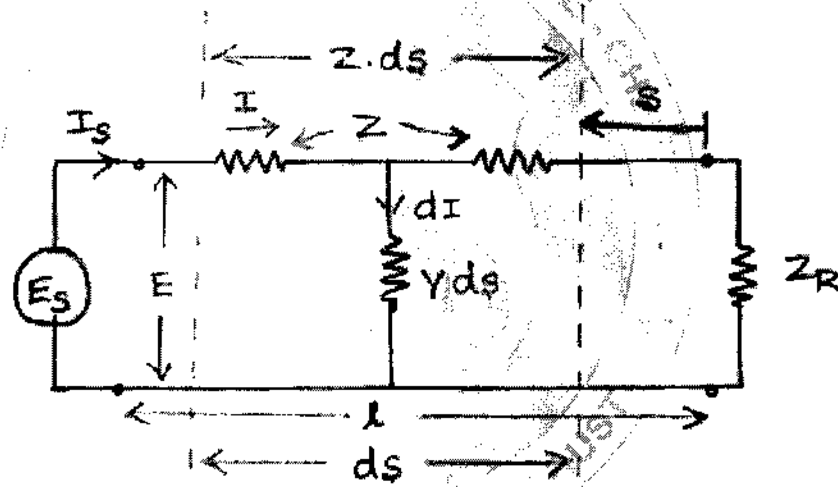
$$Z_0 = \frac{R + j\omega L}{\gamma}$$

$$\therefore R + j\omega L = Z_0 \cdot \gamma$$

General Solution Of Transmission Line :

When the voltage or current is transmitted through a transmission line, it will not be constant through out the line. There will be drop in the voltage or current.

To find the voltage and current at any point in a transmission line, Let us derive general solution of transmission line.



Let R, L, G and C be the Primary constants of transmission line.

R → Series Resistance (ohms / unit length)

L → Series Inductance (Henry / unit length)

G → Shunt Conductance (mho / unit length)

C → Shunt Capacitance (Farad / unit length)

l → Total length of the Txn. line

s → Distance from load to the point of Observati-on

ds → Small section of Txn. line

Z → Series Impedance

y → Shunt Admittance

E_S → Source Voltage

I_S → Source Current

Z_R → Impedance at Receiving end

E → Voltage at any point on the line

I → Current at any Point on the line

dE → Voltage drop in ds section

dI → Current drop in ds section

$z \cdot ds$ → Impedance of small section ds

$y \cdot ds$ → Admittance of Small section ds

Consider a small section 'ds' having the Series impedance 'zds'. Let the current flowing through this section be 'I' then the voltage drop across this section will be

$$dE = I \cdot zds$$

$$\frac{dE}{ds} = Iz \longrightarrow (1)$$

Similarly, the current drop across this section will be

$$dI = E \cdot yds$$

$$\frac{dI}{ds} = Ey \longrightarrow (2)$$

Differentiate equs. (1) & (2) w.r.to 's'

$$\frac{d^2E}{ds^2} = \frac{dI}{ds} \cdot z$$

$$\frac{d^2E}{ds^2} = Eyz \longrightarrow (3)$$

$E \leftrightarrow I$

$z \leftrightarrow y$

Similarly, $\frac{d^2I}{ds^2} = Iyz \rightarrow (4)$

Equations (3) and (4) are called the differential equations of Txn. Line.

To find solution for the differential equation, put $\frac{d}{ds} = m$, then equs. (3) & (4) becomes

$m^2E = Eyz$ $\qquad$ $m^2I = Iyz$

$m^2 = yz$ $\qquad$ $m^2 = yz$

$m = \pm\sqrt{yz} \longrightarrow (5a)$ $\qquad$ $m = \pm\sqrt{yz} \longrightarrow (5b)$

$\therefore$ The solution of the differential equations (3) & (4) are,

$$E = A \cdot e^{\sqrt{zy} \cdot s} + B \cdot e^{-\sqrt{zy} \cdot s} \longrightarrow (6a)$$

$$I = C \cdot e^{\sqrt{zy} \cdot s} + D \cdot e^{-\sqrt{zy} \cdot s} \longrightarrow (6b)$$

The voltage and current at the Receiver end is, put $I = I_R$, $E = E_R$ and $s = 0$ in (6a) & (6b)

$$E_R = A + B \longrightarrow (7a)$$

$$I_R = C + D \longrightarrow (7b)$$

Now, differentiate equs. (6a) & (6b) w.r.to 's'

$$\frac{dE}{ds} = A \cdot \sqrt{zy} \cdot e^{\sqrt{zy} \cdot s} - B \cdot \sqrt{zy} \cdot e^{-\sqrt{zy} \cdot s}$$

$$Iz = A \cdot \sqrt{zy} \cdot e^{\sqrt{zy} \cdot s} - B \cdot \sqrt{zy}\, e^{-\sqrt{zy} \cdot s}$$

$$I = A\cdot\sqrt{\frac{Y}{Z}}\cdot e^{\sqrt{ZY}\cdot s} - B\cdot\sqrt{\frac{Y}{Z}}\cdot e^{-\sqrt{ZY}\cdot s} \rightarrow (8a)$$

Similarly,

$$E = C\cdot\sqrt{\frac{Z}{Y}}\cdot e^{\sqrt{ZY}\cdot s} - D\cdot\sqrt{\frac{Z}{Y}}\cdot e^{-\sqrt{ZY}\cdot s} \rightarrow (8b)$$

The voltage and current at Receiver end is,

put $I = I_R$, $E = E_R$ and $s = 0$ in equs. (8a) & (8b)

$$I_R = A\cdot\sqrt{\frac{Y}{Z}} - B\cdot\sqrt{\frac{Y}{Z}} \rightarrow (9a)$$

$$E_R = C\cdot\sqrt{\frac{Z}{Y}} - D\cdot\sqrt{\frac{Z}{Y}} \rightarrow (9b)$$

To find the arbitary constants A, B, C and D

Solve (7a), (7b), (9a) & (9b)

$$E_R = A + B \quad \text{(from equ. (7a))}$$

$$\sqrt{\frac{Z}{Y}}\cdot I_R = A - B \quad \text{(from equ. (9a))}$$

$$E_R + \sqrt{\frac{Z}{Y}}\cdot I_R = 2A$$

$$A = \frac{E_R}{2} + \sqrt{\frac{Z}{Y}}\cdot\frac{I_R}{2} \qquad \because \sqrt{\frac{Z}{Y}} = Z_0$$

$$A = \frac{E_R}{2} + Z_0\cdot\frac{E_R}{2Z_R}$$

$$A = \frac{E_R}{2}\left(1 + \frac{Z_0}{Z_R}\right) \rightarrow (10a) \qquad \because I_R = \frac{E_R}{Z_R}$$

Similarly,

$$B = \frac{E_R}{2}\left(1 - \frac{Z_0}{Z_R}\right) \rightarrow (10b)$$

$$C = \frac{I_R}{2}\left(1 + \frac{Z_R}{Z_0}\right) \rightarrow (10c)$$

$$D = \frac{I_R}{2}\left(1 - \frac{Z_R}{Z_0}\right) \rightarrow (10d)$$

sub equs. (10a) → (10d) in equs. (6a) & (6b)

$$E = A \cdot e^{\sqrt{zy} \cdot s} + B \cdot e^{\sqrt{zy} \cdot s} \qquad (6a)$$

$$E = \frac{E_R}{2}\left(1 + \frac{Z_0}{Z_R}\right) e^{\sqrt{zy} \cdot s} + \frac{E_R}{2}\left(1 - \frac{Z_0}{Z_R}\right) e^{-\sqrt{zy} \cdot s} \rightarrow (11a)$$

$$E = \frac{E_R}{2}\left(1 + \frac{Z_0}{Z_R}\right)\left[e^{\sqrt{zy} \cdot s} + \frac{\left(1 - \frac{Z_0}{Z_R}\right)}{\left(1 + \frac{Z_0}{Z_R}\right)} \cdot e^{-\sqrt{zy} \cdot s}\right]$$

$$E = \frac{E_R}{2}\left(1 + \frac{Z_0}{Z_R}\right)\left[e^{\sqrt{zy} \cdot s} + \frac{\left(\frac{Z_R - Z_0}{Z_R}\right)}{\left(\frac{Z_R + Z_0}{Z_R}\right)} \cdot e^{-\sqrt{zy} \cdot s}\right]$$

$$E = \frac{E_R}{2}\left(\frac{Z_R + Z_0}{Z_R}\right)\left[e^{\sqrt{zy} \cdot s} + \left(\frac{Z_R - Z_0}{Z_R + Z_0}\right) e^{-\sqrt{zy} \cdot s}\right]$$

$$E = \frac{E_R}{2}\left(\frac{Z_R + Z_0}{Z_R}\right)\left[e^{\sqrt{zy} \cdot s} + K\, e^{-\sqrt{zy} \cdot s}\right] \rightarrow (12)$$

where, $K = \dfrac{Z_R - Z_0}{Z_R + Z_0}$; Reflection coefficient

Similarly,

$$I = \frac{I_R}{2}\left(\frac{Z_R + Z_0}{Z_0}\right)\left[e^{\sqrt{zy} \cdot s} - K \cdot e^{-\sqrt{zy} \cdot s}\right] \rightarrow (13)$$

Equs. (12) & (13) are the useful equs. of Txn. Line.

Eqn. (11a) can also be solved as,

$$E = \frac{E_R}{2}\left(1 + \frac{Z_0}{Z_R}\right) e^{\sqrt{zy}.s} + \frac{E_R}{2}\left(1 - \frac{Z_0}{Z_R}\right) e^{-\sqrt{zy}.s} \rightarrow (11a)$$

$$E = \frac{E_R}{2} e^{\sqrt{zy}.s} + \frac{E_R}{2}\frac{Z_0}{Z_R} . e^{\sqrt{zy}.s} + \frac{E_R}{2} . e^{-\sqrt{zy}.s} - \frac{E_R}{2}\frac{Z_0}{Z_R} . e^{-\sqrt{zy}.s}$$

$$E = \frac{E_R}{2}\left(e^{\sqrt{zy}.s} + e^{-\sqrt{zy}.s}\right) + \frac{I_R . Z_0}{2}\left(e^{\sqrt{zy}.s} - e^{-\sqrt{zy}.s}\right)$$

$$E = E_R . \left(\frac{e^{\sqrt{zy}.s} + e^{-\sqrt{zy}.s}}{2}\right) + I_R . Z_0 \left(\frac{e^{\sqrt{zy}.s} - e^{-\sqrt{zy}.s}}{2}\right)$$

$$\therefore E = E_R . \cosh . \sqrt{zy} . s + I_R . Z_0 . \sinh \sqrt{zy} . s \rightarrow (14)$$

Similarly,

$$I = I_R . \cosh \sqrt{zy} . s + \frac{E_R}{Z_0} . \sinh \sqrt{zy} . s \rightarrow (15)$$

Input Impedance of Transmission Line :

* The impedance measured at the i/p end of the Txn. line is the i/p impedance.

* The i/p impedance is defined as the ratio of source voltage to source current.

* It is denoted by Z_s, Z_{in}.

$$Z_S = \frac{E_S}{I_S} \rightarrow (1)$$

W.K.T the voltage and current at any point on a line is given by,

$$E = E_R \cosh\sqrt{ZY}.s + I_R Z_0 \sinh\sqrt{ZY}.s \rightarrow (2)$$

$$I = I_R \cosh\sqrt{ZY}.s + \frac{E_R}{Z_0} \sinh\sqrt{ZY}.s \rightarrow (3)$$

To find voltage & current at source end,

Put $E = E_s$, $I = I_s$ and $s = l$ in equs. (2) & (3)

$$E_S = E_R \cosh\sqrt{ZY}.l + I_R Z_0 \sinh\sqrt{ZY}.l \rightarrow (4)$$

$$I_S = I_R \cosh\sqrt{ZY}.l + \frac{E_R}{Z_0}.\sinh\sqrt{ZY}.l \rightarrow (5)$$

Sub (4) & (5) in (1)

$$\therefore Z_S = \frac{E_R \cosh\sqrt{ZY}.l + I_R Z_0 \sinh\sqrt{ZY}.l}{I_R \cosh\sqrt{ZY}.l + \frac{E_R}{Z_0}.\sinh\sqrt{ZY}.l}$$

$$Z_S = \frac{I_R Z_R \cosh\sqrt{ZY}.l + I_R Z_0 \sinh\sqrt{ZY}.l}{I_R \cosh\sqrt{ZY}.l + I_R Z_R \sinh\sqrt{ZY}.l} \quad (\because E_R = I_R Z_R$$

$$Z_S = \frac{I_R\left(Z_R \cosh\sqrt{zy}.l + Z_0 \sinh\sqrt{zy}.l\right)}{I_R\left(\cosh\sqrt{zy}.l + \frac{Z_R}{Z_0}\sinh\sqrt{zy}.l\right)}$$

$$Z_S = \frac{Z_R \cosh\sqrt{zy}.l + Z_0 \sinh\sqrt{zy}.l}{\cosh\sqrt{zy}.l + \frac{Z_R}{Z_0}\sinh\sqrt{zy}.l}$$

$$\therefore Z_S = Z_0 \cdot \left[\frac{Z_R \cosh\sqrt{zy}.l + Z_0 . \sinh\sqrt{zy}.l}{Z_0 . \cosh\sqrt{zy}.l + Z_R . \sinh\sqrt{zy}.l}\right] \rightarrow (6)$$

<u>Case i): I/p impedance with $Z_R = Z_0$</u>

I/p impedance of a Txn. line terminated with 'Z_0' is given as,

substitute $Z_R = Z_0$ in equ. (6)

$$Z_S = Z_0 \cdot \left[\frac{Z_0 \cosh\sqrt{zy}.l + Z_0 \sinh\sqrt{zy}.l}{Z_0 \cosh\sqrt{zy}.l + Z_0 \sinh\sqrt{zy}.l}\right]$$

$$\therefore \boxed{Z_S = Z_0} \rightarrow (7)$$

<u>Case ii): I/p impedance with short circuit end ($Z_R = 0$)</u>

I/p impedance of a Txn. line with short circuit end is given as,

put $Z_R = 0$ in equ. (6)

$$Z_S = Z_0 \left[\frac{0 + Z_0 \sinh\sqrt{zy}.l}{Z_0 \cosh\sqrt{zy}.l + 0}\right]$$

$$Z_S = Z_0 \cdot \left[\frac{Z_0 \cdot \sinh\sqrt{zy}\cdot l}{Z_0 \cdot \cosh\sqrt{zy}\cdot l}\right]$$

$$Z_S = Z_{SC} = Z_0 \cdot \tanh\sqrt{zy}\cdot l \longrightarrow (8)$$

case iii) : I/p impedance with Open circuit end ($Z_R = \infty$)

I/p impedance of a Txn. line with open circuit end is given as,

In equ. (6) take 'Z_R' as common

$$Z_S = Z_0 \cdot \frac{Z_R}{Z_R}\left[\frac{\cosh\sqrt{zy}\cdot l + \frac{Z_0}{Z_R}\sinh\sqrt{zy}\cdot l}{\frac{Z_0}{Z_R}\cosh\sqrt{zy}\cdot l + \sinh\sqrt{zy}\cdot l}\right]$$

Put $Z_R = \infty$ $\quad\left(\because \frac{1}{\infty} = 0\right)$

$$Z_S = Z_0 \cdot \left[\frac{\cosh\sqrt{zy}\cdot l + 0}{0 + \sinh\sqrt{zy}\cdot l}\right]$$

$$Z_S = Z_0 \cdot \left[\frac{\cosh\sqrt{zy}\cdot l}{\sinh\sqrt{zy}\cdot l}\right]$$

$$\therefore Z_S = Z_{OC} = Z_0 \cdot \coth\sqrt{zy}\cdot l \longrightarrow (9)$$

Transfer Impedance: 'Z_T'

It is defined as the ratio of source voltage to Receiver current. It is denoted by 'Z_T'

$$Z_T = \frac{E_S}{\ldots} \rightarrow (1)$$

W.K.T The voltage at any point on a Txn. line is given as, $E = E_R \cosh\sqrt{zy}.s + I_R . z_0 . \sinh\sqrt{zy}.s \rightarrow (2)$

To find the voltage at source end,

put $E = E_S$ and $s = l$ in equ. (2)

$$\therefore E_S = E_R . \cosh\sqrt{zy}.l + I_R z_0 . \sinh\sqrt{zy}.l$$

$$= I_R Z_R \cosh\sqrt{zy}.l + I_R . z_0 . \sinh\sqrt{zy}.l$$

$$E_S = I_R . (z_R . \cosh\sqrt{zy}.l + z_0 . \sinh\sqrt{zy}.l)$$

$$\frac{E_S}{I_R} = z_T = z_R . \cosh\sqrt{zy}.l + z_0 . \sinh\sqrt{zy}.l$$

$$\therefore \boxed{z_T = z_R . \cosh.\sqrt{zy}.l + z_0 . \sinh\sqrt{zy}.l} \rightarrow (3)$$

Distortionless line :

A Transmission line which satisfies the condition $\frac{R}{L} = \frac{G}{C}$ is called as distortionless line.

$$\boxed{RC = LG} \rightarrow (1)$$

W.K.T $\gamma = P = \alpha + j\beta = \sqrt{zy}$

$$= \sqrt{(R + j\omega L)(G + j\omega C)}$$

$$= \sqrt{LC\left(\frac{R}{L} + j\omega\right)\left(\frac{G}{C} + j\omega\right)}$$

$$= \sqrt{LC\left(\frac{R}{L} + j\omega\right)\left(\frac{R}{L} + j\omega\right)} \qquad \because \frac{R}{L} = \frac{G}{C}$$

$$\gamma = \alpha + j\beta = \left(\frac{R}{L} + j\omega\right)\sqrt{LC} \quad (\text{or}) \quad \sqrt{LC}\left(\frac{G}{C} + j\omega\right)$$

$$\alpha + j\beta = \sqrt{LC} \cdot \frac{R}{L} + j\omega\sqrt{LC} \quad (\text{or}) \quad \sqrt{LC} \cdot \frac{G}{C} + j\omega\sqrt{LC}$$

$$\alpha + j\beta = R \cdot \sqrt{\frac{C}{L}} + j\omega\sqrt{LC} \quad (\text{or}) \quad G \cdot \sqrt{\frac{L}{C}} + j\omega\sqrt{LC}$$

$$\therefore \alpha = R \cdot \sqrt{\frac{C}{L}} \quad (\text{or}) \quad G \cdot \sqrt{\frac{L}{C}} \quad \rightarrow (2)$$

$$\beta = \omega\sqrt{LC} \quad \rightarrow (3)$$

From equ. (2) & (3) we come to know that 'α' is independent of frequency and 'β' is a constant multiplied by 'ω'.

Waveform Distortion (or) Distortion line (or) line Distortion

The signal transmitted through the Txn. line will be in complex form and has many freq. components. In ideal Txn., line, the signal received at the receive end must be same as the transmitted signal. This condition is achieved only if all the freq. component are attenuated equally and transmitted with same delay. There are 2 types of waveform distortion.

1. Frequency Distortion
2. Delay (or) Phase Distortion

1. <u>Frequency Distortion</u>:

It is a type of distortion in which all the Freq. Components are not attenuated at same level (equally. This distortion can be avoided if 'α' is independent of 'ω'. In Txn. line equalizers are used at the ends to reduce the distortion.

$$\alpha = \sqrt{\frac{(RG-\omega^2 LC)+\sqrt{(RG-\omega^2 LC)^2+\omega^2(RC+LG)^2}}{2}}$$

2. <u>Delay Distortion</u>:

It is a type of distortion in which all the Freq. components are transmitted at different time intervals. This distortion can be avoided if 'β' is a constant multiplied by ω and v_p is independent of 'ω'. coaxial cables are used to reduce this distortion

$$\beta = \sqrt{\frac{(\omega^2 LC-RG)+\sqrt{(RG-\omega^2 LC)^2+\omega^2(RC+LG)^2}}{2}}$$

(or)

$$\beta = \sqrt{\frac{(\omega^2 LC-RG)+\sqrt{(R^2+\omega^2 L^2)(G^2+\omega^2 C^2)}}{2}}$$

Proof :-

W.K.T $\gamma = P = \alpha + j\beta = \sqrt{ZY} = \sqrt{(R+j\omega L)(G+j\omega C)} \rightarrow (1)$

squaring on both sides

$$(\alpha+j\beta)^2 = (R+j\omega L)(G+j\omega C)$$

$$\alpha^2 - \beta^2 + 2j\alpha\beta = RG + j\omega RC + j\omega LG - \omega^2 LC$$

$$(\alpha^2-\beta^2) + j\,2\alpha\beta = (RG - \omega^2 LC) + j\omega(RC + LG)$$

Equating real and imaginary parts

$$\alpha^2 - \beta^2 = RG - \omega^2 LC \rightarrow (2)$$

$$2\alpha\beta = \omega(RC+LG) \rightarrow (3)$$

Find magnitude for equ. (1)

$$\alpha + j\beta = \sqrt{(RG - \omega^2 LC) + j\omega(RC+LG)}$$

$$\sqrt{\alpha^2+\beta^2} = \sqrt{\sqrt{(RG-\omega^2 LC)^2 + \omega^2(RC+LG)^2}}$$

squaring on both sides

$$\alpha^2 + \beta^2 = \sqrt{(RG-\omega^2 LC)^2 + \omega^2(RC+LG)^2} \rightarrow (4)$$

Solve equ. (2) & (4) to find 'α' and 'β'

$$\alpha^2 - \beta^2 = RG - \omega^2 LC$$

$$\alpha^2 + \beta^2 = \sqrt{(RG-\omega^2 LC)^2 + \omega^2(RC+LG)^2}$$

$$2\alpha^2 = RG - \omega^2 LC + \sqrt{(RG-\omega^2 LC)^2 + \omega^2(RC+LG)^2}$$

$$\therefore \alpha = \sqrt{\frac{(RG-\omega^2 LC) + \sqrt{(RG-\omega^2 LC)^2 + \omega^2(RC+LG)^2}}{2}} \rightarrow (5)$$

Similarly,

$$\beta = \sqrt{\frac{(\omega^2 LC - RG) + \sqrt{(RG-\omega^2 LC)^2 + \omega^2 (RC+LG)^2}}{2}} \rightarrow ⑥$$

From equ. ⑤ and ⑥ we come to know that, 'α' is depending on 'ω' and 'β' is not a constant multiplied by 'ω'. So, distortion takes place in line.

Reflection Co-efficient :-

It is defined as the ratio of reflected voltage or current to the incident voltage or current.

It is represented by K, ρ, γ

$$K = \frac{V_r}{V_i} = \frac{-I_r}{I_i} \rightarrow ①$$

From general solution of Txn. line, the voltage and current can be expressed as,

$$E = \frac{E_R}{2}\left(\frac{Z_R+Z_0}{Z_R}\right)\left[e^{\sqrt{ZY}.s} + \left(\frac{Z_R-Z_0}{Z_R+Z_0}\right)e^{-\sqrt{ZY}.s}\right] \rightarrow ②$$

$$I = \frac{I_R}{2}\left(\frac{Z_R+Z_0}{Z_0}\right)\left[e^{\sqrt{ZY}.s} - \left(\frac{Z_R-Z_0}{Z_R+Z_0}\right)e^{-\sqrt{ZY}.s}\right] \rightarrow ③$$

The above 2 expressions has 2 terms. The first term represented in terms of positive 's' is called as incident wave and the term represented in terms of

negative 's' is called as reflected wave.

From equ. ②,

$$E = \frac{E_R}{2}\left(\frac{Z_R+Z_0}{Z_R}\right) \cdot e^{\sqrt{ZY}\cdot s} + \frac{E_R}{2}\left(\frac{Z_R-Z_0}{Z_R}\right) e^{-\sqrt{ZY}\cdot s} \rightarrow ④$$

(first term ↓ E_i (or) V_i; second term ↓ E_r (or) V_r)

From equ. ④

$$E_i = \frac{E_R}{2}\left(\frac{Z_R+Z_0}{Z_R}\right) e^{\sqrt{ZY}\cdot s} \rightarrow ⑤$$

$$E_r = \frac{E_R}{2}\left(\frac{Z_R-Z_0}{Z_R}\right) e^{-\sqrt{ZY}\cdot s} \rightarrow ⑥$$

sub ⑤ & ⑥ in ①

$$K = \frac{\frac{E_R}{2}\left(\frac{Z_R-Z_0}{Z_R}\right) e^{-\sqrt{ZY}\cdot s}}{\frac{E_R}{2}\left(\frac{Z_R+Z_0}{Z_R}\right) e^{+\sqrt{ZY}\cdot s}} \Rightarrow \left(\frac{Z_R-Z_0}{Z_R+Z_0}\right) \cdot e^{-2\sqrt{ZY}\cdot s}$$

At load end, $s=0$

※ $$\therefore \boxed{K = \frac{Z_R-Z_0}{Z_R+Z_0}} \rightarrow ⑦$$

$0 < K < 1$

Similarly, from equ. ③

$$I = \frac{I_R}{2}\left(\frac{Z_R+Z_0}{Z_0}\right) \cdot e^{\sqrt{ZY}\cdot s} - \frac{I_R}{2}\left(\frac{Z_R-Z_0}{Z_0}\right) \cdot e^{-\sqrt{ZY}\cdot s} \rightarrow ⑧$$

(first term ↓ I_i; second term ↓ I_r)

From equ. ⑧

$$I_i = \frac{I_R}{2}\left(\frac{Z_R + Z_0}{Z_0}\right) \cdot e^{\sqrt{ZY}.s} \rightarrow ⑨$$

$$I_r = -\frac{I_R}{2}\left(\frac{Z_R - Z_0}{Z_0}\right) e^{-\sqrt{ZY}.s} \rightarrow ⑩$$

sub ⑨ & ⑩ in ①

$$K = \frac{-\left(-\frac{I_R}{2}\left(\frac{Z_R - Z_0}{Z_0}\right) e^{-\sqrt{ZY}.s}\right)}{\frac{I_R}{2}\left(\frac{Z_R + Z_0}{Z_0}\right) e^{\sqrt{ZY}.s}}$$

$$\therefore K = \frac{Z_R - Z_0}{Z_R + Z_0} \cdot e^{-2\sqrt{ZY}.s}$$

At Load end, $s = 0$.

$$\therefore \boxed{K = \frac{Z_R - Z_0}{Z_R + Z_0}} \rightarrow ⑪$$

$< K < 1$ ✻

Reflection On a Line not Terminated in Z_0 :-

The phenomenon of setting up of reflected wave in a transmission Line is called as reflection.

Reflection is maximum in open circuit ($Z_R = \infty$) or short circuit line ($Z_R = 0$).

Reflection is zero when $Z_R = Z_0$.

From general solution of txn. line, the voltage and current can be expressed as,

$$E = \frac{E_R}{2} \cdot \left(\frac{Z_R + Z_0}{Z_R}\right) \cdot \left[e^{\sqrt{zy} \cdot s} + \left(\frac{Z_R - Z_0}{Z_R + Z_0}\right) e^{-\sqrt{zy} \cdot s} \right] \rightarrow (1)$$

$$I = \frac{I_R}{2} \cdot \left(\frac{Z_R + Z_0}{Z_0}\right) \cdot \left[e^{\sqrt{zy} \cdot s} - \left(\frac{Z_R - Z_0}{Z_R + Z_0}\right) e^{-\sqrt{zy} \cdot s} \right] \rightarrow (2)$$

The above two expressions has 2 terms. The first term represented in terms of '+s' is called as incident wave which flows from the sending end to the receiving end.

The second term represented in terms of '−s' is called as reflected wave which flows from the receiving end to the sending end.

From equ (1),

$$E_i = \frac{E_R}{2} \left(\frac{Z_R + Z_0}{Z_R}\right) e^{\sqrt{zy} \cdot s} \rightarrow (3);$$ Incident voltage wave

$$E_r = \frac{E_R}{2} \left(\frac{Z_R - Z_0}{Z_R}\right) e^{-\sqrt{zy} \cdot s} \rightarrow (4);$$ Reflected Voltage wave

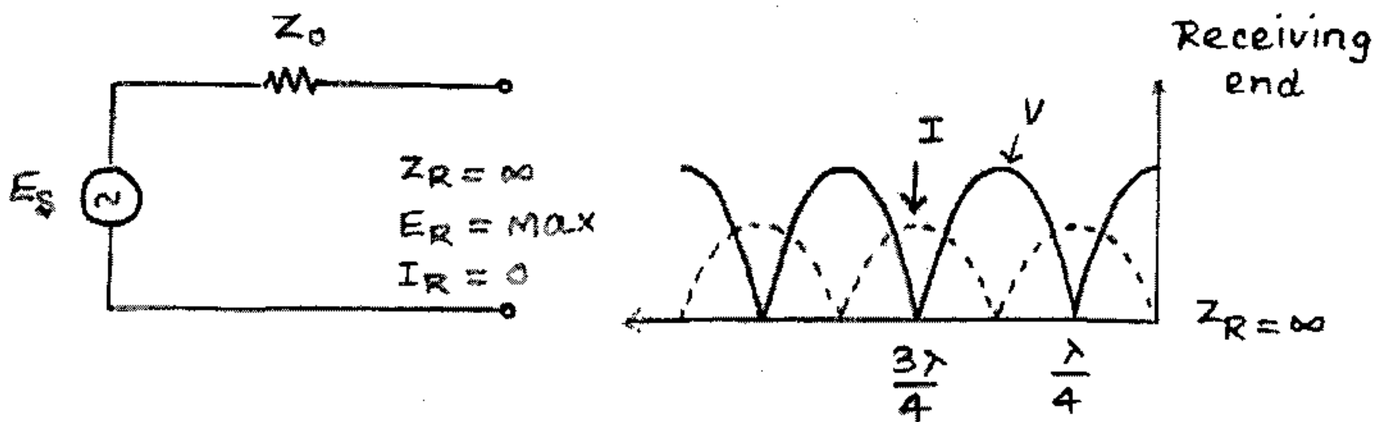
Similarly from equ (2)

$$I_i = \frac{I_R}{2} \left(\frac{Z_R + Z_0}{Z_0}\right) \cdot e^{\sqrt{zy} \cdot s} \rightarrow (5);$$ Incident Current wave

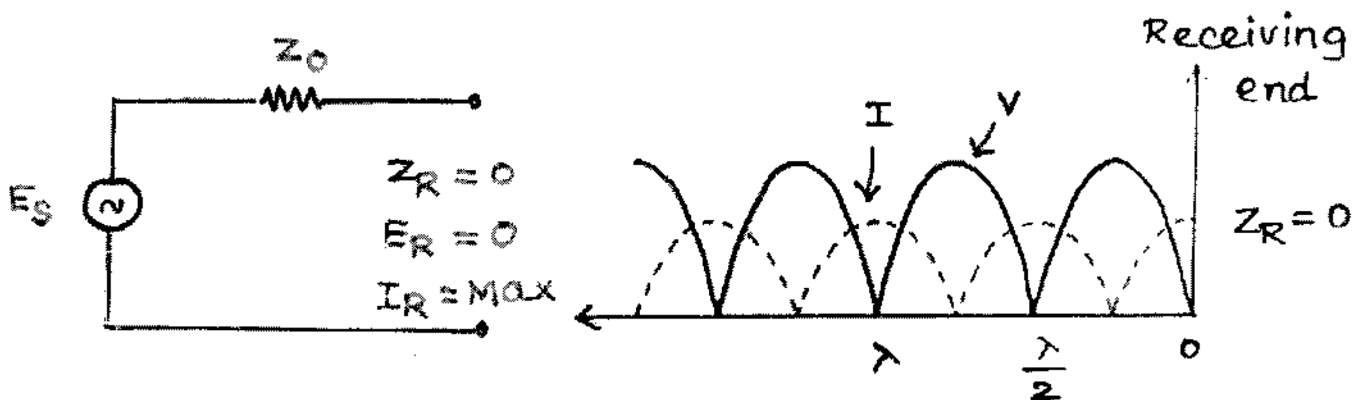
$$I_r = -\frac{I_R}{2} \left(\frac{Z_R - Z_0}{Z_0}\right) \cdot e^{-\sqrt{zy} \cdot s} \rightarrow (6);$$ Reflected Current wave

Thus the total instantaneous Voltage or current at any point on the line is the phasor sum of voltage or current of the incident and reflected waves.

In Open circuit, the magnetic field gets collapsed at the load end and increases electric field. Due to this voltage will be maximum at load end.



In Short circuit, electric field gets collapsed at load end and increases magnetic field. Due to this, current will be maximum at load end.



Loading of Lines :

The process of increasing the inductance 'L' of a line artificially is called as Loading of a Line.

Loading is introduced in telephone cables.

There are 3 types of Loading

i) Continuous Loading

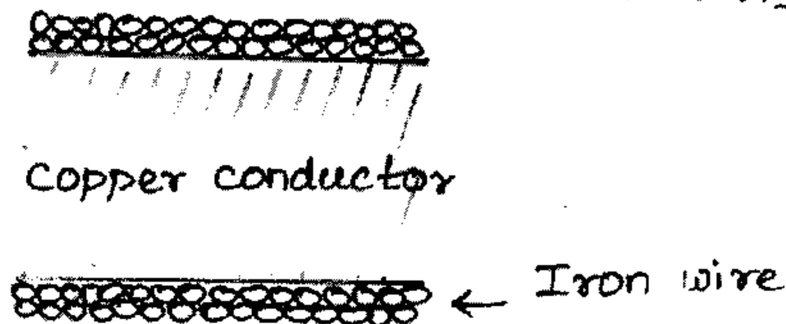
ii) Lumped Loading

iii) Patch Loading.

i) Continuous Loading :

In this method, the inductance of the line is increased uniformly along the length of the line.

In this type, iron or high permeability magnetic material in the form of a wire or tape is wound around the copper conductor as shown in figure.



Advantages :

1. Attenuation is constant over a wide range of freq.
2. continuous Loading is used ~~only~~ on submarine cables.

Advantages:

1. Due to high toroidal cores, Large values of inductanc is possible.
2. Eddy current and hysteresis losses are less.
3. cost is Less.

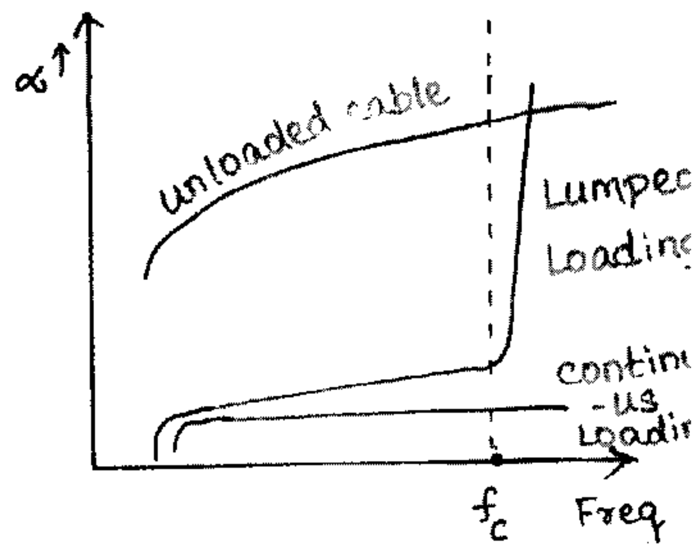
Disadvantages :-

1. Lumped Loading is useful only for voice band circuits upto 3 KHZ.
2. Inductance value is not uniform throughout the line.

iii) Patch Loading:

This type of Loading employs sections of continuously Loaded cable separated by sections of unloaded cable.

The typical length for the Patch Loading is normally 0.25 km.



Disadvantages :

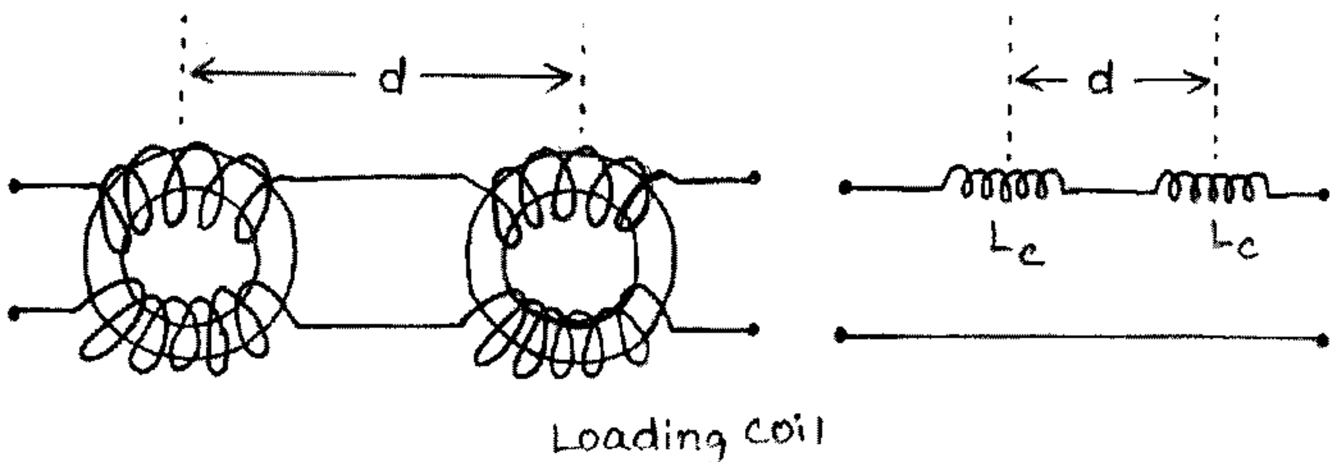
1. very expensive due to high cost of manufacture.
2. Only low inductance value is possible.
3. Since loading is done with iron wire, eddy current and hysteresis losses increases with frequency.

ii) Lumped Loading :-

In this type of loading, the inductors are introduced in lumps at uniform distances, in the line.

The inductors are introduced in both the limbs to keep the line as balanced circuit.

The lumped loading is preferred for open wire lines.



Loading coil

Telephone cable:

In the ordinary telephone cable, the wires are insulated with paper and twisted in pairs. This result in negligible values of inductance & conductance and can be neglected.

We know that,

$$\gamma = \sqrt{(R+j\omega L)(G+j\omega C)} \qquad (L \simeq G \simeq 0)$$

$$= \sqrt{R(j\omega C)}$$

$$= \sqrt{\omega RC}\sqrt{j} = \sqrt{\omega RC}\sqrt{\cos 90^\circ + j\sin 90^\circ}$$

$$= \sqrt{\omega RC}\,(\cos 45^\circ + j\sin 45^\circ)$$

$$= \sqrt{\omega RC}\left(\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}\right)$$

$$\alpha + j\beta = \sqrt{\frac{\omega RC}{2}} + j\sqrt{\frac{\omega RC}{2}}$$

Equating real & imaginary terms, we get

$$\alpha = \beta = \sqrt{\frac{\omega RC}{2}}$$

$$v_p = \frac{\omega}{\beta} = \sqrt{\frac{2\omega}{RC}}$$

From the expressions of α, β & V_p, it is obvious that both frequency and phase distortion occurs in ordinary telephone

cable. Therefore, to achive distortionless telephone cable (ie. to achieve $RC = LG$ condition), the value of Inductance has to be increased.

The process of increasing the value of inductance to achieve distortionless line is called Loading.

<u>Loaded Telephone cable:</u>

To understand the significance of Loading, consider a loaded telephone cable. (Inductance cannot be neglected). The primary constants of a loaded cable are R, L and C (only $G = 0$).

$$\therefore \quad Z = R + j\omega L$$

$$Y = j\omega C \qquad (\because G = 0)$$

$$\gamma = \sqrt{ZY} = \sqrt{|Z| \angle Z \; |Y| \angle Y} \qquad \text{—— (1)}$$

$$= \sqrt{\sqrt{R^2 + \omega^2 L^2} \angle \tan^{-1}\left(\frac{\omega L}{R}\right) \; \omega C \angle \pi/2}$$

$$= \sqrt{\sqrt{\omega^2 L^2 \left(1 + \frac{R^2}{\omega^2 L^2}\right)} \angle \frac{\pi}{2} - \tan^{-1}\left(\frac{R}{\omega L}\right) \cdot \omega C \angle \pi/2}$$

$$= \sqrt{\omega L \sqrt{1 + \frac{R^2}{\omega^2 L^2}} \cdot \omega C \angle \pi - \tan^{-1}\left(\frac{R}{\omega L}\right)}$$

$$\therefore \gamma = \omega\sqrt{LC}\sqrt{\sqrt{1+\frac{R^2}{\omega^2L^2}}}\ \angle\frac{\pi}{2}-\frac{1}{2}\tan^{-1}\left(\frac{R}{\omega L}\right)$$

Since $R << \omega L$, $\frac{R^2}{\omega^2L^2} \simeq 0$

$$\therefore \gamma = \omega\sqrt{LC}\ \angle\frac{\pi}{2}-\frac{1}{2}\tan^{-1}\left(\frac{R}{\omega L}\right)$$

Let $\frac{\pi}{2}-\frac{1}{2}\tan^{-1}\left(\frac{R}{\omega L}\right) = \theta$

$$\gamma = \omega\sqrt{LC}\ \angle\theta = \omega\sqrt{LC}\ e^{j\theta}$$

$$\gamma = \omega\sqrt{LC}\ (\cos\theta + j\sin\theta) \qquad \text{——— (2)}$$

<u>To find $\cos\theta$:</u>

$$\cos\left(\frac{\pi}{2}-\frac{1}{2}\tan^{-1}\left(\frac{R}{\omega L}\right)\right)$$

$$= \cos\left(\frac{\pi}{2}-\frac{1}{2}\left(\frac{R}{\omega L}\right)\right)$$

$$= \sin\left(\frac{1}{2}\frac{R}{\omega L}\right) = \frac{R}{2\omega L}$$

[$\because \frac{R}{\omega L}$ is very les

tan (smaller angle) = angl itsel

ie. If ϕ is very small

$\cos\phi \simeq 1$

$\sin\phi \simeq \phi$

$\tan\phi \simeq \phi$

<u>To find $\sin\theta$</u>

$$\sin\left(\frac{\pi}{2}-\frac{1}{2}\tan^{-1}\left(\frac{R}{\omega L}\right)\right)$$

$$= \sin\left(\frac{\pi}{2}-\frac{1}{2}\cdot\frac{R}{\omega L}\right)$$

$$= \cos\left(\frac{1}{2}\cdot\frac{R}{\omega L}\right) \simeq 1$$

$\therefore$ Equation (2) can be written as

$$\gamma = \omega\sqrt{LC}\left(\frac{R}{2\omega L} + j\cdot 1\right)$$

$$\alpha + j\beta = \frac{R}{2}\sqrt{\frac{C}{L}} + j\omega\sqrt{LC}$$

By equating real & imaginary terms, we get

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}} \quad , \quad \beta = \omega\sqrt{LC} \quad \& \quad v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}}$$

From the above expressions, it is obvious that by increasing the inductance value (ie. by means of Loading) distortionless line can be achieved.

Losses in a Transmission line:

The three types of losses that generally occur in a transmission line are

(1) Reflection loss & Reflection factor

(2) Return loss

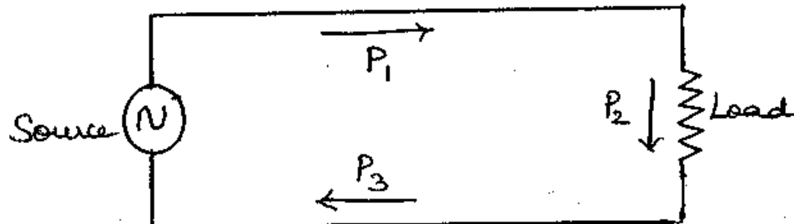
(3) Insertion loss

(1) **Reflection loss & Reflection factor:**

Reflection loss is defined as the ratio of power delivered to the load when $Z_R = Z_0$ (impedance matching condition) to power delivered to the load when $Z_R \neq Z_0$ (impedance mismatch condition).

$$\text{Reflection loss} = \frac{P_L \ (\text{when } Z_R = Z_0)}{P_L \ (\text{when } Z_R \neq Z_0)}$$

consider the following transmission line.



P_1 — Incident power

P_2 — Actual power delivered to the load

P_3 — Reflected power.

We know that $\frac{P_3}{P_1} = K^2$ —— (1)

$P_2 = P_1 - P_3$ —— (2)

$$\text{Reflection loss} = \frac{P_1}{P_2}$$

$P_1 \rightarrow$ Power that must be delivered to the load

$P_2 \rightarrow$ Actual power delivered to the load

$$= \frac{P_1}{P_1 - P_3}$$

$$= \frac{P_1}{P_1 - P_1 K^2}$$

$$= \frac{1}{1 - K^2}$$

$$= \frac{1}{1 - \left(\frac{Z_R - Z_0}{Z_R + Z_0}\right)^2}$$

$$= \frac{(Z_R + Z_0)^2}{(Z_R + Z_0)^2 - (Z_R - Z_0)^2}$$

$$\text{Ref. loss} = \frac{(Z_R + Z_0)^2}{4 Z_R Z_0}$$

$$\text{Reflection loss in dB} = 10 \log \left(\frac{(Z_R + Z_0)^2}{4 Z_R Z_0} \right)$$

$$= 10 \log \left(\frac{Z_R + Z_0}{2\sqrt{Z_R Z_0}} \right)^2$$

$$\therefore \text{Reflection loss in dB} = 20 \log \left(\frac{Z_R + Z_0}{2\sqrt{Z_R Z_0}} \right)$$

$$= 20 \log \left(\frac{1}{k_f} \right)$$

Where k_f is called Reflection factor which is given as,

$$k_f = \frac{2\sqrt{Z_R Z_0}}{(Z_R + Z_0)}$$

In general, reflection factor can be defined as the ratio of geometric mean to

arithmetic mean of any two impedances.

(2) <u>Return loss</u>:

It is defined as the ratio of incident power to reflected power.

$$\text{Return loss} = \frac{P_i}{P_r}$$

$$\text{Return loss} = \frac{1}{|K|^2} \qquad \left[\because |K|^2 = \frac{P_r}{P_i}\right.$$

$K \rightarrow$ Reflection coefficient]

$$\text{Return loss in dB} = 10 \log \frac{1}{|K|^2}$$

$$RL_{dB} = -20 \log |K|$$

(3) <u>Insertion loss</u>:

It is defined as the ratio of amount of power delivered to the load before insertion of a line to power delivered to the load after insertion of a line.

$$\text{Insertion loss} = \frac{P_L \text{ before insertion of a line}}{P_L \text{ after insertion of a line}}$$

$$\text{Insertion loss} = \frac{k_{SR}}{k_S \, k_R} e^{\alpha l}$$

Where,

k_{SR} → Reflection factor between source & receiver

$$k_{SR} = \frac{2\sqrt{Z_S Z_R}}{Z_S + Z_R}$$

k_S → Reflection factor at the sending end of a line

$$k_S = \frac{2\sqrt{Z_S Z_0}}{Z_S + Z_0}$$

k_R → Reflection factor at the receiving end of a line

$$k_R = \frac{2\sqrt{Z_R Z_0}}{Z_R + Z_0}$$

Infinite line:

A transmission line which is infinitely long is called an infinite line (i.e. $l \to \infty$). The input impedance of an infinite line is calculated as follows;

We know that, $$Z_{IN} = \frac{Z_0 (Z_R + Z_0 \tanh \gamma l)}{Z_0 + Z_R \tanh \gamma l}$$

As $l \to \infty$ in an infinite line, sub. $l = \infty$ in the above expression.

$$\therefore Z_{IN} = \frac{Z_0 (Z_R + Z_0 \tanh \gamma \infty)}{Z_0 + Z_R \tanh \gamma \infty}$$

As $\tanh \infty = 1$,

$$Z_{IN} = \frac{Z_0 (Z_R + Z_0)}{Z_0 + Z_R}$$

$$\therefore \boxed{Z_{IN} = Z_0}$$

As the infinite line is hypothetical, a finite line equivalent to infinite line has to be derived. i.e. a finite line with input impedance equal to Z_0 has to be derived. This can be achieved by considering a transmission line which is

properly terminated. ie. $Z_R = Z_0$.

If $Z_R = Z_0$ in a finite line, then its $Z_{IN} = Z_0$.

Proof:

$$Z_{IN} = \frac{Z_0 (Z_R + Z_0 \tanh \gamma l)}{Z_0 + Z_R \tanh \gamma l}$$

Sub. $Z_R = Z_0$ in the above expression,

$$Z_{IN} = \frac{Z_0 (Z_0 + Z_0 \tanh \gamma l)}{Z_0 + Z_0 \tanh \gamma l}$$

$$\therefore Z_{IN} = Z_0$$

Thus,

a properly terminated finite line (ie. $Z_R = Z_0$)	$\equiv$	Infinite line

Wavelength & velocity of propagation:

Velocity or Phase Velocity: It is defined as the velocity with which a signal of single frequency propagates along a transmission line.

ie. $$v_p = \frac{\omega}{\beta}$$

Wavelength: The distance travelled by a wave along a transmission medium in which the phase of the wave changes by 2π radians is called wavelength.

$$\therefore \lambda = \frac{2\pi}{\beta}$$

EC 8651 - Transmission lines & RF systems

Transmission lines:

It is defined as a physical conducting medium which transmits information/power in the form of electrical signal from one end to another.

Examples: Coaxial cable, Twisted pair cable, Open wire line etc. The type of wave propagation in these transmission lines is called guided wave propagation.

Classification of transmission lines:

Transmission lines are classified as

1. Metallic lines
2. Non metallic lines (optical fibres)
3. Strip lines
4. Waveguides

Metallic lines:

The various types of metallic lines are

(1) Single wire line (2) Two wire lines

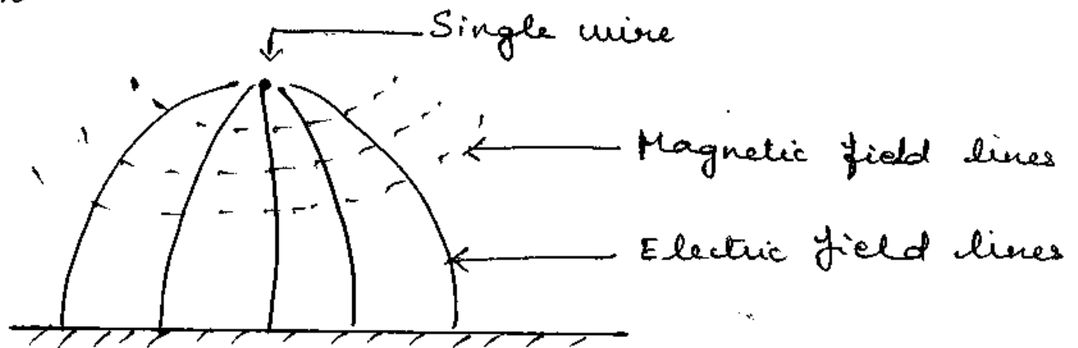
(3) Coaxial line

Single wire line:

In this type of line, a single solid conducting wire is used to connect two ends.

*. This line is acceptable at low frequencies

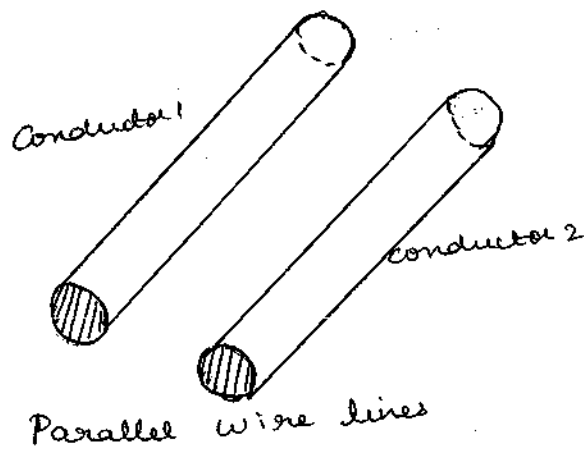
*. At high frequencies, more energy is dissipated by radiation.



Two wire lines or Parallel wire lines or open wire line:

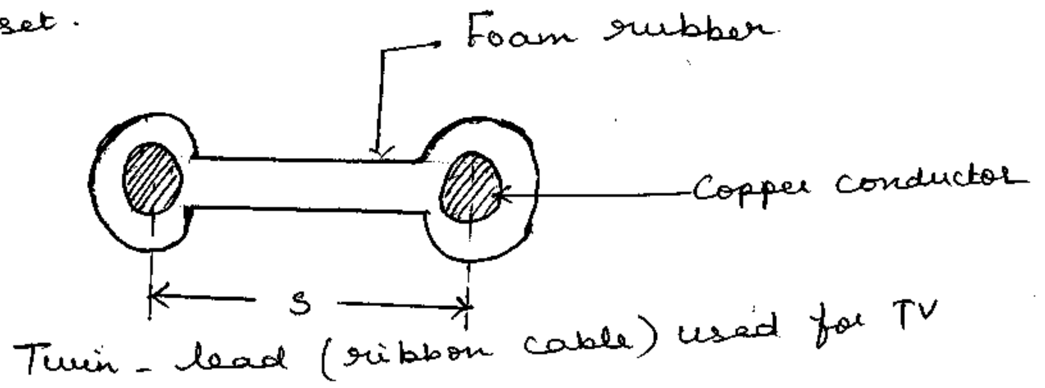
In this line, two wires or conductors are placed parallely.

*. At low frequencies (as in Voice telephony or telegraphy), the wires are supported by post or buried inside the earth.



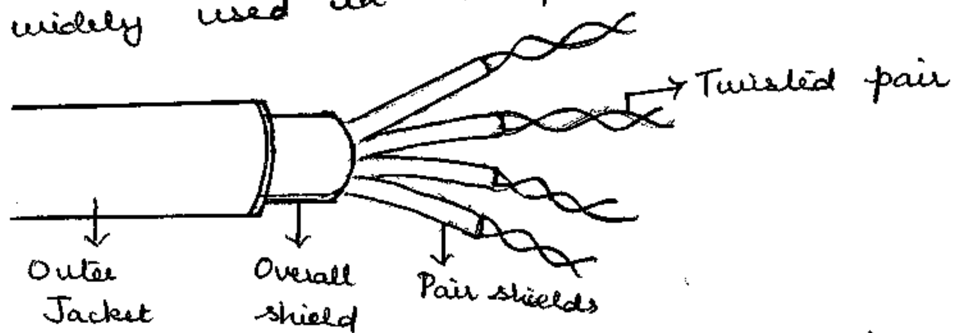
Parallel wire lines

*. One type of parallel wire line is twin lead arrangement which is used to connect an antenna to a Television set.



Twin - lead (ribbon cable) used for TV

*. Another type is twisted pair cable. This is formed by twisting 2 insulated conductors. The conductors are twisted to reduce noise interference between pairs & thus eliminate cross talk. This cable is widely used in Telephone applications.



*. Another type of parallel wire line is a Shielded pair transmission line. In this parallel wire line is placed inside a conducting pipe or metallic braid as an electromagnetic shield.

Advantages:

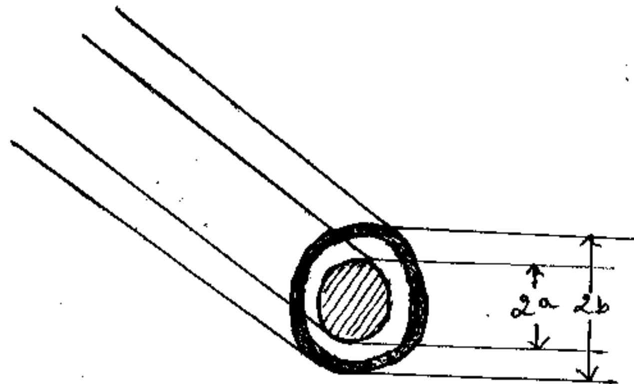
1. Cheaper
2. Easily installed
3. No dielectric between the 2 wires & hence no dielectric loss.

Disadvantages:

More radiation loss

Coaxial line:

This line consists of 2 conductors (inner conductor and outer conductor) placed coaxially.



a → Radius of the inner conductor
b → Radius of the outer conductor

The Source is connected to the inner conductor of the coaxial line. The other end of the inner conductor is connected to load Z_L. The other end of source, load & outer conductor of the coaxial line are all connected to the ground. Hence, the voltage between the inner conductor & the ground

& the outer conductor & the ground are different.
Therefore, the coaxial line is an unbalanced line.

As the two conductors are at two different potentials, the fields are entirely confined to the space between the two conductors. No field exist outside the outer conductor & similarly no external radiation can penetrate the outer conductor & propagate inside.

<u>Types of coaxial cables:</u>

1. Flexible
2. Semi-rigid
3. Rigid

The flexible coaxial cable use copper braided outer conductor, a thin center conductor & a low loss solid or foam polyethylene dielectric. Semi-rigid cables have solid dielectric and thin outer conductor so that it could be bent while laying cables. The rigid cables have solid dielectric made of Teflon

<u>Advantages</u>:

1. Low dielectric & radiation loss
2. Cheaper
3. Easy to instal & to maintain

Disadvantages:

1. Can be used upto 3 GHz. Beyond this frequency, more losses occur in solid dielectric and conductor.

Strip lines:

They are transmission lines used as microwave circuits / components in conjunction with microwave semiconductor devices. They are used over the frequency range from 100 MHz to 30GHz. The commonly used dielectrics in the structure of strip line are teflon, polystyrene etc. The mode of propagation is TEM mode. Strip lines are unbalanced lines since the ground planes are kept at ground potential.

Advantages:

1. Used in microwave applications
2. Simple construction
3. Can easily be integrated with MIC components

Disadvantages:

1. Strip lines have much less power handling capability.
2. Have greater losses.

Waveguides:

It is a hollow single conductor metallic structure used to propagate high frequency signals (in the range of GHz). Waveguide dimension is inversely proportional to frequency. TE and TM are the modes of signal propagation. Depending on the cross section, it can be classified as Rectangular, Cylindrical or Elliptical waveguide.

Advantages:

1. Higher power handling capability
2. Reduces fabrication cost
3. Lower attenuation per unit length.
4. No hysteresis or eddy current loss

Disadvantages:

1. High cost due to thicker metallic structure
2. **Difficult to instal & special couplings are required.**

Comparison of Various transmission lines

Type	Frequency range (upper / Limit)	Losses	Usable Bandwidth	Power handling capacity	Physical size	Modes of Operation
Open wire lines	Low to VHF (500 MHz)	Very high	Lowest	Very high	Small	TEM
Coaxial cable	Low to Microwave (18 GHz)	Medium	High	Medium	Medium	TEM
Waveguide	Microwave (300 GHz)	Low	High	Very high	Large	TE, TM
Strip line & Microstrip line	Microwave (30 GHz)	High	High	Low	Small	Quasi TEM
Optical fibres	Infra-red (0.8 μm - 2.5 μm)	Very low	Very high	Very low	Very Small	TE, TM & Hybrid modes

2. High Frequency Transmission Lines

The standard assumptions made for the analysis of Radio Freq. lines are,

1. At very high Freq, the skin effect is considerable. Hence it is assumed that the currents may flow on the surface of conductor. Then the internal inductance becomes zero.

2. Due to skin effect, resistance R increases with $\sqrt{f}$. But the line reactance ωL increases directly with freq. 'f'. Hence the second assumption is $\omega L >> R$. ($R \simeq 0$)

3. The third assumption is that the leakage conductance 'G' is considered as zero. (i.e., $G \simeq 0$) $\omega C >> G$

Line Constants for Zero Dissipation Line :

In general, the characteristic impedance 'Z_0' and propagation constant 'γ' of a Txn. line is given by,

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \rightarrow (1)$$

$$\gamma = \sqrt{ZY} = \sqrt{(R + j\omega L)(G + j\omega C)} \rightarrow (2)$$

According to the standard assumptions for line at high Freq. $j\omega L >> R$ and $j\omega C >> G$. (i.e., $R = G \simeq 0$)

$$\therefore Z_0 = \sqrt{\frac{j\omega L}{j\omega C}} = \sqrt{\frac{L}{C}} \quad \rightarrow (3)$$

Since, Z_0 is real and resistive, it can be represented by symbol R_0.

$$\therefore Z_0 = R_0 = \sqrt{\frac{L}{C}} \quad \rightarrow (4)$$

Similarly,

$$\gamma = \sqrt{(j\omega L)(j\omega C)}$$

$$= \sqrt{j^2\omega^2 LC}$$

$$\gamma = j\omega\sqrt{LC} \quad \rightarrow (5)$$

W.K.T $\gamma = \alpha + j\beta$

$$\therefore \gamma = \alpha + j\beta = j\omega\sqrt{LC}$$

$$\alpha = 0 \quad \rightarrow (6)$$

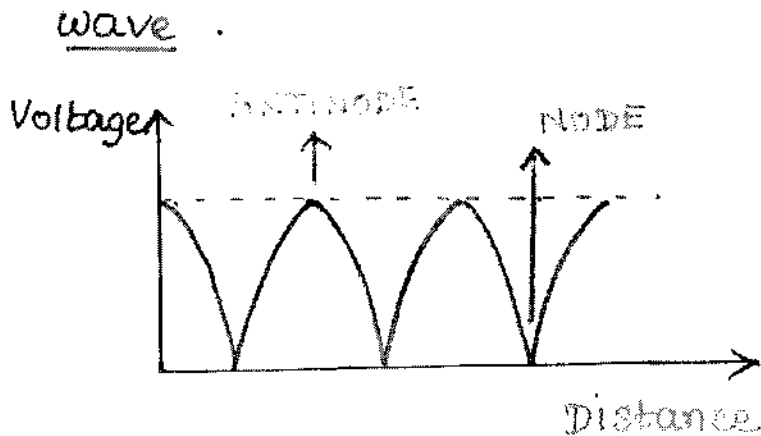
$$\beta = \omega\sqrt{LC} \quad \rightarrow (7)$$

$$\therefore v_p = \frac{\omega}{\beta} = \frac{\omega}{\omega\sqrt{LC}} = \frac{1}{\sqrt{LC}} \text{ m/sec} \quad \rightarrow (8)$$

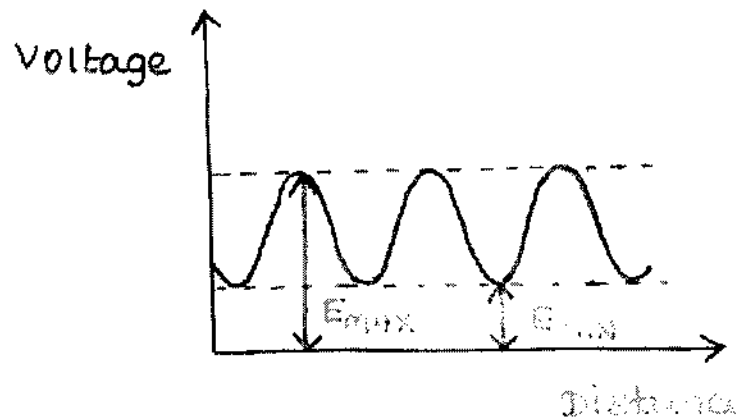
$$\lambda = \frac{2\pi}{\beta} = \frac{2\pi}{\omega\sqrt{LC}} \text{ m} \quad \rightarrow (9)$$

Standing wave :-

when $Z_R \neq Z_0$, some part of the transmitted signal from source to load will be reflected back towards the source. This reflected wave will combin with the incident wave which gives rise to <u>standing wave</u>.



* standing wave on a dissipationless line with OC or SC termination



* standing wave on a dissipationless line terminated with $Z_R \neq Z_0$

The points along the line where the magnitude of voltage or current is zero are called as <u>Nodes</u>.

The points along the line where the magnitude of voltage or current is maximum are called as <u>Antinodes or Loops</u>.

when a line is terminated in R_0, the standing waves are absent, such a line is called <u>smooth line</u>

Distance B/w <u>2 maximum</u> points is $\lambda/2$

Distance B/w <u>2 minimum</u> points is $\lambda/2$

Distance B/w <u>one</u> maximum & <u>one</u> minimum is $\lambda/4$

Standing Wave Ratio :

* The ratio of the maximum to minimum magnitudes of voltages or currents on a line having standing waves is called as standing wave ratio.

* It is denoted by 'S' or 'SWR'

* Standing wave Ratio is given by,

$$S = SWR = \frac{|V_{max}|}{|V_{min}|} = \frac{|E_{max}|}{|E_{min}|} = \frac{|I_{max}|}{|I_{min}|} \rightarrow (1)$$

* There are two types of SWR

i) VSWR — Voltage SWR

ii) CSWR — Current SWR

Relation between 'SWR' and 'K' :-

$$S = \frac{E_{max}}{E_{min}} = \frac{|E_i| + |E_r|}{|E_i| - |E_r|} \Rightarrow \frac{|E_i|}{|E_i|} \cdot \frac{1 + |E_r/E_i|}{1 - |E_r/E_i|}$$

$$\therefore \boxed{S = \frac{1 + |K|}{1 - |K|}} \rightarrow (2)$$

$$S(1 - |K|) = 1 + |K|$$

$$S - S \cdot |K| = 1 + |K|$$

$$S - 1 = S \cdot |K| + |K|$$

$$S - 1 = |K|(S + 1)$$

$$\therefore \boxed{|K| = \frac{S-1}{S+1}} \rightarrow (3)$$

Input Impedance of Zero dissipation Line :

The i/p impedance of a line can be found using the expression

$$Z_S = \frac{E_S}{I_S} \rightarrow (1)$$

From general solution of Txn. line, the current and voltage at any point on a line is expressed as,

$$E = E_R \cosh\sqrt{zy}.s + I_R \cdot Z_0 \cdot \sinh\sqrt{zy}.s \rightarrow (2)$$

$$I = I_R \cdot \cosh\sqrt{zy}.s + \frac{E_R}{Z_0} \sinh\sqrt{zy}.s \rightarrow (3)$$

For zero dissipation line,

$Z_0 = R_0$, $\gamma = j\beta$, $\alpha = 0$. So, the sending end voltage and current at a distance 's' is expressed as,

$$E_S = E_R \cdot \cosh(j\beta)s + I_R R_0 \sinh(j\beta)s$$

$$E_S = E_R \cos\beta s + j I_R R_0 \sin\beta s \rightarrow (4)$$

Similarly,

$$I_S = I_R \cos\beta s + j \frac{E_R}{R_0} \sin\beta s \rightarrow (5)$$

Substitute (4) & (5) in (1)

$$Z_S = \frac{E_R \cos\beta s + j I_R R_0 \sin\beta s}{I_R \cos\beta s + j \frac{E_R}{R_0} \sin\beta s}$$

$$Z_S = \frac{I_R \cdot Z_R \cdot \cos\beta s + j I_R R_0 \sin\beta s)}{I_R \cdot \cos\beta s + j \frac{I_R Z_R}{R_0} \sin\beta s)}$$

$$Z_S = \frac{\cancel{I_R}}{\cancel{I_R}} \frac{(Z_R \cos\beta s + j R_0 \sin\beta s)}{\left(\cos\beta s + j \frac{Z_R}{R_0} \sin\beta s \right)} \rightarrow (6)$$

W.K.T $\cos\beta s = \dfrac{e^{j\beta s} + e^{-j\beta s}}{2} \rightarrow (7)$; $\sin\beta s = \dfrac{e^{j\beta s} - e^{-j\beta s}}{2j} \rightarrow (8)$

sub (7) & (8) in (6)

$$Z_S = R_0 \left[\frac{Z_R \cdot \left(\frac{e^{j\beta s} + e^{-j\beta s}}{2} \right) + \cancel{j} R_0 \left(\frac{e^{j\beta s} - e^{-j\beta s}}{2\cancel{j}} \right)}{R_0 \left(\frac{e^{j\beta s} + e^{-j\beta s}}{2} \right) + \cancel{j} Z_R \left(\frac{e^{j\beta s} - e^{-j\beta s}}{2\cancel{j}} \right)} \right]$$

$$Z_S = \frac{R_0 \times \cancel{\frac{1}{2}}}{\cancel{\frac{1}{2}}} \left[\frac{Z_R e^{j\beta s} + Z_R e^{-j\beta s} + R_0 e^{j\beta s} - R_0 e^{-j\beta s}}{R_0 e^{j\beta s} + R_0 e^{-j\beta s} + Z_R e^{j\beta s} - Z_R e^{-j\beta s}} \right]$$

$$Z_S = R_0 \cdot \left[\frac{(Z_R + R_0)\, e^{j\beta s} + (Z_R - R_0)\, e^{-j\beta s}}{(Z_R + R_0)\, e^{j\beta s} - (Z_R - R_0)\, e^{-j\beta s}} \right]$$

$$Z_S = \frac{R_0 \cdot (Z_R + R_0) e^{j\beta s}}{(Z_R + R_0) e^{j\beta s}} \cdot \left[\frac{1 + \frac{(Z_R - R_0)\, e^{-j\beta s}}{(Z_R + R_0)\, e^{j\beta s}}}{1 - \frac{(Z_R - R_0) \cdot e^{-j\beta s}}{(Z_R + R_0) \cdot e^{j\beta s}}} \right]$$

W.K.T $\frac{Z_R - R_0}{Z_R + R_0} = K$

$$\therefore \quad Z_S = R_0 \cdot \left[\frac{1 + K \cdot e^{-j2\beta s}}{1 - K \cdot e^{-j2\beta s}} \right]$$

$$Z_S = R_0 \cdot \left[\frac{1 + |K| \cdot \angle\phi \; \angle{-2\beta s}}{1 - |K| \cdot \angle\phi \; \angle{-2\beta s}} \right] \qquad \because e^{-j\theta} = \angle{-\theta}$$

$$Z_S = R_0 \cdot \left[\frac{1 + |K| \angle{\phi - 2\beta s}}{1 - |K| \angle{\phi - 2\beta s}} \right] \rightarrow (9)$$

Input impedance will be maximum if the angle is zero

$$\phi - 2\beta s = 0$$

$$2\beta s = \phi$$

$$\boxed{s = \frac{\phi}{2\beta}} \rightarrow (10)$$

$$\therefore Z_S(max) = R_0 \cdot \frac{1 + |K|}{1 - |K|} \qquad \because \frac{1 + |K|}{1 - |K|} = S$$

$$\boxed{Z_S(max) = R_0 \cdot S} \rightarrow (11)$$

$Z_S(min)$ can be found when we move at a distance of $\frac{\lambda}{4}$ from $Z_S(max)$ point towards the source.

$$S = \frac{\phi}{2\beta} + \frac{\lambda}{4}$$

$$S = \frac{\phi}{2\beta} + \frac{2\pi}{4\beta} \qquad \because \lambda = \frac{2\pi}{\beta}$$

$$S = \frac{1}{2\beta}\left[\phi + \pi\right] \rightarrow (12)$$

$$\therefore Z_S(min) = R_o \cdot \left[\frac{1 + |K| \cdot \angle \phi - 2\beta \cdot \frac{1}{2\beta}(\phi + \pi)}{1 - |K| \cdot \angle \phi - 2\beta \cdot \frac{1}{2\beta}(\phi + \pi)}\right]$$

$$Z_S(min) = R_o \cdot \left[\frac{1 + |K| \cdot \angle \phi - \phi - \pi}{1 - |K| \angle \phi - \phi - \pi}\right]$$

$$Z_S(min) = R_o \cdot \left[\frac{1 + |K| \angle -\pi}{1 - |K| \angle -\pi}\right] \qquad \because \angle -\pi = -1$$

$$= R_o \cdot \left[\frac{1 - |K|}{1 + |K|}\right]$$

$$Z_S(min) = R_o \cdot \frac{1}{S}$$

$$\therefore Z_S(min) = \frac{R_o}{S} \rightarrow (13)$$

Input impedance of short Circuited Dissipationless Line

The input impedance of dissipationless line is expressed as,

$$Z_s = R_0 \cdot \left[\frac{Z_R \cos\beta s + j R_0 \sin\beta s}{R_0 \cos\beta s + j Z_R \sin\beta s} \right] \rightarrow (1)$$

$$Z_s = R_0 \cdot \frac{\cos\beta s}{\cos\beta s} \left[\frac{Z_R + j R_0 \frac{\sin\beta s}{\cos\beta s}}{R_0 + j Z_R \frac{\sin\beta s}{\cos\beta s}} \right]$$

$$Z_s = R_0 \cdot \left[\frac{Z_R + j R_0 \tan\beta s}{R_0 + j Z_R \tan\beta s} \right] \rightarrow (2)$$

For short circuited line $Z_R = 0$

$$\therefore Z_s = Z_{sc} = R_0 \cdot \left[\frac{j R_0 \tan\beta s}{R_0} \right]$$

$$X_s = X_{sc} = R_0 \tan\beta s \longrightarrow (3)$$

$$\frac{X_s}{R_0} = \frac{X_{sc}}{R_0} = \tan\beta s \longrightarrow (4)$$

where, $\frac{X_s}{R_0}$ is Normalised i/p reactance.

$$x_s = \frac{X_s}{R_0} \quad ; \quad x_{sc} = \frac{X_{sc}}{R_0}$$

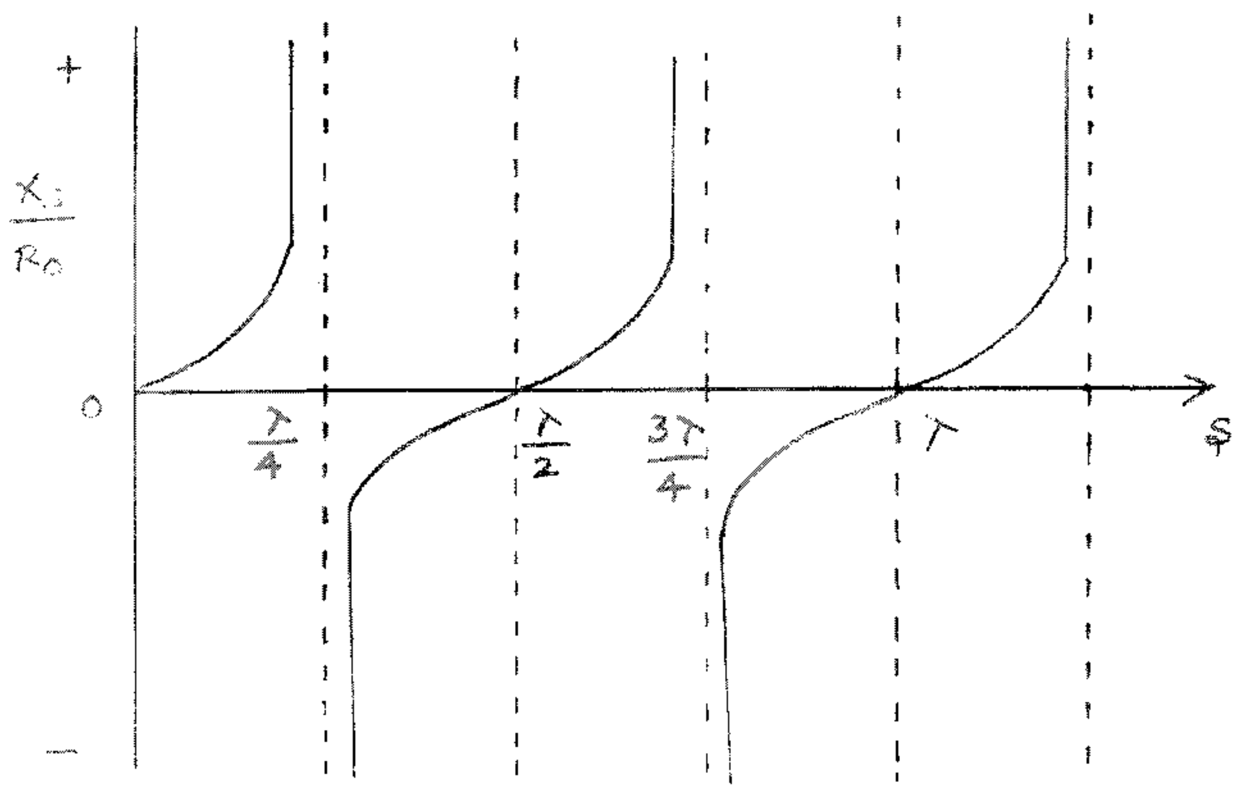
when, $S=0$, $\frac{X_s}{R_0} = \tan\left(\frac{2\pi}{\lambda}\cdot 0\right) = \tan(0) = 0$

$S=\frac{\lambda}{4}$, $\frac{X_s}{R_0} = \tan\left(\frac{2\pi}{\lambda}\cdot\frac{\lambda}{4}\right) = \tan\left(\frac{\pi}{2}\right) = \infty$

$S=\frac{\lambda}{2}$, $\frac{X_s}{R_0} = \tan\left(\frac{2\pi}{\lambda}\cdot\frac{\lambda}{2}\right) = \tan(\pi) = 0$

$S=\frac{3\lambda}{4}$, $\frac{X_s}{R_0} = \tan\left(\frac{2\pi}{\lambda}\cdot\frac{3\lambda}{4}\right) = \tan\left(\frac{3\pi}{2}\right) = -\infty$

$S=\lambda$, $\frac{X_s}{R_0} = \tan\left(\frac{2\pi}{\lambda}\cdot\lambda\right) = \tan(2\pi) = 0$



Input Impedance of Open Circuited Dissipationless line :-

The i/p impedance of dissipationless line is expressed as,

$$Z_S = R_0 \cdot \left[\frac{Z_R \cos\beta S + jR_0 \cdot \sin\beta S}{R_0 \cdot \cos\beta S + jZ_R \sin\beta S}\right] \longrightarrow (1)$$

$$Z_s = R_o \cdot \frac{\cos\beta s}{\cos\beta s}\left[\frac{Z_R + jR_o \frac{\sin\beta s}{\cos\beta s}}{R_o + jZ_R \frac{\sin\beta s}{\cos\beta s}}\right]$$

$$Z_S = R_o \left[\frac{Z_R + jR_o \tan\beta s}{R_o + jZ_R \tan\beta s}\right] \longrightarrow (2)$$

For Open circuited Line $Z_R = \infty$

$$Z_S = R_o \cdot \frac{Z_R}{Z_R}\left[\frac{1 + j\frac{R_o}{Z_R} \cdot \tan\beta s}{\frac{R_o}{Z_R} + j\tan\beta s}\right]$$

$$Z_S = R_o \cdot \left[\frac{1 + j\frac{R_o}{Z_R} \cdot \tan\beta s}{\frac{R_o}{Z_R} + j\tan\beta s}\right]$$

put $Z_R = \infty$ in (2)

$$Z_S = R_o \cdot \left[\frac{1+0}{0 + j\tan\beta s}\right]$$

$$Z_S = \frac{R_o}{j\tan\beta s}$$

$$Z_S = -jR_o \cot\beta s$$

$$\frac{Z_S}{R_o} = \frac{X_S}{R_o} = -\cot\beta s \longrightarrow (3)$$

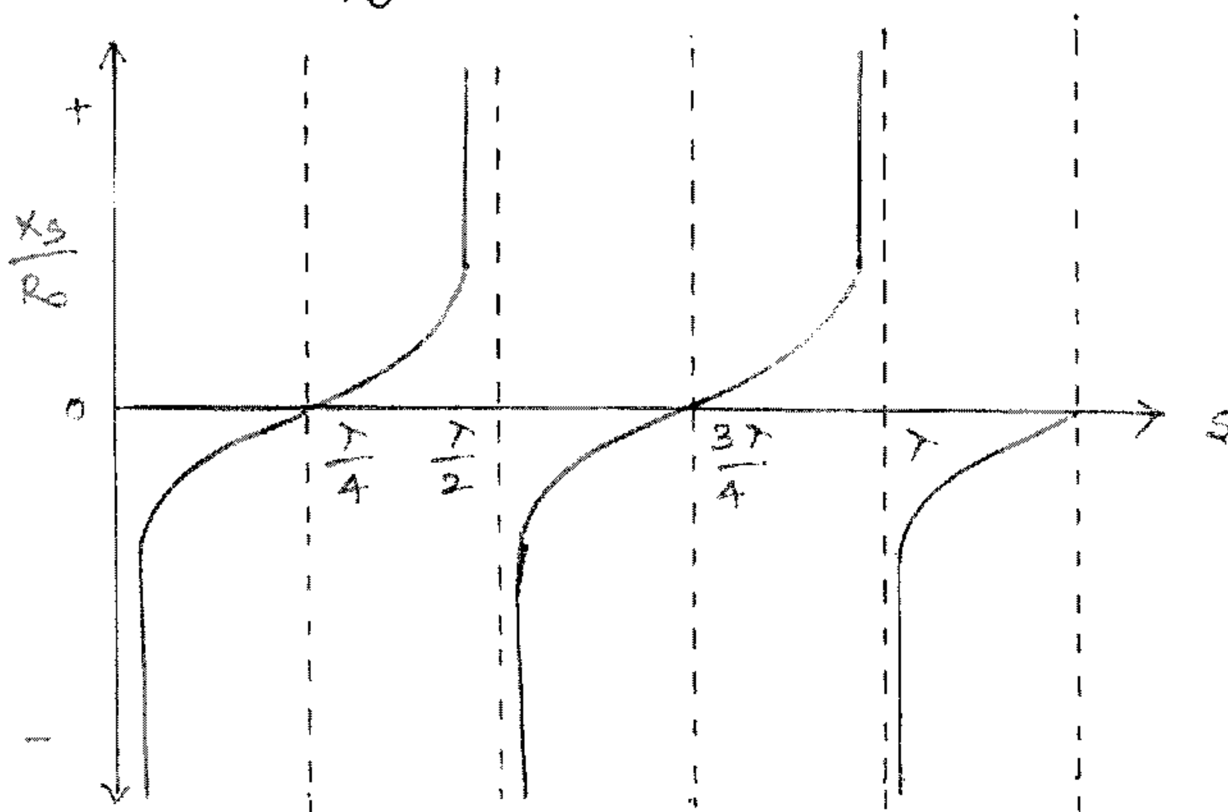
when $s = 0$, $\frac{X_s}{R_o} = -\cot\left(\frac{2\pi}{\lambda} \cdot 0\right) = -\cot(0) = -\infty$

$s = \frac{\lambda}{4}$, $\frac{X_s}{R_o} = -\cot\left(\frac{2\pi}{\lambda} \cdot \frac{\lambda}{4}\right) = -\cot\left(\frac{\pi}{2}\right) = 0$

$$s = \frac{\lambda}{2}, \quad \frac{X_s}{R_0} = -\cot\left(\frac{2\pi}{\lambda}\cdot\frac{\lambda}{2}\right) = -\cot(\pi) = +\infty$$

$$s = \frac{3\lambda}{4}, \quad \frac{X_s}{R_0} = -\cot\left(\frac{2\pi}{\lambda}\cdot\frac{3\lambda}{4}\right) = -\cot\left(\frac{3\pi}{2}\right) = 0$$

$$s = \lambda, \quad \frac{X_s}{R_0} = -\cot\left(\frac{2\pi}{\lambda}\cdot\lambda\right) = -\cot(2\pi) = -\infty$$



Voltage and Current on Dissipationless line :-

From general solution of txn. line, the voltage and current on a txn. line at any point can be expressed as,

$$E = \frac{E_R}{2}\left(\frac{Z_R + Z_0}{Z_R}\right)\left[e^{\gamma s} + K \cdot e^{-\gamma s}\right] \rightarrow (1)$$

(or)

$$E = E_R \cdot \cosh \gamma s + I_R Z_0 \cdot \sinh \gamma s \rightarrow (2)$$

Similarly,

$$I = \frac{I_R}{2}\left(\frac{Z_R + Z_0}{Z_0}\right)\left[e^{\gamma s} - K \cdot e^{-\gamma s}\right] \rightarrow (3)$$

(or)

$$I = I_R \cosh\gamma s + \frac{E_R}{Z_0} \sinh\gamma s \rightarrow (4)$$

For dissipationless line (or) zero dissipation line

$$Z_0 = R_0 \;,\; \gamma = j\beta \;,\; R = G = \alpha = 0$$

∴ equs. (1) → (4) becomes

$$E = \frac{E_R}{2}\left(\frac{Z_R + R_0}{Z_R}\right)\left[e^{j\beta s} + K \cdot e^{-j\beta s}\right] \rightarrow (5)$$

(or)

$$E = E_R \cdot \cosh(j\beta)s + I_R R_0 \sinh(j\beta)s$$

$$E = E_R \cdot \cos\beta s + j I_R R_0 \sin\beta s \rightarrow (6) ✓$$

Similarly,

$$I = \frac{I_R}{2}\left(\frac{Z_R + R_0}{R_0}\right)\left[e^{j\beta s} - K \cdot e^{-j\beta s}\right] \rightarrow (7)$$

(or)

$$I = I_R \cos\beta s + j \frac{E_R}{R_0} \sin\beta s \rightarrow (8) ✓$$

<u>For Short Circuit</u> :- $(Z_R = 0)$; $E_R = 0$ ✓

sub $Z_R = 0$ in equs. (6) & (8)

$$E_{SC} = I_R \cdot Z_R \cos\beta s + j I_R R_0 \sin\beta s$$

$$E_{SC} = 0 + j I_R R_0 \sin\beta s$$

$$\therefore E_{SC} = j I_R R_0 \sin\beta s \rightarrow (9)$$

$$I_{sc} = I_R \cos\beta s + j \frac{I_R Z_R}{R_0} \sin\beta s$$

$$\therefore I_{sc} = I_R \cos\beta s \rightarrow (10)$$

For Open Circuit : $(Z_R = \infty)$; $I_R = 0$

sub $Z_R = \infty$ in equs. (6) & (8)

$$E_{oc} = E_R \cos\beta s \rightarrow (11)$$

$$I_{oc} = j \frac{E_R}{R_0} \sin\beta s \rightarrow (12)$$

Power and Impedance measurement On dissipationless line :-

The expression for voltage and Current on the dissipationless line are given by

$$E = \frac{E_R}{2}\left(\frac{Z_R + R_0}{Z_R}\right)\left[e^{j\beta s} + K \cdot e^{-j\beta s}\right]$$

$$E = \frac{I_R}{2}(Z_R + R_0)\left[e^{j\beta s} + K \cdot e^{-j\beta s}\right] \rightarrow (1)$$

lly

$$I = \frac{I_R}{2}\left(\frac{R_0 + Z_R}{R_0}\right)\left[e^{j\beta s} - K \cdot e^{-j\beta s}\right] \rightarrow (2)$$

The Voltage and Current will be maximum when the reflected wave and incident wave are inphase.

The maximum voltage and current is expressed as,

$$E_{max} = \frac{I_R}{2}(Z_R + R_o)\left[1 + |K|\right] \rightarrow (3)$$

$$I_{max} = \frac{I_R}{2}\left(\frac{R_o + Z_R}{R_o}\right)\left[1 + |K|\right] \rightarrow (4)$$

The voltage and current will be minimum when the reflected wave and incident wave are out of phase

The minimum voltage and current is expressed a

$$E_{min} = \frac{I_R}{2}(Z_R + R_o)\left[1 - |K|\right] \rightarrow (5)$$

$$I_{min} = \frac{I_R}{2}\left(\frac{R_o + Z_R}{R_o}\right)\left\{1 - |K|\right] \rightarrow (6)$$

$$\frac{E_{max}}{I_{max}} = R_o \rightarrow (7)$$

$$\frac{E_{min}}{I_{min}} = R_o \rightarrow (8)$$

The resistive impedance seen at a voltage loop is

$$\frac{E_{max}}{I_{min}} = R_{max} = R_o \cdot \left(\frac{1 + |K|}{1 - |K|}\right) = R_o \cdot S$$

$$\therefore \boxed{R_{max} = R_o \cdot S} \rightarrow (9)$$

Since the voltage and current are again inphase at a current loop, the resistive impedance seen at a current loop is

$$\frac{E_{min}}{I} = R_{min} = R_o \cdot \left(\frac{1 - |K|}{1 + |K|}\right) = \frac{R_o}{S}$$

$$\therefore \boxed{R_{min} = \frac{R_o}{S}} \rightarrow (10)$$

The power passing a voltage loop is the power effectively flowing into a resistance R_{max} at voltage E_{max}, so that

$$P = \frac{E_{max}^2}{R_{max}} \rightarrow (11)$$

The same value of power must also pass the curren[t] loop, effectively flowing into a resistance R_{min} at voltage E_{min}, so that

$$P = \frac{E_{min}^2}{R_{min}} \rightarrow (12)$$

Then, $P^2 = \frac{E_{max}^2 \cdot E_{min}^2}{R_{max}\, R_{min}} \Rightarrow \frac{E_{max}^2 \cdot E_{min}^2}{\cancel{S} \cdot R_o \cdot \frac{R_o}{\cancel{S}}}$

$$P^2 = \frac{E_{max}^2 \cdot E_{min}^2}{R_o^2}$$

$$\therefore P = \frac{|E_{max}| \cdot |E_{min}|}{R_o} \rightarrow (13)$$

The power may also be expressed as,

$$P = |I_{max}| \cdot |I_{min}| \cdot R_o \rightarrow (14)$$

Measurement of Unknown Impedance :

The i/p impedance of a dissipationless line is given as,

$$Z_s = R_0 \cdot \left[\frac{Z_R + jR_0 \tan\beta s}{R_0 + jZ_R \tan\beta s} \right] \rightarrow (1)$$

R_{min} at a distance s' is given by,

$$Z_s(\min) = R_{min} = \frac{R_0}{S} \rightarrow (2)$$

sub (2) in (1)

$$\frac{R_0}{S} = R_0 \cdot \left[\frac{Z_R + jR_0 \tan\beta s'}{R_0 + jZ_R \tan\beta s'} \right]$$

$$\frac{1}{S} = \frac{Z_R + jR_0 \tan\beta s'}{R_0 + jZ_R \tan\beta s'}$$

$$R_0 + jZ_R \tan\beta s' = SZ_R + jSR_0 \tan\beta s'$$

$$R_0 - jSR_0 \tan\beta s' = SZ_R - jZ_R \tan\beta s'$$

$$R_0 (1 - jS \tan\beta s') = Z_R (S - j\tan\beta s')$$

$$\therefore Z_R = \frac{R_0 (1 - jS \tan\beta s')}{(S - j\tan\beta s')} \rightarrow (3)$$

Reflection loss in High frequency lines:

Reflection loss is defined as the ratio of power delivered to the load to the incident power.

$$\text{Reflection loss} = \frac{P_L}{P_i}$$

$$\text{Reflection loss in dB} = 10 \log \frac{P_L}{P_i}$$

$$= 10 \log \frac{P_i - P_r}{P_i}$$

$$= 10 \log \left(1 - \frac{P_r}{P_i}\right)$$

$$= 10 \log \left(1 - |K|^2\right)$$

$$= 10 \log \left(1 - \left(\frac{S-1}{S+1}\right)^2\right) \qquad \left(\because |K| = \frac{S-1}{S+1}\right)$$

$$= 10 \log \left(\frac{(S+1)^2 - (S-1)^2}{(S+1)^2}\right)$$

$$= 10 \log \left(\frac{4S}{(S+1)^2}\right)$$

$$= 10 \log \left(\frac{2\sqrt{S}}{S+1}\right)^2$$

$$\therefore \text{Reflection loss in dB} = 20 \log \left(\frac{2\sqrt{S}}{S+1}\right)$$

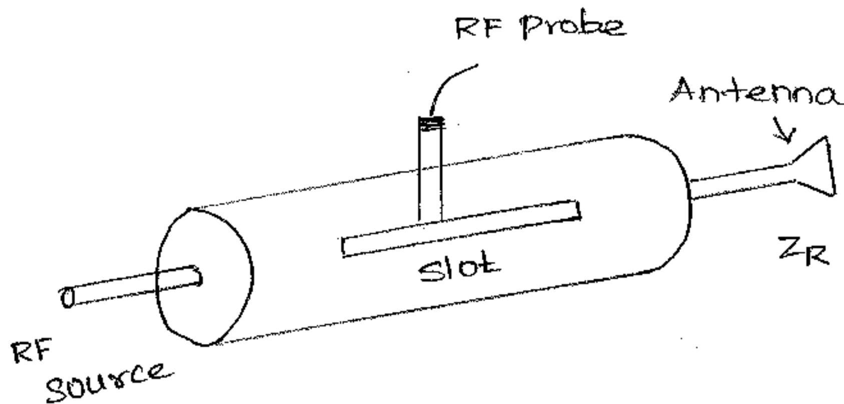
Measurement of VSWR and wavelength:

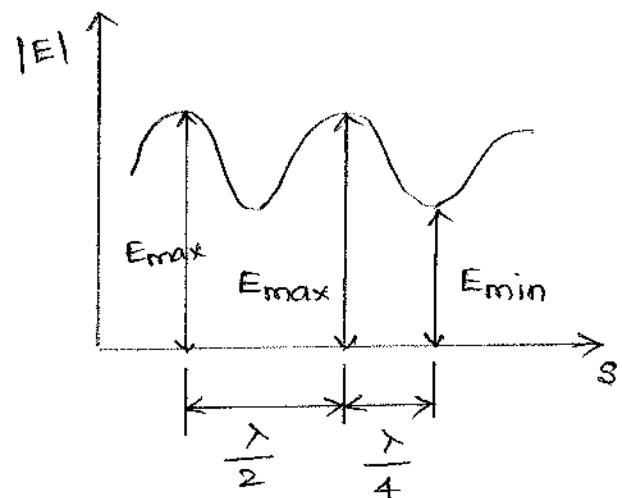
VSWR can be measured using 2 techniques.

1. Slotted line measurement
2. Directional coupler measurement

(1) Slotted line measurement:

This method is used to measure SWR and wavelength.





A longitudinal slot of length $n\lambda/2$ is cut on the coaxial line. A wire probe is inserted into the air dielectric of the line as a pickup device, a voltmeter or other detector connected between probe and sheath. Since the distance between V_{min} and V_{max} is $\lambda/4$, by placing the probe at a point which is $\lambda/4$ away

from Vmin, Vmax can be obtained. The ratio of Vmax to Vmin gives the value of SWR.

$$SWR = \frac{V_{max}}{V_{min}}$$

From the value of SWR, |K| can be calculated using equation,

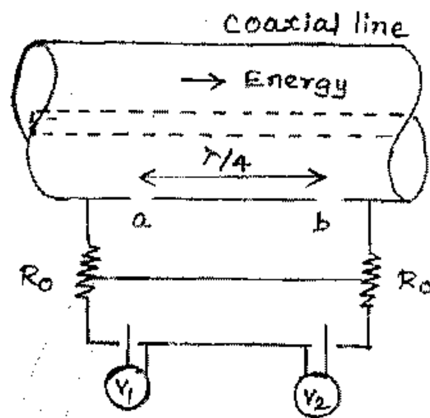
$$|K| = \frac{S-1}{S+1}$$

The same technique can also be used to measure the wavelength on the line. Wavelength can be calculated by considering the distance between two successive Vmax or Vmin as $\lambda/2$ or by considering the distance between Vmin & Vmax as $\lambda/4$. This kind of measurements are called Lecher measurements.

(2) Directional coupler measurement:

It consists of a section of coaxial transmission line, having two small holes in the outer sheath spaced by $\lambda/4$.

i) Directional coupler:



Clamped over these holes is a small section of line, terminated in its R_o value at both ends to prevent reflections.

Some energy will leak through the holes, and will set up a wave traveling to both left and right in the second line.

If the main line is transmitting energy to the right, then a wave entering the secondary line through hole a and traveling to the right will be in phase, setting up a wave traveling to the right in the secondary line. This gives an indication on V_2 & is considered as V_i.

When a wave entering the secondary line through hole 'a' travels in left direction, V_1 shows indication & this can be considered as V_r (reflected voltage).

The ratio of V_r to V_i gives the value of K (Reflection coefficient). From the value of K, SWR can be calculated using the equation,

$$SWR = \frac{1 + |K|}{1 - |K|}$$

Parameters of open wire line at high frequencies:

At high frequencies, current is considered to flow on the surface of the conductor and hence internal flux and internal inductance are reduced nearly to zero.

The inductance of open wire line is given as,

$$L = \frac{\mu_0}{2\pi} \ln\left(\frac{d}{a}\right) = 4 \times 10^{-7} \ln\left(\frac{d}{a}\right) \quad H/m$$

or $$L = 9.21 \times 10^{-7} \log\left(\frac{d}{a}\right) \quad H/m$$

The capacitance of a line is not affected by skin effect or frequency and hence it is given as,

$$C = \frac{\pi \epsilon}{\ln\left(\frac{d}{a}\right)} \quad F/m$$

$$= \frac{27.7}{\ln(d/a)} \quad \mu\mu F/m = \frac{12.07}{\log(d/a)} \quad \mu\mu F/m$$

The effective thickness of the surface layer of current may be considered as

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} \quad \text{meters}$$

where $\mu = \mu_0 = 4\pi \times 10^{-7}$ H/m , $\sigma = 5.75 \times 10^{7}$ mho/m

$$\therefore \quad \delta = \frac{0.0664}{\sqrt{f}}$$

The resistance of a round conductor of radius a meters to direct current is inversely proportional to the area as

$$R_{dc} = \frac{K}{\pi a^2}$$

While that of a round conductor with alternating current flowing in a skin of thickness δ is

$$R_{ac} = \frac{K}{2\pi a \delta}$$

Therefore the ratio of resistance to alternating current to resistance to direct current is

$$\frac{R_{ac}}{R_{dc}} = \frac{a\sqrt{\pi f \mu \sigma}}{2}$$

Which becomes, for copper,

$$\frac{R_{ac}}{R_{dc}} = 7.53\, a\sqrt{f}$$ (a → radius of the conductor)

The above equation shows that increase in resistance with increasing frequency is greater for

large radius than for small radius conductors.

The resistance of an open wire line of copper, with spacing greater than 20a, is given as

$$R_{ac} = \frac{8.33 \times 10^{-8} \sqrt{f}}{a} \text{ ohms/meter of line}$$

Parameters of coaxial line at high frequencies:

In coaxial line, due to skin effect at high frequencies, internal flux and internal inductance can be neglected. ∴ The inductance of the coaxial line is given as

$$L = \frac{\mu}{2\pi} \ln\left(\frac{b}{a}\right) = 2 \times 10^{-7} \ln\left(\frac{b}{a}\right) \text{ H/m}$$

$$L = 4.6 \times 10^{-7} \log\left(b/a\right) \text{ H/m}$$

The capacitance of the coaxial line, which is not affected by frequency, is given as

$$C = \frac{2\pi\epsilon}{\ln(b/a)} \text{ Fd/m}$$

$$= \frac{55.5\, \epsilon_r}{\ln(b/a)} = \frac{24.14\, \epsilon_r}{\log(b/a)} \ \mu\mu\text{F/m}$$

The resistance of the coaxial line is given as

$$R_{ac} = 4.16 \times 10^{-8} \sqrt{f} \left(\frac{1}{b} + \frac{1}{a} \right) \text{ ohms/m of the line}$$

where a → Outer radius of the inner conductor

b → Inner radius of the outer conductor

The shunt susceptance of the coaxial line is given as

$$y = g + j\omega c$$

The quality of the insulating material is measured in terms of power factor & is given as,

$$pf = \frac{g}{\sqrt{g^2 + \omega^2 c^2}}$$

For good insulating material, $g << \omega c$

$$\therefore \ pf = \frac{g}{\omega c}$$

The quality of the dielectric is also expressed in terms of the dissipation factor, which is the ratio of energy dissipated to energy stored in the dielectric per cycle.

UNIT 3

Impedance matching in High frequency lines

Impedance matching is the process of matching the load impedance to the characteristic impedance of the line or matching the line impedance to source impedance at the load or source side respectively.

Impedance matching can be achieved using

(1) Half wave line

(2) One Eighth wave line

(3) Quarter wave line

(4) Stub

(1) Half wave line or $\lambda/2$ line:

If the length of the transmission line is exactly equal to half the wavelength of the signal $(\lambda/2)$, then the line can be called as Half wave line.

To calculate its Z_{IN}, let us consider.

$$Z_{IN} = \frac{R_0 (Z_R + j R_0 \tan \beta s)}{R_0 + j Z_R \tan \beta s}$$

$s = l = \lambda/2$

$\beta s = \frac{2\pi}{\lambda} \cdot \frac{\lambda}{2} = \pi$

$\therefore \tan \beta s = \tan \pi = 0$

$$\therefore \boxed{Z_{IN} = Z_R}$$

Since $Z_{IN} = Z_R$, half wave line is also called as one to one transformer.

(2) <u>Quarter wave line or $\lambda/4$ line:</u>

If the length of the transmission line is exactly equal to $\lambda/4$, then the line can be called as quarter wave line or $\lambda/4$ line.

$$Z_{IN} = \frac{R_0 (Z_R + j R_0 \tan \beta s)}{R_0 + j Z_R \tan \beta s}$$

$s = l = \lambda/4$

$\beta s = \frac{2\pi}{\lambda} \cdot \frac{\lambda}{4} = \pi/2$

$\tan \beta s = \tan \pi/2 = \infty$

$$\therefore Z_{IN} = \frac{R_0 \cdot \tan \beta s \left(\frac{Z_R}{\tan \beta s} + j R_0 \right)}{\tan \beta s \left(\frac{R_0}{\tan \beta s} + j Z_R \right)}$$

If $\tan \beta s \to \infty$, then

$$\boxed{Z_{IN} = \frac{R_0^2}{Z_R}}$$

or $R_0 = \sqrt{Z_{IN} Z_R}$

A quarter wave line is used as a transformer to match a load of Z_R ohms to a source of Z_s ohms. As $Z_{IN} \propto \frac{1}{Z_R}$, $\lambda/4$ line is also called as impedance inverter & can be widely used in achieving impedance matching.

<u>Applications of $\lambda/4$ line:</u>

1. It is used as impedance inverter
2. It is used in impedance matching at source & load of a line
3. It is used to match the impedance of a transmission line to a resistive load such as antenna.
4. It is used to match any complex load to a line with Z_0.

5. It is used as an insulator as the input impedance of a short circuited $\lambda/4$ line is infinite.

(3) <u>One-eighth wave line or $\lambda/8$ line:</u>

If the length of the transmission line is exactly equal to $\lambda/8$, then the line can be called as $\lambda/8$ line.

$$Z_{IN} = \frac{R_0 (Z_R + jR_0 \tan \beta s)}{R_0 + jZ_R \tan \beta s}$$

$$s = l = \frac{\lambda}{8}$$

$$\beta s = \frac{2\pi}{\lambda} \cdot \frac{\lambda}{8} = \pi/4$$

$$\tan \beta s = \tan \pi/4 = 1$$

$$\therefore Z_{IN} = \frac{R_0 (Z_R + jR_0)}{R_0 + jZ_R}$$

$$|Z_{IN}| = \frac{R_0 \sqrt{Z_R^2 + R_0^2}}{\sqrt{Z_R^2 + R_0^2}} = R_0$$

Thus, $\lambda/8$ line is used to obtain magnitude match between R_0 & source impedance & therefore called as magnitude matching line.

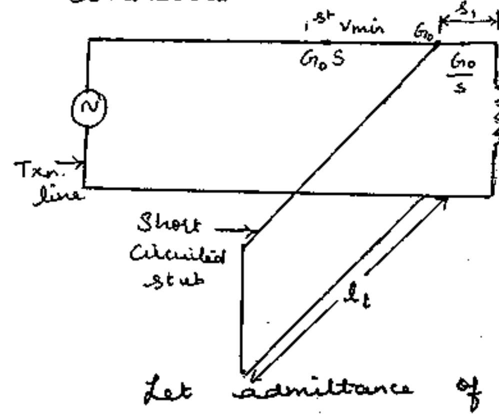
<u>Stub Matching:</u>

The process of achieving impedance matching using stub is called stub matching.

* Stub is a section of transmission line which can be used in achieving impedance matching.
* Stub has to be connected in parallel with the transmission line and nearer to the load.

<u>How stub is used in achieving impedance matching?</u>

Consider a transmission line which is improperly terminated ($Z_R \neq Z_0$). To match the impedance at the load, a stub has to be connected nearer to the load.



At 1st V_{min} point, $Y = G_0 S$

At 1st V_{max} point $Y = G_0 S$

$\therefore$ In between these two points, there must be a point on the line with $Y = G_0$. Exactly at that point stub has to be connected.

Let admittance of the line be,

$$Y_{line} = G \pm jB$$

(Where,
$G \to$ Conductance
$G = \frac{1}{R}$
$B \to$ Susceptance)

Let at point s_1, the Conductance of the line is G_0.

$$\therefore \text{ At } s_1, \quad Y_{line} = G_0 \pm jB$$

A stub has to be connected at point s_1, such that its susceptance is chosen to be equal & opposite to that of line.

$$\text{ie. } Y_{stub} = \mp jB$$

Since, line and stub are connected in parallel, the total admittance is given as,

$$Y_{total} = Y_{line} + Y_{stub}$$

$$= G_0 \pm jB \mp jB$$

$$Y = G_0$$

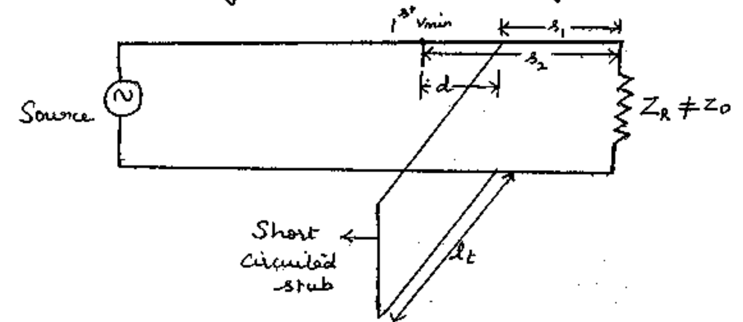
$$\Rightarrow Z = R_0$$

Thus, when stub is connected in parallel with the line at the point where conductance of the line is G_0, $Z = R_0$ & $\therefore$ impedance matching is achieved at the load.

A short circuited stub is preferred over open circuited stub because of less radiation loss & easy tuning.

Single stub matching

The process of achieving impedance matching by connecting single stub in parallel with the line is called Single stub matching.



Let s_2 → Distance between load & 1st V_{min}

l_s or s_1 → Distance between load and the stub (Location of stub)

d → distance between s_2 and s_1

L or l_t → Length of the stub

WKT,

$$Z_{IN} = \frac{R_0 \left(1 + |K| \angle \phi - 2\beta s\right)}{1 - |K| \angle \phi - 2\beta s} \quad \text{——— (1)}$$

For the design of stub, admittance can be considered.

$$\therefore \quad Y_{IN} = \frac{1}{R_0} \frac{\left(1 - |K| \angle \phi - 2\beta s\right)}{1 + |K| \angle \phi - 2\beta s}$$

$$Y_{IN} = G_0 \frac{(1 - |K| \angle \varphi - 2\beta s)}{(1 + |K| \angle \varphi - 2\beta s)}$$

$$\frac{Y_{IN}}{G_0} = Y_{in} = \frac{1 - |K| \angle \varphi - 2\beta s}{1 + |K| \angle \varphi - 2\beta s} \quad \text{---} (2)$$

[where Y_{in} is called normalised admittance]

Let $\varphi - 2\beta s = \theta$

$$\therefore Y_{in} = \frac{1 - |K| \angle \theta}{1 + |K| \angle \theta} = \frac{1 - |K| e^{j\theta}}{1 + |K| e^{j\theta}}$$

$$= \frac{1 - |K| (\cos\theta + j \sin\theta)}{1 + |K| (\cos\theta + j\sin\theta)}$$

$$Y_{in} = \frac{1 - |K| \cos\theta - j |K| \sin\theta}{1 + |K| \cos\theta + j |K| \sin\theta}$$

Taking complex conjugate of RHS we get,

$$Y_{in} = \frac{1 - |K| \cos\theta - j |K| \sin\theta}{1 + |K| \cos\theta + j |K| \sin\theta} \times \frac{1 + |K| \cos\theta - j |K| \sin\theta}{1 + |K| \cos\theta - j |K| \sin\theta}$$

On simplifying the above expression, we get,

$$Y_{in} = \frac{1 - |K|^2 - 2j |K| \sin\theta}{1 + |K|^2 + 2 |K| \cos\theta} \quad \text{---} (3)$$

As $Y_{in} = \frac{Y_{IN}}{G_0} = \frac{G_{IN} + jB_{IN}}{G_0}$, equation (3) can be written as,

$$\frac{G_{IN} + j B_{IN}}{G_0} = \frac{1 - |K|^2 - 2j |K| \sin\theta}{1 + |K|^2 + 2 |K| \cos\theta}$$

Equating real & imaginary terms, we get,

$$\frac{G_{IN}}{G_0} = \frac{1 - |K|^2}{1 + |K|^2 + 2 |K| \cos\theta} \quad \text{---} (4)$$

&

$$\frac{B_{IN}}{G_0} = \frac{-2 |K| \sin\theta}{1 + |K|^2 + 2 |K| \cos\theta} \quad \text{---} (5)$$

where $\theta = \varphi - 2\beta s$.

To find s_1:

At point s_1 (ie. the point of connecting the stub with the line), $G_{IN} = G_0$

$$\therefore \frac{G_{IN}}{G_0} = 1$$

Sub. $\frac{G_{IN}}{G_0} = 1$ in equation (4), we get

$$1 = \frac{1 - |K|^2}{1 + |K|^2 + 2 |K| \cos(\varphi - 2\beta s_1)}$$

($\because$ at $s = s_1$, $\theta = \varphi - 2\beta s_1$)

$$\Rightarrow 1 + |K|^2 + 2 |K| \cos(\varphi - 2\beta s_1) = 1 - |K|^2$$

$$2 |K| \cos(\varphi - 2\beta s_1) = -2 |K|^2$$

$$\cos(\varphi - 2\beta s_1) = -|K|$$

$$\varphi - 2\beta s_1 = \cos^{-1}(-|K|)$$

$$\varphi - 2\beta s_1 = \cos^{-1}|K| - \pi$$

$$\therefore \boxed{s_1 = \frac{\varphi + \pi - \cos^{-1}|K|}{2\beta}} \quad \text{---} (I)$$

(or)

$$\therefore \boxed{l_s = s_1 = \frac{\lambda}{4\pi}(\varphi + \pi - \cos^{-1}|K|)}$$

To find s_2:

At $s = s_2$, $G_{IN} = S G_0$.

$\therefore$ Sub. $\frac{G_{IN}}{G_0} = S$ in equation (4) [At $s = s_2$, $\theta = \varphi - 2\beta s_2$)

$$S = \frac{1 - |K|^2}{1 + |K|^2 + 2|K| \cos(\varphi - 2\beta s_2)}$$

$$\frac{1 + |K|}{1 - |K|} = \frac{1 - |K|^2}{1 + |K|^2 + 2|K| \cos(\varphi - 2\beta s_2)}$$

$$\Rightarrow 1 + |K|^2 + 2|K| \cos(\varphi - 2\beta s_2) = 1 - 2|K| + |K|^2$$

$$2|K| \cos(\varphi - 2\beta s_2) = -2|K|$$

$$\cos(\varphi - 2\beta s_2) = -1$$

$$\therefore \varphi - 2\beta s_2 = \cos^{-1}(-1)$$

$$\varphi - 2\beta s_2 = -\pi$$

$$\therefore \boxed{s_2 = \frac{\varphi + \pi}{2\beta}} \quad \text{—— (II)}$$

To find d:

$$d = s_2 - s_1$$

$$= \frac{\varphi + \pi}{2\beta} - \frac{\varphi + \pi - \cos^{-1}|K|}{2\beta}$$

$$\boxed{d = \frac{\cos^{-1}|K|}{2\beta}} \quad \text{—— (III)}$$

To find l_t (length of the stub):

Equation (5) $\Rightarrow$ $\frac{B_{IN}}{G_0} = \frac{-2|K| \sin\theta}{1 + |K|^2 + 2|K| \cos\theta}$ [where $\theta = \varphi - 2\beta s$]

At $s = s_1$, $\cos(\varphi - 2\beta s_1) = -|K|$

Let $\theta = \varphi - 2\beta s_1$; $\cos\theta = -|K|$

$$\sin^2\theta + \cos^2\theta = 1$$

$$\therefore \sin\theta = \sqrt{1 - |K|^2}$$

Sub. $\sin\theta = \sqrt{1 - |K|^2}$ & $\cos\theta = -|K|$ in eqn (5),

$$\frac{B_{IN}}{G_0} = \frac{-2|K| \cdot \sqrt{1 - |K|^2}}{1 + |K|^2 + 2|K|(-|K|)}$$

$$= \frac{-2|K| \sqrt{1 - |K|^2}}{1 - |K|^2}$$

$$\frac{B_{IN}}{G_0} = \frac{-2|K|}{\sqrt{1 - |K|^2}}$$

$$\therefore B_{IN} \text{ of a line} = G_0 \cdot \frac{-2|K|}{\sqrt{1 - |K|^2}}$$

B_{stub} (B of a stub) must be equal to $G_0 \cdot \frac{2|K|}{\sqrt{1 - |K|^2}}$ —— (6)

[Bcoz The susceptance of the stub must be equal & opposite to that of line].

For a short circuit stub, $Z_{sc} = j R_0 \tan\beta s$

$\therefore$ For a S.C stub of length 'l_t',

$$Y = \frac{1}{j R_0 \tan\beta l_t} = \frac{-j G_0}{\tan\beta l_t}$$

$$j B_{sc} = \frac{-j G_0}{\tan\beta l_t}$$

$$\therefore |B_{sc}| = \frac{G_0}{\tan\beta l_t} \quad \text{——— ⑦}$$

Equating equation ⑥ and ⑦ we get,

$$G_0 \cdot \frac{2|K|}{\sqrt{1-|K|^2}} = \frac{G_0}{\tan\beta l_t}$$

$$\tan\beta l_t = \frac{\sqrt{1-|K|^2}}{2|K|}$$

$$\therefore \beta l_t = \tan^{-1}\left(\frac{\sqrt{1-|K|^2}}{2|K|}\right)$$

$$\boxed{l_t = \frac{\lambda}{2\pi} \tan^{-1}\left(\frac{\sqrt{1-|K|^2}}{2|K|}\right)} \quad \left[\because \beta = \frac{2\pi}{\lambda}\right]$$

where $|K| = \frac{S-1}{S+1}$

Summary:

Location of stub $\rightarrow l_s$ or $s_1 = \dfrac{\phi + \pi - \cos^{-1}|K|}{2\beta}$

Location of 1st $V_{min} \rightarrow s_2 = \dfrac{\phi+\pi}{2\beta}$

$$d = \frac{\cos^{-1}|K|}{2\beta}$$

Length of the stub, $l_t = \dfrac{\lambda}{2\pi} \tan^{-1}\left(\dfrac{\sqrt{1-|K|^2}}{2|K|}\right)$

or $l_t = \dfrac{\lambda}{2\pi} \tan^{-1}\left(\dfrac{\sqrt{S}}{S-1}\right)$

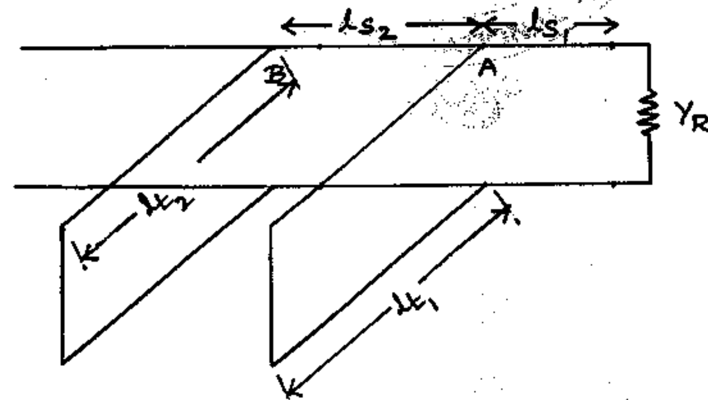
Disadvantages :-

1. The single stub matching is useful for fixed frequency only because as the frequency changes, the location of the stub will have to be changed.

2. For final adjustment, the stub has to be moved along the line slightly. This is possible only in open wire lines and not on coaxial lines.

DOUBLE STUB MATCHING :-

To Overcome the disadvantages of single stub matching two short circuited stubs whose lengths are adjustable independently but whose positions are fixed



Let the 1st stub whose length is l_t, be located at point 'A'. at a distance of l_{s_1} from the load end.

The input admittance at point A is given by

$$Y_A = Y_0 \cdot \left[\frac{Y_R + jY_0 \tan\beta s}{Y_0 + jY_R \tan\beta s}\right] \rightarrow ①$$

Find the normalized admittance

$$\frac{Y_A}{Y_0} = y_A = \left[\frac{y_r + j\tan\beta s}{1 + jy_r \tan\beta s}\right] \rightarrow (2)$$

Put $s = l_{s_1}$ and rationalize equ (2)

$$y_A = \frac{y_r + j\tan\beta l_{s_1}}{1 + jy_r\tan\beta l_{s_1}} \times \frac{1 - jy_r\tan\beta l_{s_1}}{1 - jy_r\tan\beta l_{s_1}}$$

$$= \frac{y_r(1 - jy_r\tan\beta l_{s_1}) + j\tan\beta l_{s_1}(1 - jy_r\tan\beta l_{s_1})}{(1 + jy_r\tan\beta l_{s_1})(1 - jy_r\tan\beta l_{s_1})}$$

$$= \frac{y_r - jy_r^2\tan\beta l_{s_1} + j\tan\beta l_{s_1} + y_r\tan^2\beta l_{s_1}}{1 + y_r^2\tan\beta l_{s_1}}$$

$$y_A = \frac{y_r(1 + \tan^2\beta l_{s_1}) + j\tan\beta l_{s_1}(1 - y_r^2)}{1 + y_r^2\tan\beta l_{s_1}}$$

Equate real and imaginary parts

$$g_A = \frac{y_r(1 + \tan^2\beta l_{s_1})}{1 + y_r^2\tan\beta l_{s_1}} \rightarrow (3)$$

$$b_A = \frac{\tan\beta l_{s_1}(1 - y_r^2)}{1 + y_r^2\tan\beta l_{s_1}} \rightarrow (4)$$

When a stub having a susceptance b_1 is added at this point A, the new admittance value will be, $y_A' = g_A + jb_A' \rightarrow (5)$

Since only the susceptance value is altered by the addition of the stub, the conductance part remain: unchanged. Here $b_A' = b_A \pm b_1$.

The input admittance of line at point 'B' should be equal to 'G_0' so that the line appears to be terminated with Y_0. The normalized admittance at point B is

$$y_B = 1 + jb_B \rightarrow (6)$$

Finally the length of stub 2 is adjusted to produc a susceptance $-jb_B$ and the desired admittance is '1' at point B.

Typically 2 stubs are separated by fixed distances $\frac{\lambda}{4}, \frac{\lambda}{8}, \frac{\lambda}{16}, \frac{3\lambda}{8}$. The most commonly preferred is $\frac{\lambda}{4}$ and $\frac{3\lambda}{8}$.

Matching takes place between point B and the generator. So, there are chances of having reflection loss in between point B and the Load. In order to minimize the loss, the stubs are located very close to the load. Sometimes the first stub is located at load itself. But the common practice is to keep distance of 0.1λ to 0.15λ between load and 1st stub.

SMITH CHART PROCEDURE

1. Calculation of SWR:

Step 1: Locate the given normalized load impedance point and let it be 'P'.

Step 2: With '0' as centre and OP as radius draw a circle.

Step 3: The right hand intersection of the circle and the horizontal axis gives the value of SWR.

2. Calculation of Reflection coefficient, k:

Step 1: Locate the given normalized load impedance point and let it be 'P'.

Step 2: Draw the line OP and extent it to cut the 'Angle of reflection coefficient circle' at point P'. Note down the angle corresponding to this point. This gives angle of k.

Step 3: Measure the line length of OP and OP'. The ratio of OP to OP' gives the value of magnitude of k.

3. Procedure for single stub matching:

Step 1: Find the normalized load impedance and mark it as point A.

Step 2: With OA as radius draw S circle

Step 3: To find the normalized load admittance, draw a line joining OA and extend the same line till it cuts the S Circle. Mark the intersection point as B.

Step 4: To find the position of load, extend the OA line till it cuts the wavelength scale (outer circle). Mark it as point C

Step 5: Draw a unit circle (R=1)

Step 6: Select the intersection point of S circle with unit circle nearer to the load. Mark it as point D

Step 7: To find the position of stub on the main line, draw a line joining OD and extend the same line till it cuts the outer circle. Mark the point as E.

Step 8: To find the position of the stub 'l_s' from the load subtract D from E (i.e., E – D).

Step 9: Find the imaginary value at point D and mark the opposite value of it as F.

Step 10: Draw a line joining OF till it cuts the outer circle. Mark the intersection point on outer circle as G

Step 11: To find the total length of the stub 'l_t', subtract 0.25 from G point.

WAVELENGTH TOWARDS GENERATOR

RHS

Angle of Reflection coefficient

+ve Reactance circle value

+ve Reactance (Inductance)

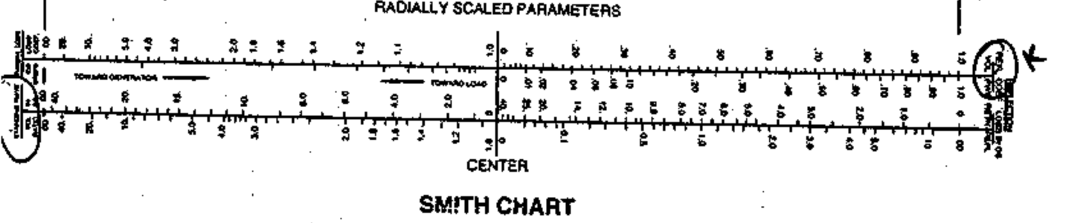
S circle

-ve Reactance (capacitance)

-ve Reactance circle value

WAVELENGTH TOWARDS LOAD

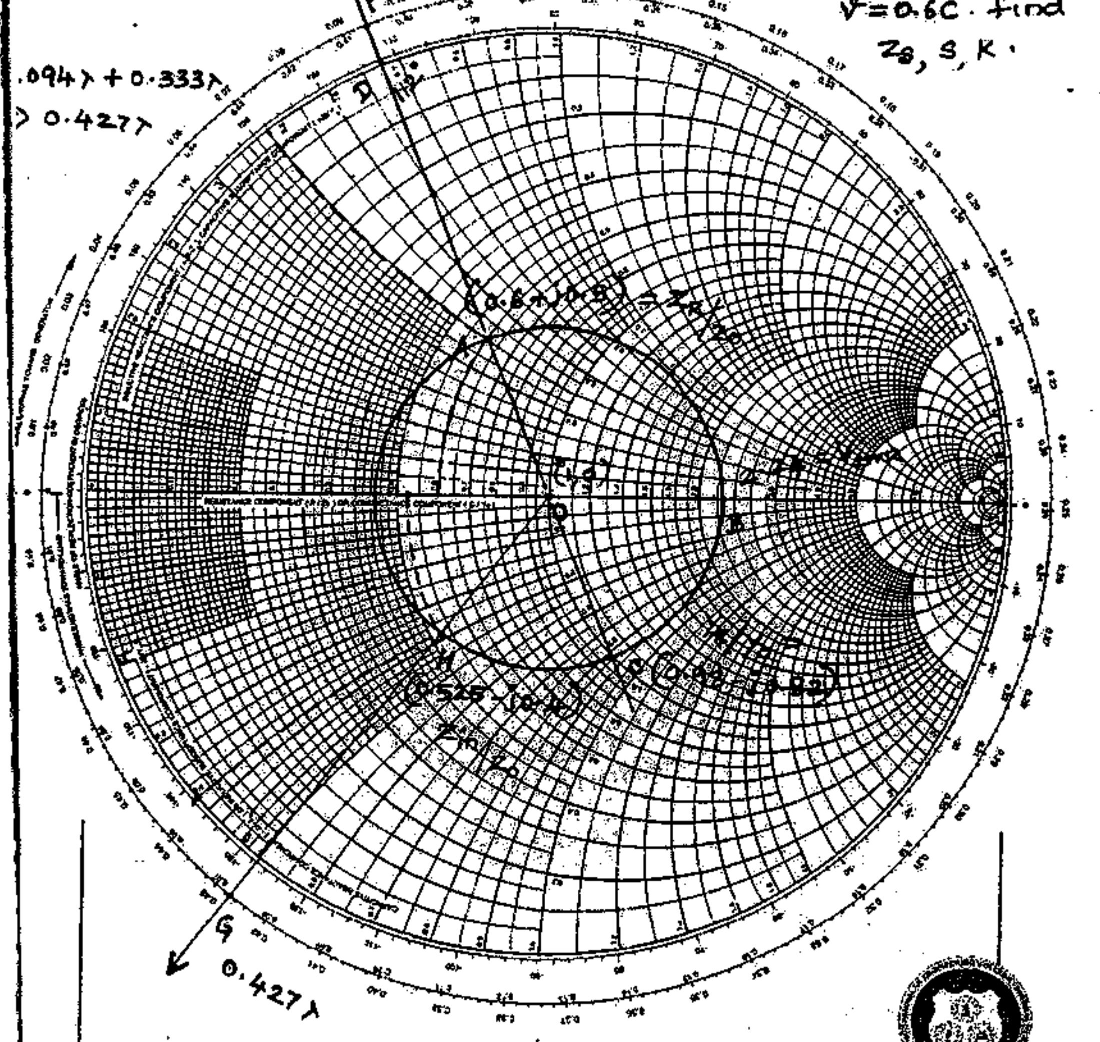
ANGLE OF REFLECTION IN DEGREES

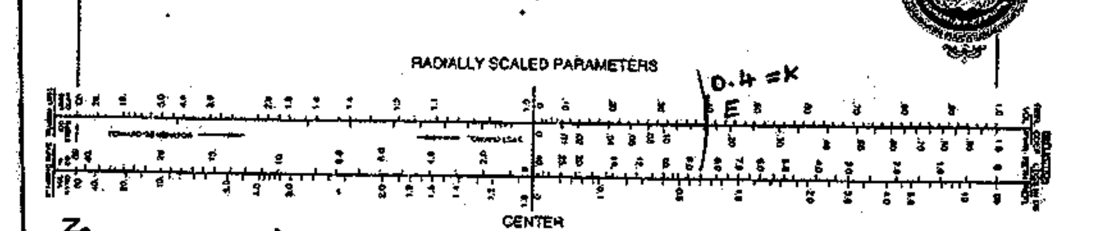


APPLICATIONS OF SMITH CHART

$(180+j150)\Omega$. If the length of the line is 30m long operating at a freq of 2MHz & $v=0.6c$ find Z_s, S, K.

$0.094\lambda + 0.333\lambda = 0.427\lambda$





$$\frac{Z_s}{Z_0} = 0.525 - j0.4$$

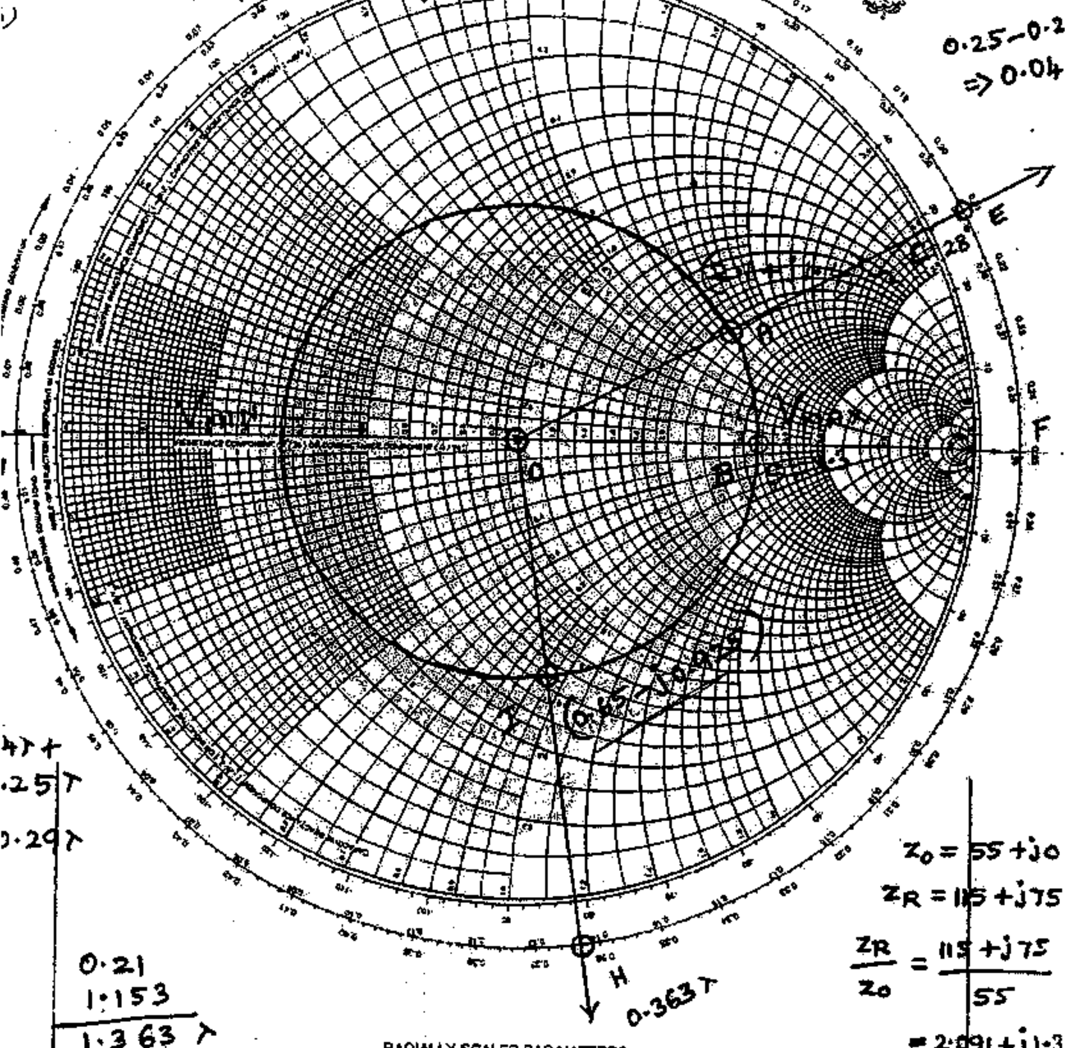
$$Z_s = 300\,(0.525 - j0.4)$$

$$= 158 - j120\,\Omega$$

a Load imp., of $(115+j75)\Omega$.

i) what is K in polar form?

ii) " " value of S?

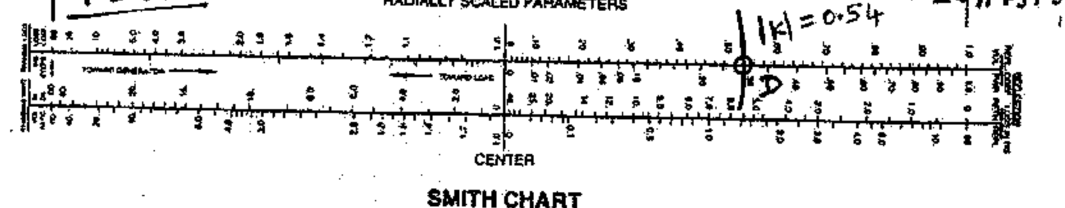


$$\frac{Z_R}{Z_0} = \frac{115+j75}{55} = 2.091+j1.3$$

$|K| = 0.54$

0.21
1.153
1.363 λ

.47 +
.25 λ
0.29 λ



iii) At what distance from the load will the 1st maximum and minimum of voltage will occur?

iv) If the line is 1.153 λ long what will be i/p Impedance?

(11)

$l_s = 0.238\lambda$, Distance Between stub and Load

$l_s = 0.238\lambda$

Admittance chart

S-circle

0.182 λ E

1+j1.8

D G

A

F 0.444 λ

G

1-j1.8 B

H 0.331 λ

Length of the stub $l_L = 0.081\lambda$

$l_s = E - F$

$= 0.182\lambda - 0.444\lambda$

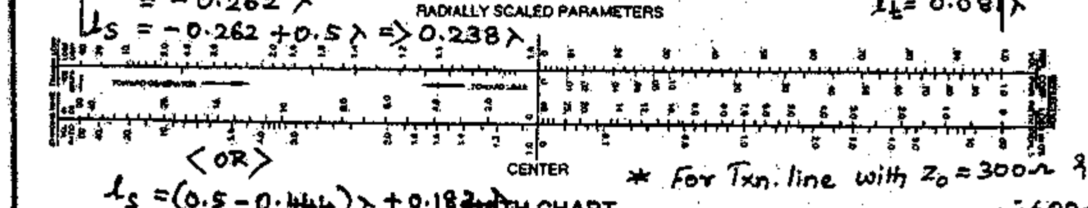
$= -0.262\lambda$

$l_s = -0.262 + 0.5\lambda \Rightarrow 0.238\lambda$

H - I

$l_L = 0.331\lambda - 0.25$

$l_L = 0.081\lambda$



<OR>

$l_s = (0.5 - 0.444)\lambda + 0.182\lambda$

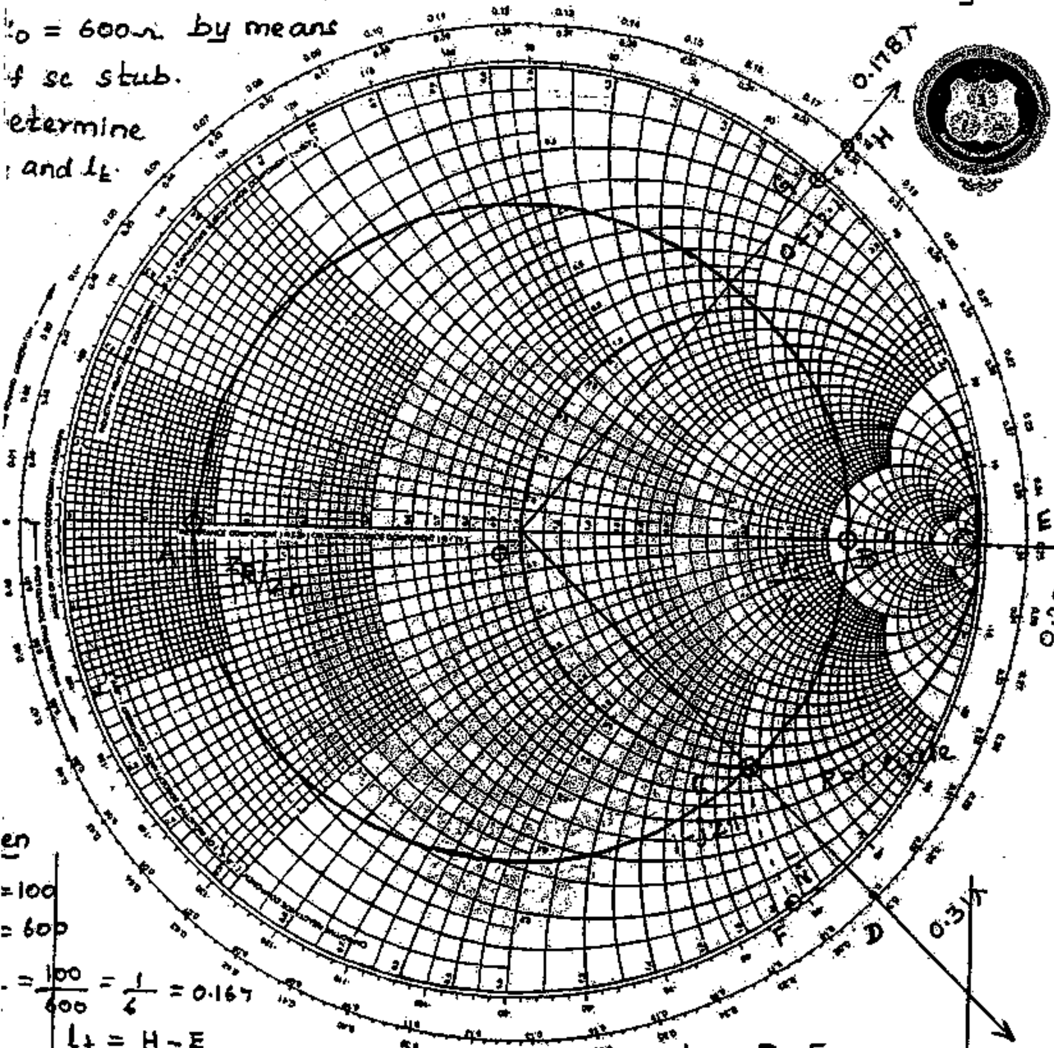
$= 0.056\lambda + 0.182\lambda$

$= 0.238\lambda$

* For Txn. line with $Z_0 = 300\Omega$ & Load Impedance $Z_L = 390+j600\Omega$.

Determine SWR, the distance a SC stub must be placed from the Load &

...be matched at a freq. of 100 MHz to a Txn. line having $Z_0 = 600\,\Omega$ by means of sc stub. Determine l_s and l_t.



Given

$Z_R = 100$

$Z_0 = 600$

$z_R = \frac{100}{600} = \frac{1}{6} = 0.167$

$l_t = H - E$

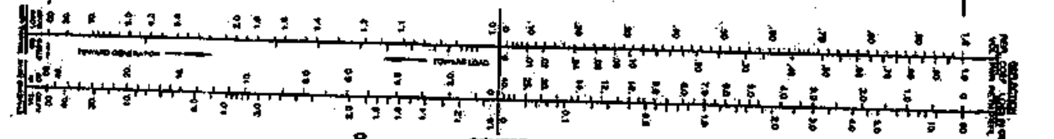
$= 0.178 - 0.25\lambda$

$= 0.428\lambda$

$l_s = D - E$

$0.31\lambda - 0.25\lambda$

0.06λ



$\lambda = \frac{c}{f} = \frac{3\times10^8}{100\times10^6}$

$= 3\,m$

$l_s = 0.06\lambda$

$= 0.06\times3$

$= 0.18\,m$

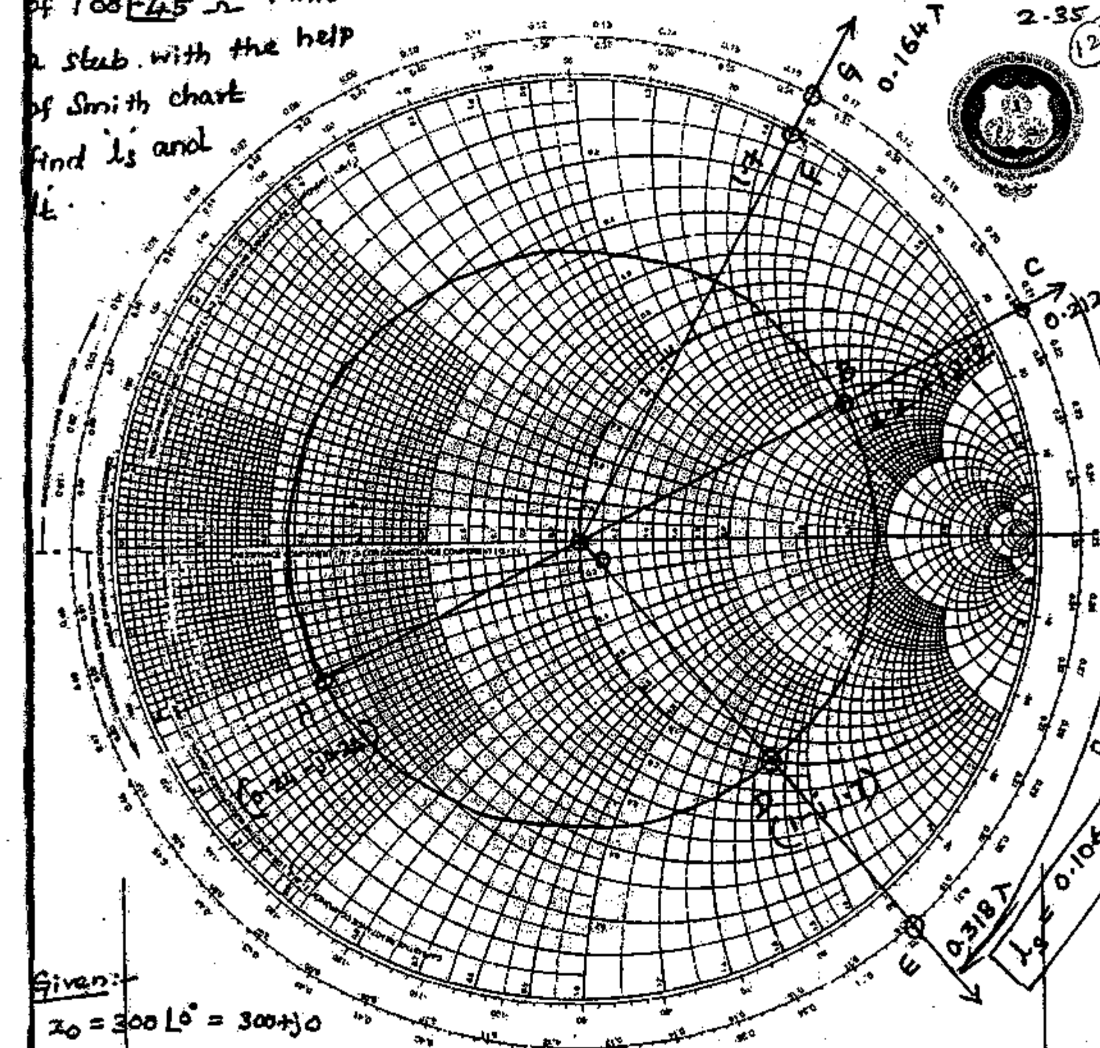
$l_t = 0.428\lambda$

$= 0.428\times3$

$= 1.284\,m$

...of $100\angle{-45^\circ}\,\Omega$. This load is to be matched to a txn. line by using a stub with the help of Smith chart. Find l_s and l_t.

2.35 (12)



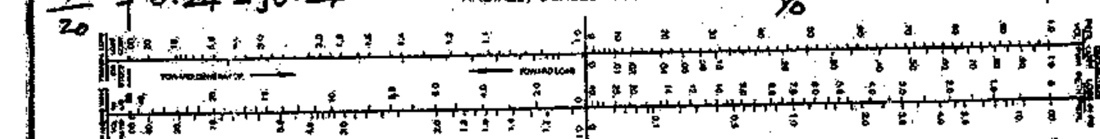
Given:

$z_0 = 300\angle0^\circ = 300 + j0$

$z_R = 100\angle{-45^\circ} = 70.71 - j70.71$

$\frac{z_R}{z_0} = 0.24 - j0.24$

$\frac{Y_R}{Y_0} = 2.12 + j2.12$



$l_t = (0.25 + 0.164)\lambda$

$l_t = 0.414\lambda$

$l_s = 0.122\lambda$

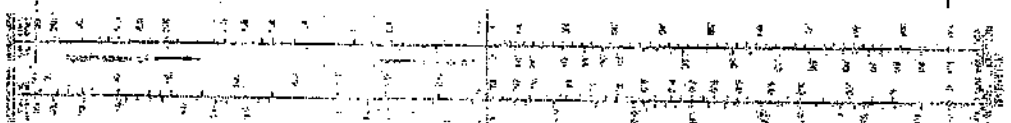
0.185
0.063
0.122

0.063λ   0.185λ

$l_s = 0.076\lambda$

F = 0.326λ

0.326
0.25
0.076λ



2) A Load $(50 - j100)\Omega$ is connected across a 50Ω line. Design a short circuited stub to provide matching between the 2 lines at a signal Freq. of 30 MHz using Smith chart.

DOUBLE STUB MATCHING

(13)

in an impedance of $40 - j120\Omega$. This load is to be matched to a tx line by using stub with the help of Smith chart find l_{t_1}, l_{t_2}.

D-E

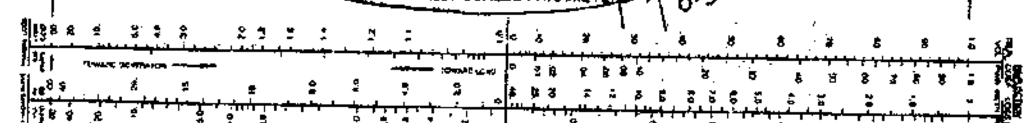
$l_{t_1} = 0.097 - 0.25$
$= 0.347\lambda$

Length of stub 1 = 0.347λ

Length of stub 2 = 0.11λ

I-E

$l_{t_2} = 0.36 - 0.25$
$= 0.11\lambda$



$l_{s_1} = F - J = 0.418\lambda - 0.355\lambda$
$= 0.063\lambda$

$l_{s_2} = K - F = 0.168\lambda - 0.418\lambda$

$\dfrac{Z_L}{Z_0} = \dfrac{40 - j120}{100} = 0.4 - j1.2$

Advantages of short circuited stub over open circuited stub:

1. S.c stub has lower loss of energy due to radiation
2. Greater ease in construction
3. Length of S.c stubs can easily be changed by using a sliding short circuit.

Disadvantages of Single stub matching:

1. In single stub matching, location and length of the stub depend on load impedance & signal frequency. Any change in this will require corresponding change in length and position of the stub.
2. Single stub matching requires that the stub be located at a definite point on the line. This requirement often calls for placement of the stub at an undesirable place from mechanical consideration.

Transmission of TM Waves Between Parallel Planes:

To find the field configuration or components inside the parallel planes, consider 2 plates are placed at a distance 'a' in the x-axis from 'o' to 'a'. Assume the wave is propagating in z-direction and there is no boundary in 'y' and 'z' direction.

The maxwell's equation to be satisfied by the electric and magnetic field at the boundary are,

$$\nabla \times E = -j\omega\mu H \rightarrow (1)$$

$$\nabla \times H = j\omega\varepsilon E \rightarrow (2)$$

equation (1) can be written as,

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -j\omega\mu H_x \rightarrow (3)$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \rightarrow (4)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \rightarrow (5)$$

similarly, equ. (2) can be written as

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = j\omega\varepsilon E_x \rightarrow (6)$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = j\omega\varepsilon E_y \rightarrow (7)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\varepsilon E_z \rightarrow (8)$$

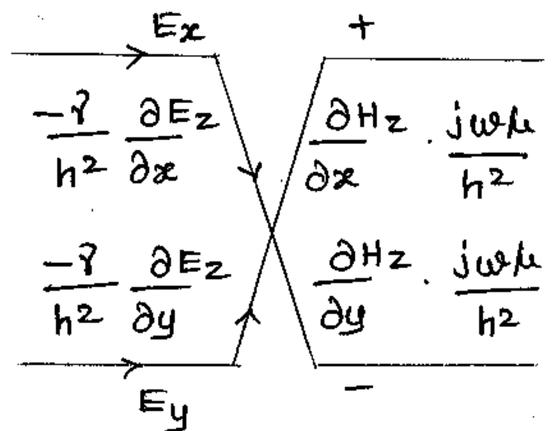
Manipulating equs. ③ → ⑧, the relation between field components inside the guide can be obtained as,

$$E_x = \frac{-\gamma}{h^2}\frac{\partial E_z}{\partial x} - \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial y} \rightarrow (9)$$

$$E_y = \frac{-\gamma}{h^2}\frac{\partial E_z}{\partial y} + \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x} \rightarrow (10)$$

$$H_x = \frac{-\gamma}{h^2}\frac{\partial H_z}{\partial x} + \frac{j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial y} \rightarrow (11)$$

$$H_y = \frac{-\gamma}{h^2}\frac{\partial H_z}{\partial y} - \frac{j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial x} \rightarrow (12)$$



where, $h^2 = \gamma^2 + \omega^2\mu\varepsilon$

For TM waves, $H_z = 0$

and for $\parallel^{el}$ plates, $\frac{\partial}{\partial y} = 0$ & $\gamma = \gamma_m$

so, equs ⑨ → ⑫ becomes

$$E_x = \frac{-\gamma_m}{h^2}\frac{\partial E_z}{\partial x} \rightarrow (13)$$

$$E_y = 0 \rightarrow (14)$$

$$H_x = 0 \rightarrow (15)$$

$$H_y = \frac{-j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial x} \rightarrow (16)$$

From equs. ⑬ → ⑯, the existing field components inside the parallel plates are,

E_x, E_z and H_y

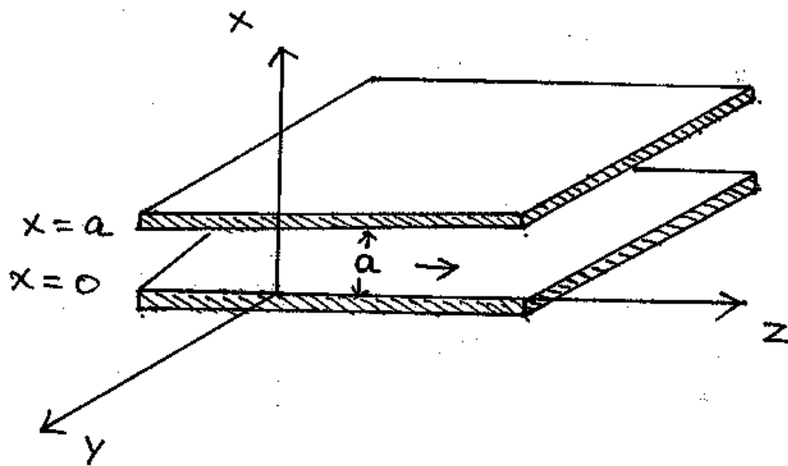
To find the field components of TM waves inside the parallel plates, we can assume a value for one of the field component and from this value, we can get the value of other field components.

Let us assume a sine wave is propagating in z-direction. Let the value of E_z be,

$$E_z = \frac{h}{j\omega\varepsilon}\left[B_1 \cosh x - B_2 \sinh x\right] \rightarrow (17)$$

where, $h = \frac{m\pi}{a}$

'a' is the distance between 2 planes in x-axis and and 'm' is the integer having the value $m = 0, 1, 2, 3 \ldots$



The boundary conditions are,

$x = 0 \quad E_z = 0$

$x = a \quad E_z = 0$

Differentiate equ. (17) w.r.to 'x' and 'y'

$$\frac{\partial E_z}{\partial x} = \frac{-h^2}{j\omega\varepsilon}\left[B_1 \sinh x + B_2 \cosh x\right] \rightarrow (18)$$

sub equ. (18) in (13) → (16)

$$E_x = \frac{-\gamma_m}{h^2} \times \frac{-h^2}{j\omega\varepsilon} \left[B_1 \sin hx + B_2 \cos hx \right]$$

$$E_x = \frac{\gamma_m}{j\omega\varepsilon} \left[B_1 \sin hx + B_2 \cos hx \right] \rightarrow (19)$$

$$E_y = 0 \rightarrow (20)$$

$$H_x = 0 \rightarrow (21)$$

$$H_y = \frac{-j\omega\varepsilon}{h^2} \times \frac{-h^2}{j\omega\varepsilon} \left[B_1 \sin hx + B_2 \cos hx \right]$$

$$H_y = B_1 \sin hx + B_2 \cos hx \rightarrow (22)$$

At HF, $\gamma_m = j\beta_m$

so, sub $\gamma_m = j\beta_m$ and $h = \frac{m\pi}{a}$ in equs. (19) → (22), (17)

$$E_x = \frac{j\beta_m}{j\omega\varepsilon} \left[B_1 \sin\left(\frac{m\pi}{a} x\right) + B_2 \cos\left(\frac{m\pi}{a} x\right) \right]$$

$$E_x = \frac{\beta_m}{\omega\varepsilon} \left[B_1 \sin\left(\frac{m\pi}{a} x\right) + B_2 \cos\left(\frac{m\pi}{a} x\right) \right] \rightarrow (23)$$

$$E_y = 0 \rightarrow (24)$$

$$H_x = 0 \rightarrow (25)$$

$$H_y = B_1 \sin\left(\frac{m\pi}{a} x\right) + B_2 \cos\left(\frac{m\pi}{a} x\right) \rightarrow (26)$$

$$E_z = \frac{m\pi}{ja\omega\varepsilon} \left[B_1 \cos\left(\frac{m\pi}{a} x\right) - B_2 \sin\left(\frac{m\pi}{a} x\right) \right] \rightarrow (27)$$

The field components of TM wave inside the $\parallel^{el}$

Planes can be represented interms of time and propagation variation.

$$E_x = \frac{\beta_m}{\omega\varepsilon}\left[B_1 \sin\left(\frac{m\pi}{a}x\right) + B_2 \cos\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} \cdot \sin\omega t \rightarrow (28)$$

$$E_y = 0 \rightarrow (29)$$

$$E_z = \frac{-m\pi}{a\omega\varepsilon}\left[B_1 \cos\left(\frac{m\pi}{a}x\right) - B_2 \sin\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} \cdot \cos\omega t \rightarrow (30)$$

$$H_x = 0 \rightarrow (31)$$

$$H_y = \left[B_1 \sin\left(\frac{m\pi}{a}x\right) + B_2 \cos\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} \cdot \sin\omega t \rightarrow (32)$$

The field components of TM waves inside the Parallel plates can be summarized as follows.

$$E_x = \frac{\beta_m}{\omega\varepsilon}\left[B_1 \sin\left(\frac{m\pi}{a}x\right) + B_2 \cos\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} \cdot \sin\omega t$$

$$E_y = 0$$

$$E_z = \frac{-m\pi}{a\omega\varepsilon}\left[B_1 \cos\left(\frac{m\pi}{a}x\right) - B_2 \sin\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} \cdot \cos\omega t$$

$$H_x = 0$$

$$H_y = \left[B_1 \sin\left(\frac{m\pi}{a}x\right) + B_2 \cos\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} \cdot \sin\omega t$$

$$H_z = 0$$

Transmission of TE Waves Between Parallel Planes:

To find the field configuration or components inside the Parallel Planes, consider 2 plates are placed at a distance 'a' in the x-axis from '0' to 'a'. Assume the wave is propagating in the z-direction and there is no boundary in the y and z direction.

The maxwells equation to be satisfied by the electric and magnetic field at the boundary are,

$$\nabla \times E = -j\omega\mu H \rightarrow (1)$$

$$\nabla \times H = j\omega\varepsilon E \rightarrow (2)$$

x y z

equation (1) can be written as

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -j\omega\mu H_x \rightarrow (3)$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \rightarrow (4)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \rightarrow (5)$$

Similarly, equation (2) can be written as

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = j\omega\varepsilon E_x \rightarrow (6)$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = j\omega\varepsilon E_y \rightarrow (7)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\varepsilon E_z \rightarrow (8)$$

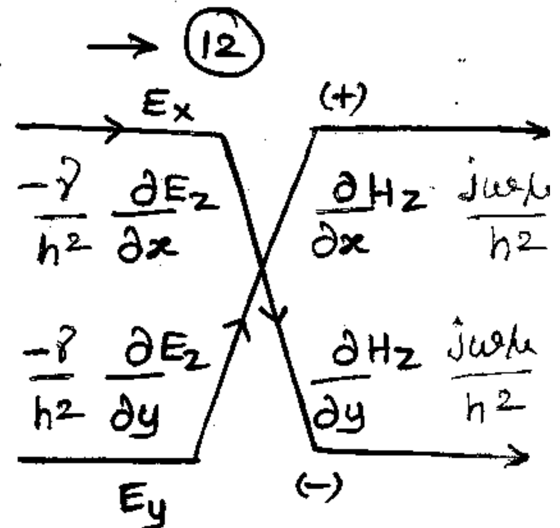
Manipulating equs. ③ → ⑧, the relation B/W field components inside the guide can be obtained as,

$$E_x = \frac{-\gamma}{h^2}\frac{\partial E_z}{\partial x} - \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial y} \rightarrow (9)$$

$$E_y = \frac{-\gamma}{h^2}\frac{\partial E_z}{\partial y} + \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x} \rightarrow (10)$$

$$H_x = \frac{-\gamma}{h^2}\frac{\partial H_z}{\partial x} + \frac{j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial y} \rightarrow (11)$$

$$H_y = \frac{-\gamma}{h^2}\frac{\partial H_z}{\partial y} - \frac{j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial x} \rightarrow (12)$$



where, $h^2 = \gamma^2 + \omega^2\mu\cdot\varepsilon$

For TE Waves, $E_z = 0$

and for $\|^{el}$ plates, $\frac{\partial}{\partial y} = 0$

$$E_x = 0 \rightarrow (13)$$

$$E_y = \frac{j\omega\mu}{h^2}\cdot\frac{\partial H_z}{\partial x} \rightarrow (14)$$

$$H_x = \frac{-\gamma}{h^2}\cdot\frac{\partial H_z}{\partial x} \rightarrow (15)$$

$$H_y = 0 \rightarrow (16)$$

From equs. ⑬ → ⑯, the existing field components inside the parallel plates are,

H_x, H_z and E_y

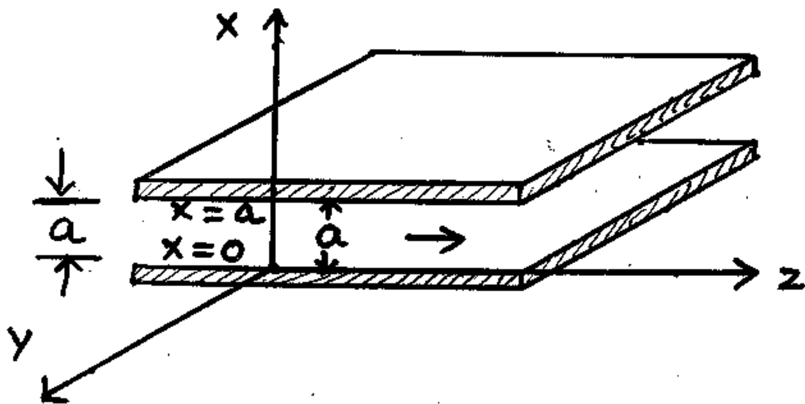
To find the field components of TE waves inside the parallel plates, we can assume a value for one of the field component and from this value, we can get the value of other field components.

Let us assume a sine wave is propagating in the z-direction. Let the value of H_z be,

$$H_z = \frac{h}{-j\omega\mu}\left[B_1 \cos hx - B_2 \sin hx\right] \rightarrow (17)$$

where, $h = \frac{m\pi}{a}$

where, 'a' is the distance between 2 planes in x-axis and 'm' is the integer having the value $m = 0, 1, 2, 3 \ldots$



The Boundary conditions are.

$x = 0 \qquad H_z = 0$

$x = a \qquad H_z = 0$

$$\frac{\partial H_z}{\partial x} = \frac{h^2}{+j\omega\mu}\left[B_1 \sin hx + B_2 \cos hx\right] \rightarrow (18)$$

sub equ (18) in (14) and (15)

$$E_y = \frac{\gamma j\omega\mu}{h^2} \cdot \frac{h^2}{\gamma j\omega\mu}\left[B_1 \sin hx + B_2 \cos hx\right]$$

$$\therefore E_y = B_1 \sin hx + B_2 \cos hx \rightarrow (19)$$

$$H_x = \frac{-\gamma}{h^2} \cdot \frac{h^2}{j\omega\mu}\left[B_1 \sin hx + B_2 \cos hx\right]$$

$$\therefore H_x = \frac{-\gamma}{j\omega\mu}\left[B_1 \sin hx + B_2 \cos hx\right] \rightarrow (20)$$

At High Freq. $\gamma = j\beta$

$\therefore$ put $\gamma = j\beta$ and $h = \frac{m\pi}{a}$ in equs. (17), (19), (20)

$$E_y = B_1 \sin\left(\frac{m\pi}{a} \cdot x\right) + B_2 \cdot \cos\left(\frac{m\pi}{a} x\right) \rightarrow (21)$$

$$H_x = \frac{-j\beta}{j\omega\mu}\left[B_1 \sin\left(\frac{m\pi}{a} x\right) + B_2 \cos\left(\frac{m\pi}{a} x\right)\right]$$

$$H_x = \frac{-\beta}{\omega\mu}\left[B_1 \sin\left(\frac{m\pi}{a} x\right) + B_2 \cos\left(\frac{m\pi}{a} x\right)\right] \rightarrow (22)$$

$$H_z = \frac{m\pi}{-ja\omega\mu}\left[B_1 \cos\left(\frac{m\pi}{a} x\right) - B_2 \sin\left(\frac{m\pi}{a} x\right)\right] \rightarrow (23)$$

The field components of TE wave inside the $\parallel^{el}$ planes can be represented in terms of <u>time and propagation</u> variation.

$$E_y = \left[B_1 \sin\left(\frac{m\pi}{a}x\right) + B_2 \cos\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} . \sin\omega t \rightarrow (24)$$

$$H_x = \left[\frac{-\beta_m}{\omega\mu}\left(B_1 \sin\left(\frac{m\pi}{a}x\right) + B_2 \cos\left(\frac{m\pi}{a}x\right)\right)\right] e^{-j\beta_m z} . \sin\omega t \rightarrow (25)$$

$$H_z = \left[\frac{-m\pi}{ja\omega\mu}\left(B_1 \cos\left(\frac{m\pi}{a}x\right) - B_2 \sin\left(\frac{m\pi}{a}x\right)\right)\right] e^{-j\beta_m z} . \sin\omega t$$

$$= \frac{\not{j} m\pi}{a\omega\mu}\left[B_1 \cos\left(\frac{m\pi}{a}x\right) - B_2 \sin\left(\frac{m\pi}{a}x\right)\right] . e^{-j\beta_m z} . \frac{\cos\omega t}{\not{j}}$$

$$H_z = \frac{m\pi}{a\omega\mu}\left[B_1 \cos\left(\frac{m\pi}{a}x\right) - B_2 \sin\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} . \cos\omega t \rightarrow (26)$$

The field components of TE waves inside the Parallel plates can be summarized as follows.

$$E_x = 0$$

$$E_y = \left[B_1 \sin\left(\frac{m\pi}{a}x\right) + B_2 \cos\left(\frac{m\pi}{a}x\right)\right] . e^{-j\beta_m z} . \sin\omega t$$

$$E_z = 0$$

$$H_x = \frac{-\beta_m}{\omega\mu}\left[B_1 \sin\left(\frac{m\pi}{a}x\right) + B_2 \cos\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} . \sin\omega t$$

$$H_y = 0$$

$$H_z = \frac{m\pi}{a\omega\mu}\left[B_1 \cos\left(\frac{m\pi}{a}x\right) - B_2 \sin\left(\frac{m\pi}{a}x\right)\right] e^{-j\beta_m z} . \cos\omega t$$

Characteristics (or) Properties of TE, TM Waves :-

1. **Propagation Constant, γ_m :-**

W.K.T $h^2 = \gamma_m^2 + \omega^2 \mu \varepsilon \rightarrow (1)$

$$\gamma_m^2 = h^2 - \omega^2 \mu \varepsilon$$

$$\gamma_m^2 = \left(\frac{m\pi}{a}\right)^2 - \omega^2 \mu \varepsilon$$

$$\gamma_m = \sqrt{\left(\frac{m\pi}{a}\right)^2 - \omega^2 \mu \varepsilon} \rightarrow (2)$$

$\gamma_m \rightarrow$ parallel, $\gamma_{mn} \rightarrow$ Rectangular

$\gamma_{nm} \rightarrow$ circular, γ

$\gamma_{mnp} \rightarrow$ Rectangular cavity Resonator

$\gamma_{nmp} \rightarrow$ circular cavity Resonator

parallel, $h^2 = \left(\frac{m\pi}{a}\right)^2$

Rectangular, $h^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2$

Rectangular cavity Resonator, $h^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2$

2. **Cut-off frequency, f_c :-**

At $f = f_c$, $\gamma_m = 0$ and $\omega = \omega_c$

$\therefore$ equ. (2) becomes

$$0 = \sqrt{\left(\frac{m\pi}{a}\right)^2 - \omega_c^2 \mu \varepsilon}$$

$$0 = \left(\frac{m\pi}{a}\right)^2 - \omega_c^2 \mu \varepsilon$$

$$\left(\frac{m\pi}{a}\right)^2 = \omega_c^2 \mu \varepsilon \rightarrow (3)$$

$$\omega_c^2 = \frac{1}{\mu \varepsilon} \cdot \left(\frac{m\pi}{a}\right)^2$$

$$\omega_c = \frac{1}{\sqrt{\mu \varepsilon}} \sqrt{\left(\frac{m\pi}{a}\right)^2}$$

$$f_c = \frac{1}{2\pi \sqrt{\mu \varepsilon}} \sqrt{\left(\frac{m\pi}{a}\right)^2} \rightarrow (4)$$

for air medium,

$$\frac{1}{\sqrt{\mu\varepsilon}} = c$$

$$\therefore f_c = \frac{c}{2\pi} \cdot \frac{m\pi}{a}$$

$$f_c = \frac{mc}{2a} \rightarrow (5)$$

$\mu = \mu_0 \cdot \mu_r$

$\mu_0 = 4\pi \times 10^{-7}$

$\mu_r = 1$

$\varepsilon = \varepsilon_0 \cdot \varepsilon_r$

$\varepsilon_0 = 8.854 \times 10^{-12}$

$\varepsilon_r = 1$

3. Phase constant, β_m :-

At HF, $\alpha_m = 0$, $\gamma_m = j\beta_m$

From equ (2)

$$j\beta_m = \sqrt{\left(\frac{m\pi}{a}\right)^2 - \omega^2\mu\varepsilon}$$

$$\beta_m = \frac{1}{j}\sqrt{\left(\frac{m\pi}{a}\right)^2 - \omega^2\mu\varepsilon}$$

$$= \sqrt{\frac{1}{j^2}\left[\left(\frac{m\pi}{a}\right)^2 - \omega^2\mu\varepsilon\right]}$$

$$= \sqrt{-\left[\left(\frac{m\pi}{a}\right)^2 - \omega^2\mu\varepsilon\right]}$$

$$\beta_m = \sqrt{\omega^2\mu\varepsilon - \left(\frac{m\pi}{a}\right)^2} \rightarrow (6)$$

'β_m' in terms of 'f_c'

substitute equ. (3) in (6)

$$\beta_m = \sqrt{\omega^2 \mu \varepsilon - \omega_c^2 \mu \varepsilon}$$

$$= \omega\sqrt{\mu\varepsilon}\ \sqrt{1 - \frac{\omega_c^2 \mu\varepsilon}{\omega^2 \mu\varepsilon}}$$

$$= \omega\sqrt{\mu\varepsilon} \cdot \sqrt{1-\left(\frac{\omega_c}{\omega}\right)^2}$$

$$\beta_m = \omega\sqrt{\mu\varepsilon} \cdot \sqrt{1-\left(\frac{f_c}{f}\right)^2} \longrightarrow (7)$$

'β_m' in terms of 'λ_c'

$$\beta_m = \omega\sqrt{\mu\varepsilon}\ \sqrt{1-\left(\frac{\lambda}{\lambda_c}\right)^2} \longrightarrow (8)$$

$\because \lambda = \frac{c}{f}$

$\lambda_c = \frac{c}{f_c}$

4. Guide wavelength, 'λ_g' :-

$$\lambda_g = \frac{2\pi}{\beta_m} \longrightarrow (9)$$

5. Cut-off wavelength, 'λ_c' :-

$$\lambda_c = \frac{c}{f_c} \longrightarrow (10)$$

6. Phase Velocity, V_p :-

$$V_p = \frac{\omega}{\beta_m} \longrightarrow (11)$$

7. Group velocity, v_g :-

$$v_p \times v_g = c^2 \rightarrow (12)$$

$$v_g = \frac{c^2}{v_p} \rightarrow (13)$$

8. Angle of Incidence, θ :-

$$\theta = \cos^{-1}\left(f_c/f\right) \rightarrow (14)$$

9. Intrinsic Impedance, η :-

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \rightarrow (15)$$

10. Characteristic Impedance, Z :-

$$Z_{TE} = \frac{\eta}{\sqrt{1-\left(f_c/f\right)^2}} \rightarrow (16)$$

$$\text{(or)} \quad Z_{TE} = \frac{\omega\mu}{\beta_m} \rightarrow (17)$$

$$Z_{TM} = \eta \cdot \sqrt{1-\left(f_c/f\right)^2} \rightarrow (18)$$

$$\text{(or)} \quad Z_{TM} = \frac{\beta_m}{\omega\varepsilon} \rightarrow (19)$$

Evanescant Mode :-

When Operating frequency is lower than cut-off frequency, the propagation constant becomes real. (i.e., $\gamma_m = \alpha_m$). The waves will not propagate. The modes that does not propagate is called as Evanescant Mode.

Propagating Mode :-

When cut-off frequency is lower than Operating frequency, the propagation constant becomes imaginary. (i.e., $\gamma_m = j\beta_m$). The waves will propagate. The modes that can propagate is called as propagating mode.

Dominant Mode :

The lowest order mode is called as dominant mode. This is the mode where wave starts to propagate.

The dominant mode of TE wave for parallel waveguide is TE_1 and TM_1.

Transmission of TE Waves Inside Rectangular Waveguide

To find the field configuration or components of TE waves inside Rectangular Waveguide, consider 2 planes are placed at a distance 'a' in x-axis and 'b' in y-axis. Assume the wave is propagating in z-direction.

The maxwell's equation to be satisfied by the electric and magnetic field at the boundary are,

$$\nabla \times E = -j\omega\mu H \rightarrow (1)$$

$$\nabla \times H = j\omega\varepsilon E \rightarrow (2)$$

Equation (1) can be written as,

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -j\omega\mu H_x \rightarrow (3)$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \rightarrow (4)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \rightarrow (5)$$

similarly equation (2) can be written as,

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = j\omega\varepsilon E_x \rightarrow (6)$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = j\omega\varepsilon E_y \rightarrow (7)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\varepsilon E_z \rightarrow (8)$$

Manipulating equs. (3) → (8) the relation between field components inside the guide can be obtained as,

$$E_x = \frac{-\gamma}{h^2}\frac{\partial E_z}{\partial x} - \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial y} \rightarrow (9)$$

$$E_y = \frac{-\gamma}{h^2}\frac{\partial E_z}{\partial y} + \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x} \rightarrow (10)$$

$$H_x = \frac{-\gamma}{h^2}\frac{\partial H_z}{\partial x} + \frac{j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial y} \rightarrow (11)$$

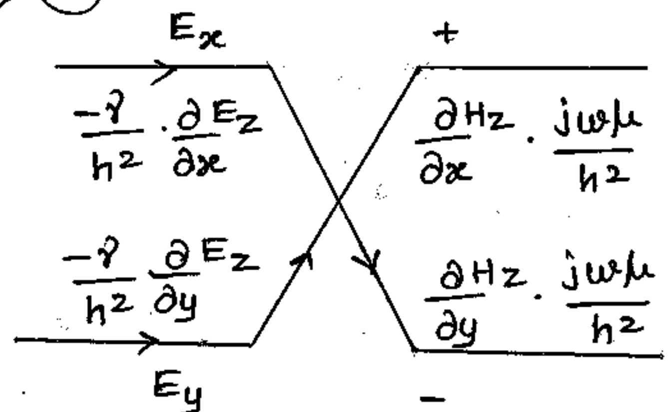
$$H_y = \frac{-\gamma}{h^2}\frac{\partial H_z}{\partial y} - \frac{j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial x} \rightarrow (12)$$

where, $h^2 = \gamma^2 + \omega^2\mu\varepsilon \rightarrow (13)$

For TE waves, $E_z = 0$

and for Rectangular waveguide $\gamma = \gamma_{mn}$

so, equs. (9) → (12) becomes



$$E_x = \frac{-j\omega\varepsilon}{h^2} . \frac{\partial H_z}{\partial y} \rightarrow (14)$$

$$E_y = \frac{j\omega\mu}{h^2} . \frac{\partial H_z}{\partial x} \rightarrow (15)$$

$$H_x = \frac{-\gamma_{mn}}{h^2} . \frac{\partial H_z}{\partial x} \rightarrow (16)$$

$$H_y = \frac{-\gamma_{mn}}{h^2} \frac{\partial H_z}{\partial y} \rightarrow (17)$$

From equs. (14) → (17), the existing field components inside Rectangular waveguide are,

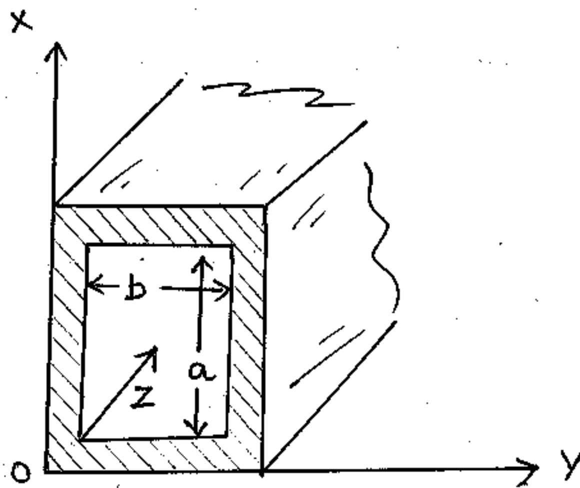
E_x, E_y, H_x, H_y and H_z

To find the field components of TE wave inside the Rectangular waveguide, we can assume a value for one of the field component and from this value, we can get the value of other field components.

Let us assume a sine wave is propagating in z-direction. Let the value of H_z component be

$$H_z = H_0 \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) \rightarrow (18)$$

where, 'a' and 'b' are dimensions of Rectangular waveguide and m, n are integers having the value $m, n = 0, 1, 2, 3, \ldots$



The Boundary conditions are

$$\left.\begin{aligned} \text{at } x = a, \quad & H_z = 0 \\ y = b, \quad & H_z = 0 \end{aligned}\right\} \rightarrow (19)$$

Differentiate equ. (18) w.r.to x and y

$$\frac{\partial H_z}{\partial x} = -H_0 \cdot \frac{m\pi}{a} \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) \rightarrow (20)$$

$$\frac{\partial H_z}{\partial y} = -H_0 \cdot \frac{n\pi}{b} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} \cdot y\right) \rightarrow (21)$$

sub equs. (20) & (21) in (14) → (17)

$$E_x = \frac{-j\omega\varepsilon}{h^2} \times -H_0 \cdot \frac{n\pi}{b} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} \cdot y\right)$$

$$E_x = \frac{j\omega\varepsilon}{h^2} \cdot H_0 \cdot \frac{n\pi}{b} \cos\left(\frac{m\pi}{a} x\right) \sin\left(\frac{n\pi}{b} y\right) \rightarrow (22)$$

$$E_y = \frac{-j\omega\mu}{h^2} \cdot H_0 \cdot \frac{m\pi}{a} \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) \rightarrow (23)$$

$$H_x = \frac{+\gamma_{mn}}{h^2} \cdot H_0 \cdot \frac{m\pi}{a} \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) \rightarrow (24)$$

$$H_y = \frac{\gamma_{mn}}{h^2} \cdot H_0 \cdot \frac{n\pi}{b} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) \rightarrow (25)$$

At high Frequency, $\gamma_{mn} = j\beta_{mn}$

∴ put $\gamma_{mn} = j\beta_{mn}$ and $h^2 = \omega_c^2 \mu\varepsilon$ in equs. (22) → (25)

$$E_x = \frac{j\omega\cancel{\varepsilon}}{\omega_c^2 \mu \cancel{\varepsilon}} \cdot H_0 \cdot \frac{n\pi}{b} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right)$$

$$E_x = \frac{j\omega}{\omega_c^2 \mu} \cdot H_0 \cdot \frac{n\pi}{b} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) \rightarrow (26)$$

$$E_y = \frac{-j\omega\cancel{\mu}}{\omega_c^2 \cancel{\mu}\varepsilon} \cdot H_0 \cdot \frac{m\pi}{a} \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right)$$

$$E_y = \frac{-j\omega}{\omega_c^2 \varepsilon} \cdot H_0 \cdot \frac{m\pi}{a} \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} \cdot y\right) \rightarrow (27)$$

$$H_x = \frac{j\beta_{mn}}{\omega_c^2 \mu \varepsilon} \cdot H_0 \cdot \frac{m\pi}{a} \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) \rightarrow (28)$$

$$H_y = \frac{j\beta_{mn}}{\omega_c^2 \mu \varepsilon} \cdot H_0 \cdot \frac{n\pi}{b} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) \rightarrow (29)$$

The field components of TE wave inside the rectangular waveguide can be represented in terms of time and propagation variation. so, equs. (18), (26) → (29) becomes,

$$E_x = \frac{\cancel{j}\omega}{\omega_c^2 \mu} \cdot H_0 \cdot \frac{n\pi}{b} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) \cdot e^{-j\beta_{mn} z} \cdot \frac{\cos\omega t}{\cancel{j}}$$

$$E_x = \frac{\omega}{\omega_c^2 \mu} \cdot H_0 \cdot \frac{n\pi}{b} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) \cdot e^{-j\beta_{mn} z} \cdot \cos\omega t \rightarrow (30)$$

$$E_y = \frac{-\omega}{\omega_c^2 \varepsilon} \cdot H_0 \cdot \frac{m\pi}{a} \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) \cdot e^{-j\beta_{mn} z} \cdot \cos\omega t \rightarrow (31)$$

$$H_x = \frac{\beta_{mn}}{\omega_c^2 \mu \varepsilon} \cdot H_0 \cdot \frac{m\pi}{a} \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) \cdot e^{-j\beta_{mn} z} \cdot \cos\omega t \rightarrow (32)$$

$$H_y = \frac{\beta_{mn}}{\omega_c^2 \mu \varepsilon} \cdot H_0 \cdot \frac{n\pi}{b} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) e^{-j\beta_{mn} z} \cdot \cos\omega t \rightarrow (33)$$

$$H_z = H_0 \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) \cdot e^{-j\beta_{mn} z} \cdot \sin\omega t \rightarrow (34)$$

Transmission Of TM waves Inside Rectangular Waveguide

To find the field configuration or Components inside Rectangular Waveguide, consider 2 planes are placed at a distance 'a' in x-axis and 'b' in y-axis. Assume the wave is propagating in z-direction.

The maxwell's equation to be satisfied by the electric and magnetic field at the boundary are,

$$\nabla \times E = -j\omega\mu H \rightarrow (1)$$

$$\nabla \times H = j\omega\varepsilon E \rightarrow (2)$$

Equation (1) can be written as,

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -j\omega\mu H_x \rightarrow (3)$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \rightarrow (4)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \rightarrow (5)$$

Similarly, equation (2) can be written as,

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = j\omega\varepsilon E_x \rightarrow (6)$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = j\omega\varepsilon E_y \rightarrow (7)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\varepsilon E_z \rightarrow (8)$$

Manipulating equs. ③ → ⑧, the relation between field Components inside the guide can be obtained as,

$$E_x = \frac{-\gamma}{h^2}\frac{\partial E_z}{\partial x} - \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial y} \rightarrow (9)$$

$$E_y = \frac{-\gamma}{h^2}\frac{\partial E_z}{\partial y} + \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x} \rightarrow (10)$$

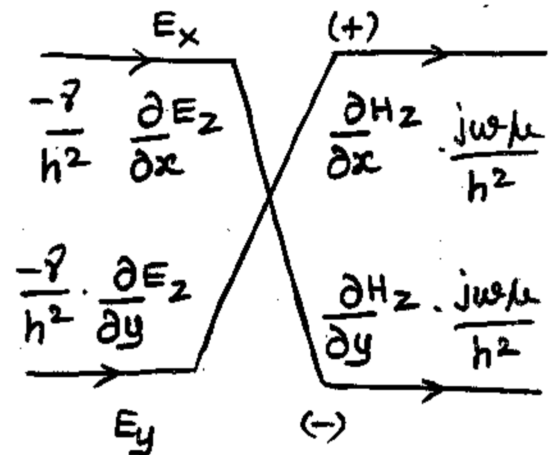
$$H_x = \frac{-\gamma}{h^2}\frac{\partial H_z}{\partial x} + \frac{j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial y} \rightarrow (11)$$

$$H_y = \frac{-\gamma}{h^2}\frac{\partial H_z}{\partial y} - \frac{j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial x} \rightarrow (12)$$

where, $h^2 = \gamma^2 + \omega^2\mu\cdot\varepsilon \rightarrow (13)$

For TM waves, $H_z = 0$

and for Rectangular Waveguide $\gamma = \gamma_{mn}$



so, equations ⑨ → ⑫ becomes

$$E_x = \frac{-\gamma_{mn}}{h^2}\cdot\frac{\partial E_z}{\partial x} \rightarrow (14)$$

$$E_y = \frac{-\gamma_{mn}}{h^2}\cdot\frac{\partial E_z}{\partial y} \rightarrow (15)$$

$$H_x = \frac{j\omega\varepsilon}{h^2}\cdot\frac{\partial E_z}{\partial y} \rightarrow (16)$$

$$H_y = \frac{-j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial x} \rightarrow (17)$$

From equs. (14) → (17) the existing field components inside Rectangular Waveguide are,

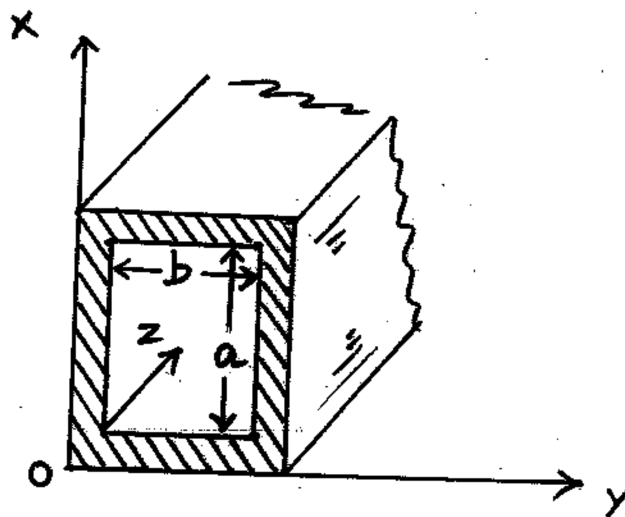
E_x, E_y, E_z, H_x and H_y

To find the field components of TM wave inside the Rectangular waveguide, we can assume a value for one of the field components and from this value, we can get the value of other field components.

Let us assume a sine wave is propagating in z-direction. Let the value of E_z component be

$$E_z = E_0 \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{b}y\right) \cdot \rightarrow (18)$$

where, 'a' and 'b' are dimensions of Rectangular waveguide and m, n are integers having the value $m, n = 0, 1, 2, 3, \ldots$



The Boundary conditions are,

$$\left.\begin{aligned} \text{at } x = a, \quad & E_z = 0 \\ y = b, \quad & E_z = 0 \end{aligned}\right\} \rightarrow (19)$$

Differentiate equ. 18 w.r.to x and y.

$$\frac{\partial E_z}{\partial x} = E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{b}y\right) \rightarrow (20)$$

(annotations written above the next equation: H_z above E_z, $-H_o$ above E_0, cos above sin, sin above cos)

$$\frac{\partial E_z}{\partial y} = E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b}y\right) \rightarrow (21)$$

sub equs. (20) & (21) in (14) → (17)

$$E_x = -\frac{\gamma_{mn}}{h^2} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{b}y\right) \rightarrow (22)$$

$$E_y = -\frac{\gamma_{mn}}{h^2} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b}y\right) \rightarrow (23)$$

$$H_x = \frac{j\omega\varepsilon}{h^2} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b}y\right) \rightarrow (24)$$

$$H_y = -\frac{j\omega\varepsilon}{h^2} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{b}y\right) \rightarrow (25)$$

At High freq. $\gamma_{mn} = j\beta_{mn}$

$\therefore$ Put $\gamma_{mn} = j\beta_{mn}$ and $h^2 = \omega_c^2 \mu\varepsilon$ in equs. (22) → (25)

$$E_x = \frac{-j\beta_{mn}}{\omega_c^2 \mu\varepsilon} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a}x\right) \sin\left(\frac{n\pi}{b}y\right) \rightarrow (26)$$

$$E_y = \frac{-j\beta_{mn}}{\omega_c^2 \mu\varepsilon} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b}y\right) \rightarrow (27)$$

$$H_x = \frac{j\omega\not{\varepsilon}}{\omega_c^2 \mu\not{\varepsilon}} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b}y\right)$$

$$H_x = \frac{j\omega}{\omega_c^2 \mu} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b}y\right) \rightarrow (28)$$

$$H_y = \frac{-j\omega\varepsilon}{\omega_c^2 \mu\varepsilon} \cdot E_0 \cdot \frac{m\pi}{a} \cos\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{b}y\right)$$

$$H_y = \frac{-j\omega}{\omega_c^2 \mu} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{b}y\right) \rightarrow (29)$$

The field components of TM wave inside the Rectangular waveguide can be represented in terms of time and propagation variation. so, equs. (26) → (29) becomes,

$$E_x = \frac{-j\beta_{mn}}{\omega_c^2 \mu\varepsilon} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{b}y\right) \cdot e^{-j\beta_{mn}z} \cdot \frac{\cos\omega t}{j}$$

$$E_x = \frac{-\beta_{mn}}{\omega_c^2 \mu\varepsilon} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{b}y\right) \cdot e^{-j\beta_{mn}z} \cdot \cos\omega t \rightarrow (30)$$

$$E_y = \frac{-j\beta_{mn}}{\omega_c^2 \mu\varepsilon} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b}y\right) \cdot e^{-j\beta_{mn}z} \cdot \frac{\cos\omega t}{j}$$

$$E_y = \frac{-\beta_{mn}}{\omega_c^2 \mu\varepsilon} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b} \cdot y\right) \cdot e^{-j\beta_{mn}z} \cdot \cos\omega t \rightarrow (31)$$

$$H_x = \frac{j\omega}{\omega_c^2 \mu} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b}y\right) \cdot e^{-j\beta_{mn}z} \cdot \frac{\cos\omega t}{j}$$

$$H_x = \frac{\omega}{\omega_c^2 \mu} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \cos\left(\frac{n\pi}{b} \cdot y\right) \cdot e^{-j\beta_{mn}z} \cdot \cos\omega t \rightarrow (32)$$

$$H_y = \frac{-\cancel{\gamma}\omega}{\omega_c^2 \mu} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) \cdot e^{-j\beta_{mn} z} \cdot \frac{\cos\omega t}{\cancel{\gamma}}$$

$$H_y = \frac{-\omega}{\omega_c^2 \mu} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} \cdot y\right) \cdot e^{-j\beta_{mn} z} \cdot \cos\omega t \quad \rightarrow (33)$$

The field components of TM waves inside the Rectangular Waveguide is summarized as,

$$E_x = \frac{-\beta_{mn}}{\omega_c^2 \mu \varepsilon} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) \cdot e^{-j\beta_{mn} z} \cdot \cos\omega t$$

$$E_y = \frac{-\beta_{mn}}{\omega_c^2 \mu \varepsilon} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) e^{-j\beta_{mn} z} \cdot \cos\omega t$$

$$E_z = E_0 \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) e^{-j\beta_{mn} z} \cdot \sin\omega t$$

$$H_x = \frac{\omega}{\omega_c^2 \mu \varepsilon} \cdot E_0 \cdot \frac{n\pi}{b} \cdot \sin\left(\frac{m\pi}{a} x\right) \cdot \cos\left(\frac{n\pi}{b} y\right) e^{-j\beta_{mn} z} \cdot \cos\omega t$$

$$H_y = \frac{-\omega}{\omega_c^2 \mu \varepsilon} \cdot E_0 \cdot \frac{m\pi}{a} \cdot \cos\left(\frac{m\pi}{a} x\right) \cdot \sin\left(\frac{n\pi}{b} y\right) \cdot e^{-j\beta_{mn} z} \cdot \cos\omega t$$

$$H_z = 0$$

characteristics (or) properties of TE, TM waves of Rectangular Waveguide :

1. Propagation constant, γ_{mn} :-

W.K.T $h^2 = \gamma_{mn}^2 + \omega^2\mu\varepsilon \longrightarrow (1)$

$$\gamma_{mn}^2 = h^2 - \omega^2\mu\varepsilon$$

$$\gamma_{mn}^2 = \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] - \omega^2\mu\varepsilon$$

$$\gamma_{mn} = \sqrt{\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] - \omega^2\mu\varepsilon} \longrightarrow (2)$$

2. cut-off frequency, f_c :-

At $f = f_c$, $\gamma_{mn} = 0$ and $\omega = \omega_c$

$\therefore$ equ (2) becomes

$$0 = \sqrt{\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] - \omega_c^2\mu\varepsilon}$$

Squaring on both sides

$$0 = \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] - \omega_c^2\mu\varepsilon$$

$$\omega_c^2\mu\varepsilon = \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] \longrightarrow (3)$$

$$\omega_c^2 = \frac{1}{\mu\varepsilon}\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]$$

$$\omega_c = \frac{1}{\sqrt{\mu\varepsilon}}\sqrt{\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]}$$

$$\therefore f_c = \frac{1}{2\pi\sqrt{\mu\varepsilon}}\sqrt{\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]} \longrightarrow (4)$$

For free space, $\frac{1}{\sqrt{\mu\varepsilon}} = c$

So, equ ④ becomes,

$$f_c = \frac{c}{2\pi}\sqrt{\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]} \rightarrow ⑤$$

3. Phase constant, β_{mn} :-

At HF, $\alpha_{mn} = 0$; $\gamma_{mn} = j\beta_{mn}$

So, equ. ② becomes

$$j\beta_{mn} = \sqrt{\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] - \omega^2\mu\varepsilon}$$

$$\beta_{mn} = \frac{1}{j}\sqrt{\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] - \omega^2\mu\varepsilon}$$

$$= \sqrt{\frac{1}{j^2}\left[\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] - \omega^2\mu\varepsilon\right]}$$

$$= \sqrt{-\left[\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] - \omega^2\mu\varepsilon\right]}$$

$$\beta_{mn} = \sqrt{\omega^2\mu\varepsilon - \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]} \rightarrow ⑥$$

β_{mn} in terms of cut-off frequency

from equ. ③, $\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] = \omega_c^2\mu\varepsilon$

∴ equ ⑥ becomes,

$$\beta_{mn} = \sqrt{\omega^2\mu\varepsilon - \omega_c^2\mu\varepsilon}$$

$$= \omega\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{\omega_c^2\mu\varepsilon}{\omega^2\mu\varepsilon}\right)}$$

$$\beta_{mn} = \omega\sqrt{\mu\varepsilon}\cdot\sqrt{1-\left(\frac{f_c}{f}\right)^2} \longrightarrow (7)$$

β_{mn} in terms of λ_c

$$\beta_{mn} = \omega\sqrt{\mu\varepsilon}\cdot\sqrt{1-\left(\lambda/\lambda_c\right)^2} \rightarrow (8)$$

4. Guide Wavelength, λ_g :-

$$\lambda_g = \frac{2\pi}{\beta_{mn}} \rightarrow (9)$$

5. cut - off Wavelength, λ_c :-

$$\lambda_c = \frac{c}{f_c} \longrightarrow (10)$$

6. Phase velocity, V_p :-

$$V_p = \frac{\omega}{\beta_{mn}} \longrightarrow (11)$$

7. Group velocity, V_g :-

$$V_p \times V_g = c^2 \longrightarrow (12)$$

$$V_g = \frac{c^2}{V_p} \longrightarrow (13)$$

8. Angle of Incidence, θ :-

$$\theta = \cos^{-1}(f_c/f) \rightarrow (14)$$

9. Intrinsic Impedance, η :-

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \rightarrow (15)$$

For air medium,

$$\eta = 377\,\Omega$$

10. characteristic Impedance, :-

$$Z_{TE} = \frac{\eta}{\sqrt{1-(f_c/f)^2}} \rightarrow (16)$$

(or) $$Z_{TE} = \frac{\omega\mu}{\beta_{mn}} \rightarrow (17)$$

$$Z_{TM} = \eta \cdot \sqrt{1-(f_c/f)^2} \rightarrow (18)$$

(or)

$$Z_{TM} = \frac{\beta_{mn}}{\omega\varepsilon} \rightarrow (19).$$

Impossibility of TEM wave in Waveguide :

Transverse Magnetic (TM) waves and Transverse Electric (TE) waves can propagate through the rectangular waveguide. For TM waves, no component of the magnetic field exists in z-direction and TE waves has no component of the Electric field in z-direction.

Consider that TEM wave exists within a hollow guide of any shape. By the property, the lines of magnetic field intensity H must lie entirely in the transverse plane. For a non-magnetic material with condition $\nabla \cdot H = 0$, the lines of H must lie in closed loops so, to have existence of the TEM waves inside the guide this H lines must be in a plane transverse to the axis of the guide.

According to Maxwell's first equation, the magnetomotive force (mmf) around each closed loop must be equal to the axial current. In a guide consisting inner conductor, the axial current is nothing but the conduction current in the inner inductor. But in a hollow waveguide like rectangular waveguide

there is no inner conductor present. In this case the axial current must be equal to the displacement current. By the property, the displacement current needs the component of the electric field E in axial direction. But such axial component of E is not present in TEM waves, hence it cannot exist in rectangular or circular waveguide.

Transmission Of TM waves inside Circular Waveguide

To find field configuration or Components inside Circular waveguide, Consider a circular waveguide with inner radius 'r' and assume the wave is propagating in z-direction.

The maxwell's equation to be satisfied by the electric and magnetic field at the boundary are,

$$\nabla \times E = -j\omega\mu H \rightarrow (1)$$

$$\nabla \times H = j\omega\varepsilon E \rightarrow (2)$$

Equation (1) can be written as r ϕ z

$$\frac{1}{r}\cdot\frac{\partial E_z}{\partial \phi} - \frac{\partial E_\phi}{\partial z} = -j\omega\mu H_r \rightarrow (3)$$

$$\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} = -j\omega\mu H_\phi \rightarrow (4)$$

$$\frac{\partial E_\phi}{\partial r} - \frac{1}{r}\cdot\frac{\partial E_r}{\partial \phi} = -j\omega\mu H_z \rightarrow (5)$$

Equation (2) can be written as

$$\frac{1}{r}\cdot\frac{\partial H_z}{\partial \phi} - \frac{\partial H_\phi}{\partial z} = j\omega\mu E_r \rightarrow (6)$$

$$\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} = j\omega\mu E_\phi \rightarrow (7)$$

$$\frac{\partial H_\phi}{\partial r} - \frac{1}{r}\cdot\frac{\partial H_r}{\partial \phi} = j\omega\mu E_z \rightarrow (8)$$

Manipulating equs. (3) → (8), the relation between field components inside the guide can be

obtained as,

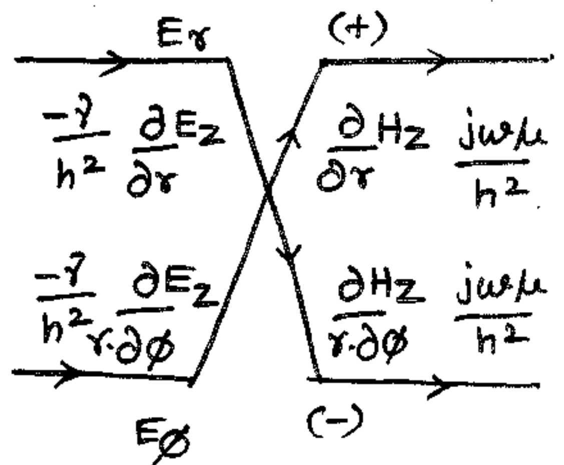
$$E_r = \frac{-\gamma}{h^2}\frac{\partial E_z}{\partial r} - \frac{j\omega\mu}{r.h^2}\frac{\partial H_z}{\partial \phi} \rightarrow (9)$$

$$E_\phi = \frac{-\gamma}{r.h^2}\frac{\partial E_z}{\partial \phi} + \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial r} \rightarrow (10)$$

$$H_r = \frac{-\gamma}{h^2}\frac{\partial H_z}{\partial r} + \frac{j\omega\varepsilon}{r.h^2}\frac{\partial E_z}{\partial \phi} \rightarrow (11)$$

$$H_\phi = \frac{-\gamma}{r.h^2}\frac{\partial H_z}{\partial \phi} - \frac{j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial r} \rightarrow (12)$$

where, $h^2 = \gamma^2 + \omega^2\mu\varepsilon \rightarrow (13)$



For TM waves, $H_z = 0$ and

For Circular WG. $\gamma = \gamma_{nm}$

∴ equs (9) → (12) becomes

$$E_r = -\frac{\gamma_{nm}}{h^2}\frac{\partial E_z}{\partial r} \rightarrow 14$$

$$E_\phi = \frac{-\gamma_{nm}}{r.h^2}\frac{\partial E_z}{\partial \phi} \rightarrow 15$$

$$H_r = \frac{j\omega\varepsilon}{r.h^2}\frac{\partial E_z}{\partial \phi} \rightarrow 16$$

$$H_\phi = \frac{-j\omega\varepsilon}{h^2}\frac{\partial E_z}{\partial r} \rightarrow 17$$

From equs. 14 → 17 the existing field components inside circular waveguide are E_r, E_ϕ, E_z, H_r and H_ϕ.

To find the field components of TM wave inside the circular waveguide, we can assume a value for one

one of the field components and from this value, we can get the value of other field components.

Let us assume a sine wave is propagating in z-direction. Let the value of E_z component be,

$$E_z = E_0 \cdot J_n(hr) \cos(n\phi) \rightarrow (18)$$

$$H_z = H_0 \cdot J_n(h'r) \cos(n\phi)$$

Differentiate equ. (18) w.r.to r and ϕ

$$\frac{\partial E_z}{\partial r} = E_0 \cdot \frac{\partial}{\partial r} J_n(hr) \cdot h \cdot \cos n\phi$$

$$\frac{\partial E_z}{\partial r} = E_0 \cdot h \cdot J_n'(hr) \cdot \cos n\phi \rightarrow (19)$$

$$\frac{\partial E_z}{\partial \phi} = -E_0 \cdot n \cdot J_n(hr) \sin n\phi \rightarrow (20)$$

sub equs. (19) & (20) in (14) → (17)

$$E_r = -\frac{\gamma_{nm}}{h^{\cancel{2}}} \cdot \cancel{h} \cdot E_0 \cdot J_n'(hr) \cdot \cos(n\phi)$$

$$E_r = -\frac{\gamma_{nm}}{h} \cdot E_0 \cdot J_n'(hr) \cdot \cos(n\phi) \rightarrow (21)$$

$$E_\phi = \frac{\gamma_{nm}}{r \cdot h^2} \cdot n \cdot E_0 \cdot J_n(hr) \cdot \sin(n\phi) \rightarrow (22)$$

$$H_r = \frac{-j\omega\varepsilon}{r \cdot h^2} \cdot n \cdot E_0 \cdot J_n(hr) \cdot \sin(n\phi) \rightarrow (23)$$

$$H_\phi = \frac{-j\omega\varepsilon}{h^{\cancel{2}}} \cdot E_0 \cdot \cancel{h} \cdot J_n'(hr) \cdot \cos(n\phi)$$

$$H_\phi = \frac{-j\omega\varepsilon}{h} \cdot E_0 \cdot J_n'(hr) \cdot \cos(n\phi) \rightarrow (24)$$

The field components of TM wave inside the Circular waveguide can be represented in terms of time and propagation variations. so, equs. 21 → 24 becomes

$$E_r = \frac{-\gamma_{nm}}{h} \cdot E_0 \cdot J_n'(hr) \cos(n\phi) \cdot e^{-\gamma_{nm}z} \cdot \sin\omega t \rightarrow (25)$$

$$E_\phi = \frac{\gamma_{nm}}{r \cdot h^2} \cdot n \cdot E_0 \cdot J_n(hr) \sin(n\phi) \cdot e^{-\gamma_{nm}z} \cdot \sin\omega t \rightarrow (26)$$

$$H_r = \frac{-\cancel{j}\omega\varepsilon}{r \cdot h^2} \cdot n \cdot E_0 \cdot J_n(hr) \cdot \sin(n\phi)\, e^{-\gamma_{nm}z} \cdot \frac{\cos\omega t}{\cancel{j}}$$

$$H_r = \frac{-\omega\varepsilon}{rh^2} \cdot n \cdot E_0 \cdot J_n(hr) \cdot \sin(n\phi)\, e^{-\gamma_{nm} \cdot z} \cdot \cos\omega t \rightarrow (27)$$

$$H_\phi = \frac{-\cancel{j}\omega\varepsilon}{h} \cdot E_0 \cdot J_n'(hr) \cdot \cos(n\phi) \cdot e^{-\gamma_{nm} \cdot z} \cdot \frac{\cos\omega t}{\cancel{j}}$$

$$H_\phi = \frac{-\omega\varepsilon}{h} \cdot E_0 \cdot J_n'(hr) \cdot \cos(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \cos\omega t \rightarrow (28)$$

The field components of TM waves inside the circular waveguide is summarized as,

$$E_r = -\frac{\gamma_{nm}}{h} \cdot E_0 \cdot J_n'(hr) \cdot \cos(n\phi)\, e^{-\gamma_{nm}z} \cdot \sin\omega t \rightarrow 20$$

$$E_\phi = \frac{\gamma_{nm}}{r \cdot h^2} \cdot n \cdot E_0 \cdot J_n(hr) \sin(n\phi)\, e^{-\gamma_{nm}z} \cdot \sin\omega t \rightarrow$$

$$E_z = E_0 \cdot J_n(hr) \cdot \cos(n\phi) \cdot e^{-\gamma_{nm} \cdot z} \cdot \sin\omega t \rightarrow$$

$$H_r = \frac{-\omega\varepsilon}{rh^2} \cdot n \cdot E_0 \cdot J_n(hr) \cdot \sin(n\phi) \cdot e^{-\gamma_{nm}z} \cdot \cos\omega t \rightarrow$$

$$H_\phi = \frac{-\omega\varepsilon}{h} \cdot E_0 \cdot J_n'(hr) \cdot \cos(n\phi) \cdot e^{-\gamma_{nm}z} \cdot \cos\omega t \rightarrow$$

$$H_z = 0$$

Transmission of TE Waves inside Circular Waveguide

To find field configuration or components inside circular waveguide consider a circular waveguide with inner radius 'r' and assume the wave is propagating in z-direction.

The maxwell's equation to be satisfied by the electric and magnetic field at the boundary are,

$$\nabla \times E = -j\omega\mu H \quad \rightarrow (1)$$

$$\nabla \times H = j\omega\varepsilon E \quad \rightarrow (2)$$

Equation (1) can be written as

$$\frac{1}{r}\frac{\partial E_z}{\partial \phi} - \frac{\partial E_\phi}{\partial z} = -j\omega\mu H_r \quad \rightarrow (3)$$

$$\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} = -j\omega\mu H_\phi \quad \rightarrow (4)$$

$$\frac{\partial E_\phi}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial \phi} = -j\omega\mu H_z \quad \rightarrow (5)$$

Equation (2) can be written as

$$\frac{1}{r}\frac{\partial H_z}{\partial \phi} - \frac{\partial H_\phi}{\partial z} = j\omega\mu E_r \quad \rightarrow (6)$$

$$\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} = j\omega\mu E_\phi \quad \rightarrow (7)$$

$$\frac{\partial H_\phi}{\partial r} - \frac{1}{r}\frac{\partial H_r}{\partial \phi} = j\omega\mu E_z \quad \rightarrow (8)$$

Manipulating equs. (3) → (8), the relation between field components inside the guide can obtained as,

$$E_r = -\frac{\gamma}{h'^2}\frac{\partial E_z}{\partial r} - \frac{j\omega\mu}{r.h'^2}\cdot\frac{\partial H_z}{\partial \phi} \rightarrow (9)$$

$$E_\phi = \frac{-\gamma}{r.h'^2}\frac{\partial E_z}{\partial \phi} + \frac{j\omega\mu}{h'^2}\cdot\frac{\partial H_z}{\partial r} \rightarrow (10)$$

$$H_r = \frac{-\gamma}{h'^2}\frac{\partial H_z}{\partial r} + \frac{j\omega\varepsilon}{r.h'^2}\frac{\partial E_z}{\partial \phi} \rightarrow (11)$$

$$H_\phi = \frac{-\gamma}{r.h'^2}\frac{\partial H_z}{\partial \phi} - \frac{j\omega\varepsilon}{h'^2}\cdot\frac{\partial E_z}{\partial r} \rightarrow (12)$$

where, $h'^2 = \gamma^2 + \omega^2\mu\varepsilon \rightarrow (13)$

For TE waves, $E_z = 0$ and for circular WG. $\gamma = \gamma_{nm}$

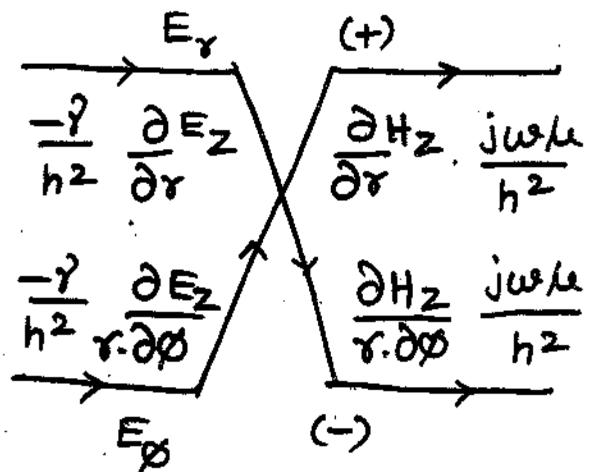
$\therefore$ equs. (9) $\rightarrow$ (12) becomes

$$E_r = \frac{-j\omega\mu}{r.h'^2}\cdot\frac{\partial H_z}{\partial \phi} \rightarrow 14$$

$$E_\phi = \frac{j\omega\mu}{h'^2}\cdot\frac{\partial H_z}{\partial r} \rightarrow 15$$

$$H_r = \frac{-\gamma}{h'^2}\frac{\partial H_z}{\partial r} \rightarrow 16$$

$$H_\phi = \frac{-\gamma}{r.h'^2}\frac{\partial H_z}{\partial \phi} \rightarrow 17$$



From equs. 14 → 17 the existing field components inside circular waveguide are,

E_r, E_ϕ, H_r, H_ϕ and H_z

To find the field components of TE wave inside the circular waveguide, we can assume a value for one of the field components and from this value, we can get

the value of other field components.

Let us assume a sine wave is propagating in z-direction. Let the value of H_z component be,

$$H_z = H_0 \cdot J_n(h'r) \cos(n\phi) \rightarrow (18)$$

Differentiate equ (18) w.r.to r and ϕ

$$\frac{\partial H_z}{\partial r} = H_0 \cdot \frac{\partial}{\partial r} J_n(h'r) \cdot h' \cdot \cos(n\phi)$$

$$\frac{\partial H_z}{\partial r} = H_0 \cdot h' \cdot J_n'(h'r) \cos(n\phi) \rightarrow (19)$$

$$\frac{\partial H_z}{\partial \phi} = -H_0 \cdot J_n(h'r) \sin(n\phi) \cdot n$$

$$\frac{\partial H_z}{\partial \phi} = -H_0 \cdot n \cdot J_n(h'r) \sin(n\phi) \rightarrow (20)$$

sub. equs. (19) & (20) in (14) → (17)

$$E_r = \frac{j\omega\mu}{r \cdot h'^2} \cdot H_0 \cdot n \cdot J_n(h'r) \sin(n\phi) \rightarrow (21)$$

$$E_\phi = \frac{j\omega\mu}{h'^2} \cdot H_0 \cdot h' \cdot J_n'(h'r) \cos(n\phi) \rightarrow (22)$$

$$H_r = \frac{-\gamma_{nm}}{h'^2} \cdot H_0 \cdot h' \cdot J_n'(h'r) \cos(n\phi)$$

$$H_r = \frac{-\gamma_{nm}}{h'} \cdot H_0 \cdot J_n(h'r) \cdot \cos(n\phi) \rightarrow (23)$$

$$H_\phi = \frac{\gamma_{nm}}{r \cdot h'^2} \cdot H_0 \cdot n \cdot J_n(h'r) \sin(n\phi) \rightarrow (24)$$

The field components of TE wave inside the circular waveguide can be represented interms of time and

propagation variations. so, equs. (21) → (24) becomes

$$E_r = \frac{j\omega\mu}{r \cdot h'^2} \cdot H_0 \cdot n \cdot J_n(h'r) \cdot \sin(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \frac{\cos\omega t}{j}$$

$$E_r = \frac{\omega\mu}{r \cdot h'^2} H_0 \cdot n \cdot J_n(h'r) \sin(n\phi)\, e^{-\gamma_{nm} z} \cdot \cos\omega t \rightarrow (25)$$

$$E_\phi = \frac{j\omega\mu}{h'} \cdot H_0 \cdot J_n'(h'r) \cdot \cos(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \frac{\cos\omega t}{j}$$

$$E_\phi = \frac{\omega\mu}{h'} \cdot H_0 \cdot J_n'(h'r) \cos(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \cos\omega t \rightarrow (26)$$

$$H_r = \frac{-\gamma_{nm}}{h'} \cdot H_0 \cdot J_n(h'r) \cdot \cos(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \sin\omega t \rightarrow (27)$$

$$H_\phi = \frac{\gamma_{nm}}{r \cdot h'^2} \cdot H_0 \cdot n \cdot J_n(h'r) \cdot \sin(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \sin\omega t \rightarrow (28)$$

The field components of TE waves inside the circular waveguide is summarized as,

$$E_r = \frac{\omega\mu}{r \cdot h'^2} \cdot H_0 \cdot n \cdot J_n(h'r) \cdot \sin(n\phi)\, e^{-\gamma_{nm} z} \cdot \cos\omega t$$

$$E_\phi = \frac{\omega\mu}{h'} \cdot H_0 \cdot J_n'(h'r) \cos(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \cos\omega t$$

$$E_z = 0$$

$$H_r = \frac{-\gamma_{nm}}{h'} \cdot H_0 \cdot J_n(h'r) \cdot \cos(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \sin\omega t$$

$$H_\phi = \frac{\gamma_{nm}}{r \cdot h'^2} \cdot H_0 \cdot n \cdot J_n(h'r) \cdot \sin(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \sin\omega t$$

$$H_z = H_0 \cdot J_n(h'r) \cdot \cos(n\phi) \cdot e^{-\gamma_{nm} z} \cdot \sin\omega t$$

Characteristics (or) Properties of TM waves of Circular Waveguide

1. Propagation constant, γ_{nm} :-

W.K.T $h^2 = \gamma_{nm}^2 + \omega^2\mu\varepsilon \rightarrow (1)$

$$\gamma_{nm}^2 = h^2 - \omega^2\mu\varepsilon$$

$$\gamma_{nm}^2 = \frac{(ha)^2}{a^2} - \omega^2\mu\varepsilon$$

$$\gamma_{nm} = \sqrt{\frac{(ha)^2}{a^2} - \omega^2\mu\varepsilon} \rightarrow (2)$$

2. Cut-off frequency, f_c :-

At $f = f_c$, $\gamma_{nm} = 0$ and $\omega = \omega_c$

$\therefore$ equ (2) becomes

$$0 = \sqrt{\frac{(ha)^2}{a^2} - \omega_c^2\mu\varepsilon}$$

Squaring on both sides

$$0 = \frac{(ha)^2}{a^2} - \omega_c^2\mu\varepsilon$$

$$\omega_c^2\mu\varepsilon = \frac{(ha)^2}{a^2} \rightarrow (3)$$

$$\omega_c^2 = \frac{1}{\mu\varepsilon}\cdot\frac{(ha)^2}{a^2}$$

$$\omega_c = \frac{1}{\sqrt{\mu\varepsilon}}\cdot\frac{(ha)}{a}$$

$$f_c = \frac{1}{2\pi\sqrt{\mu\varepsilon}} \cdot \frac{(ha)}{a} \rightarrow (4)$$

for free space,

$$f_c = \frac{c}{2\pi} \frac{(ha)}{a} \rightarrow (5)$$

3. Phase constant, β_{nm} :

At HF, $\alpha_{nm} = 0$, $\gamma_{nm} = j\beta_{nm}$

so, equ (2) becomes

$$j\beta_{nm} = \sqrt{\frac{(ha)^2}{a^2} - \omega^2\mu\varepsilon}$$

$$\beta_{nm} = \frac{1}{j}\sqrt{\left[\frac{(ha)^2}{a^2} - \omega^2\mu\varepsilon\right]}$$

$$= \sqrt{\frac{1}{j^2}\left[\frac{(ha)^2}{a^2} - \omega^2\mu\varepsilon\right]}$$

$$\beta_{nm} = \sqrt{\left[\omega^2\mu\varepsilon - \frac{(ha)^2}{a^2}\right]} \rightarrow (6)$$

β_{nm} in terms of cut-off frequency

from equ. (3) $\frac{(ha)^2}{a^2} = \omega_c^2\mu\varepsilon$

$$\therefore \beta_{nm} = \sqrt{\omega^2\mu\varepsilon - \omega_c^2\mu\varepsilon}$$

$$= \omega\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{\omega_c^2\mu\varepsilon}{\omega^2\mu\varepsilon}\right)}$$

$$\beta_{nm} = \omega\sqrt{\mu\varepsilon}\cdot\sqrt{1-(f_c/f)^2} \rightarrow (7)$$

β_{nm} in terms of λ_c.

$$\beta_{nm} = \omega\sqrt{\mu\varepsilon}\cdot\sqrt{1-(\lambda/\lambda_c)^2} \rightarrow (8)$$

4. Guide Wavelength, λ_g :-

$$\lambda_g = \frac{2\pi}{\beta_{nm}} \rightarrow (9)$$

5. cut-off Wavelength, λ_c :-

$$\lambda_c = \frac{c}{f_c} \rightarrow (10)$$

6. Phase Velocity, v_p :-

$$v_p = \frac{\omega}{\beta_{nm}} \rightarrow (11)$$

7. Group velocity, v_g :-

$$v_p \times v_g = c^2 \rightarrow (12)$$

$$v_g = \frac{c^2}{v_p} \rightarrow (13)$$

8. Intrinsic Impedance, η :-

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \rightarrow (14)$$

For air medium,

$$\eta = 377\,\Omega$$

9. Angle of Incidence, θ :-

$$\theta = \cos^{-1}\left(\frac{f_c}{f}\right) \rightarrow (15)$$

10. Characteristic Impedance, Z_{TM} :-

$$Z_{TM} = \frac{\beta_{nm}}{\omega\varepsilon} \rightarrow (16)$$

(or) $$Z_{TM} = \eta \cdot \sqrt{1-(f_c/f)^2} \rightarrow (17)$$

Order of the Bessel function n	$(ha)_{n1}$	$(ha)_{n2}$	$(ha)_{n3}$
0	✓ 2.405	5.52	8.65
1	3.83	7.06	10.17
2	5.13	8.41	11.62

Characteristics (or) Properties Of TE Waves Of Circular Waveguide

1. Propagation constant, γ_{nm} :

W.K.T $\quad h^2 = \gamma_{nm}^2 + \omega^2\mu\varepsilon \rightarrow (1)$

$$\gamma_{nm}^2 = h^2 - \omega^2\mu\varepsilon$$

$$\gamma_{nm}^2 = \frac{(h'a)^2}{a^2} - \omega^2\mu\varepsilon$$

$$\gamma_{nm} = \sqrt{\frac{(h'a)^2}{a^2} - \omega^2\mu\varepsilon} \rightarrow (2)$$

2. cut-off frequency, f_c :-

At $f = f_c$, $\gamma_{nm} = 0$ and $\omega = \omega_c$

$\therefore$ equ (2) becomes

$$0 = \sqrt{\frac{(h'a)^2}{a^2} - \omega_c^2\mu\varepsilon}$$

squaring on both sides

$$0 = \frac{(h'a)^2}{a^2} - \omega_c^2\mu\varepsilon$$

$$\omega_c^2\mu\varepsilon = \frac{(h'a)^2}{a^2} \rightarrow (3)$$

$$\omega_c^2 = \frac{1}{\mu\varepsilon} \cdot \frac{(h'a)^2}{a^2}$$

$$\omega_c = \frac{1}{\sqrt{\mu\varepsilon}} \cdot \frac{(h'a)}{a}$$

$$f_c = \frac{1}{2\pi\sqrt{\mu\varepsilon}} \cdot \frac{(h'a)}{a} \rightarrow (4)$$

for free space

$$f_c = \frac{c}{2\pi} \frac{(h'a)}{a} \rightarrow (5)$$

3. Phase constant, β_{nm} :

At HF, $\alpha_{nm} = 0$, $\gamma_{nm} = j\beta_{nm}$

so, equ. (2) becomes

$$j\beta_{nm} = \sqrt{\frac{(h'a)^2}{a^2} - \omega^2\mu\varepsilon}$$

$$\beta_{nm} = \frac{1}{j}\sqrt{\left[\frac{(h'a)^2}{a^2} - \omega^2\mu\varepsilon\right]}$$

$$= \sqrt{\frac{1}{j^2}\left[\frac{(h'a)^2}{a^2} - \omega^2\mu\varepsilon\right]}$$

$$\beta_{nm} = \sqrt{\left[\omega^2\mu\varepsilon - \frac{(h'a)^2}{a^2}\right]} \rightarrow (6)$$

β_{nm} in terms of cut-off frequency

from equ. (3) $\frac{(h'a)^2}{a^2} = \omega_c^2\mu\varepsilon$

$$\therefore \beta_{nm} = \sqrt{\omega^2\mu\varepsilon - \omega_c^2\mu\varepsilon}$$

$$= \omega\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{\omega_c^2 \cancel{\mu\varepsilon}}{\omega^2 \cancel{\mu\varepsilon}}\right)}$$

$$\beta_{nm} = \omega\sqrt{\mu\varepsilon}\sqrt{1-(f_c/f)^2} \rightarrow (7)$$

β_{nm} in terms of λ_c

$$\beta_{nm} = \omega\sqrt{\mu\varepsilon}\cdot\sqrt{1-(\lambda/\lambda_c)^2} \rightarrow (8)$$

4. Guide Wavelength, λ_g :-

$$\lambda_g = \frac{2\pi}{\beta_{nm}} \rightarrow (9)$$

5. cut-off Wavelength, λ_c :-

$$\lambda_c = \frac{c}{f_c} \rightarrow (10)$$

6. Phase velocity, v_p :-

$$v_p = \frac{\omega}{\beta_{nm}} \rightarrow (11)$$

7. Group velocity, v_g :-

$$v_g = \frac{c^2}{v_p} \rightarrow (12)$$

8. Intrinsic Impedance, 'η' :-

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \rightarrow (13)$$

For air medium,

$$\eta = 377\,\Omega$$

9. Angle of Incidence, 'θ' :-

$$\theta = \cos^{-1}(f_c/f) \rightarrow (14)$$

10. characteristic Impedance, Z_{TE} :-

$$Z_{TE} = \frac{\omega\mu}{\beta_{nm}} \rightarrow (15)$$

(or)

$$Z_{TE} = \frac{\eta}{\sqrt{1-(f_c/f)^2}} \rightarrow (16)$$

Order of the Bessel function n	$(h'a)_{n1}$	$(h'a)_{n2}$	$(h'a)_{n3}$
0	3.83	7.01	10.17
1	1.841 ✓	5.33	8.53
2	3.05	6.73	9.97

EXCITATION OF RECTANGULAR WAVEGUIDES :

* In Order to launch a particular mode, a type of probe is chosen which will produce lines of E and H that are roughly parallel to the lines of E and H for that mode.

* Generally, a guide is closed at one end by a conducting wall. An antenna probe is inserted through the end or side of the guide.

* The end of the waveguide closed by a conducting wall acts a reflector. By properly adjusting the distance between the probe and the end, we can make transmitted wave inphase with reflected wave. so, that both the waves will propagate as a single wave.

* The excitation methods of rectangular waveguide for various modes is as shown in fig. below.

* In Figure (a), the probe is connected in parallel to y-axis and produces lines of 'E' in y-direction. Lines of 'H' lie in x-z plane. This is the correct field configuration for TE_{10} mode.

* In Figure (b), two probes are connected in parallel to y-axis and produces lines of E in

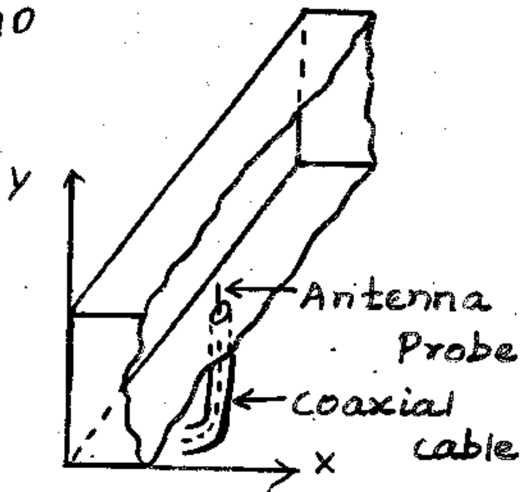
y-direction. The distance between the probes is $\frac{\lambda}{2}$. This is the correct field configuration for TE_{20} mode.

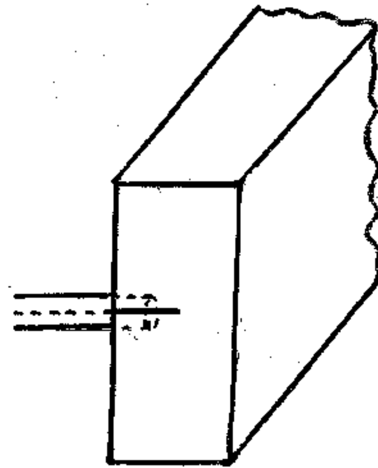
In Figure (c), the probe is connected at the terminated end. The probe is connected in parallel to z-axis and produces lines of E in z-direction. Lines of H lie in x-y plane. This is the correct field configuration for TM_{11} mode.

In Figure (d), two probes are connected in parallel to z-axis and produces lines of E in z-direction. The distance between the probes is $\lambda/2$. This is the correct field configuration for TM_{21} mode.

TE_{10}

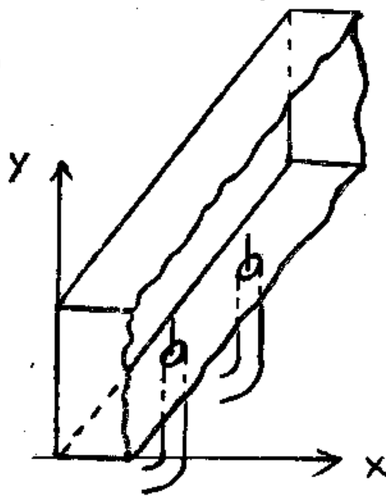
a)

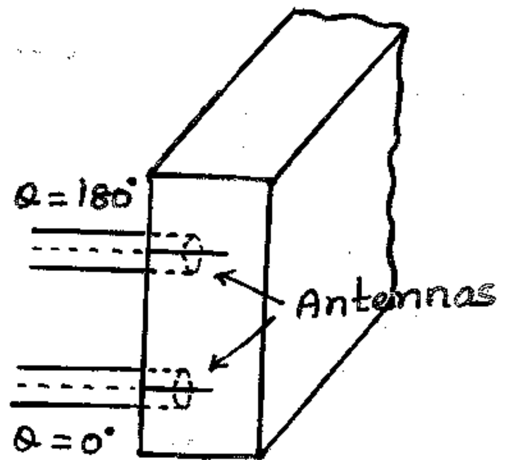




TE_{20}

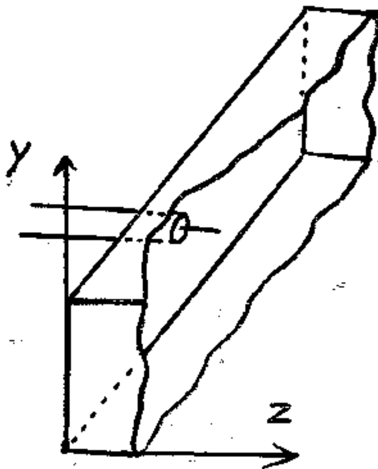
b)

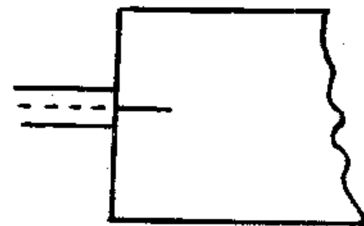




TM_{11}

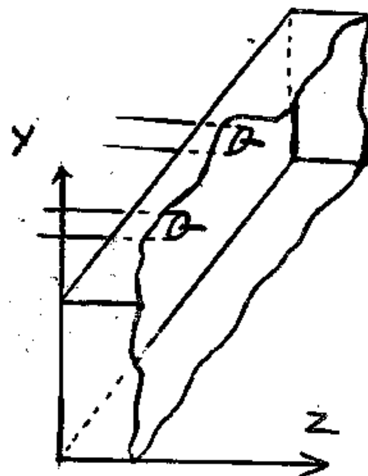
c)

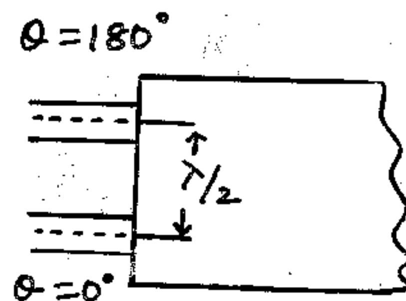




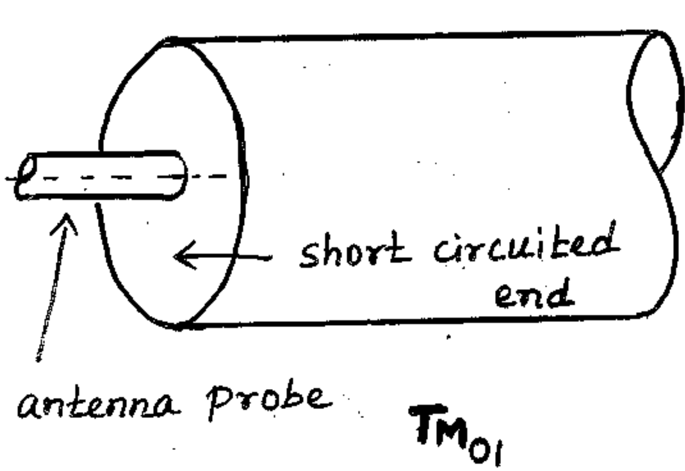
TM_{21}

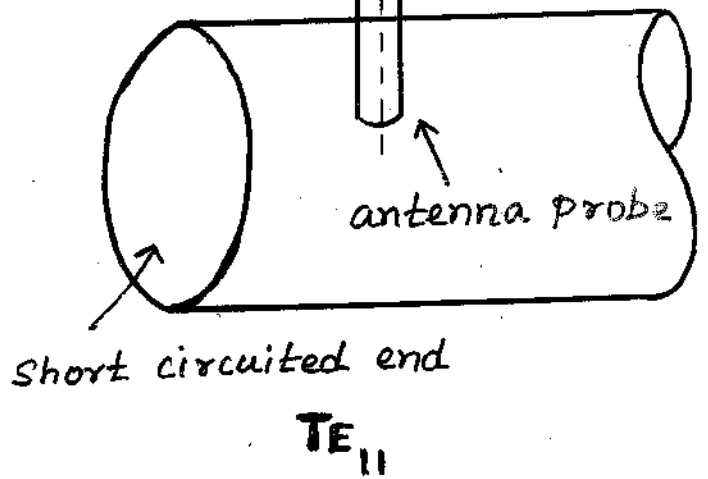
d)



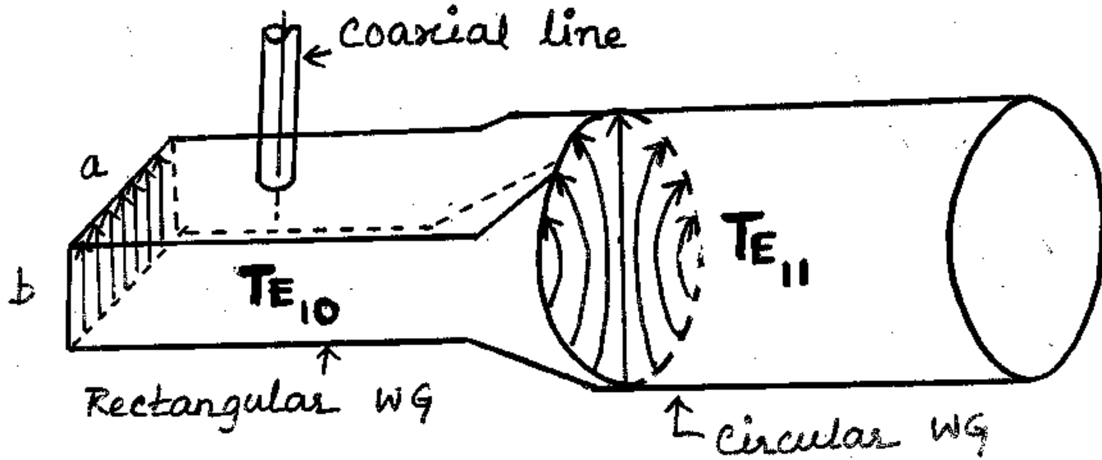


EXCITATION OF MODES IN CIRCULAR WAVEGUIDES

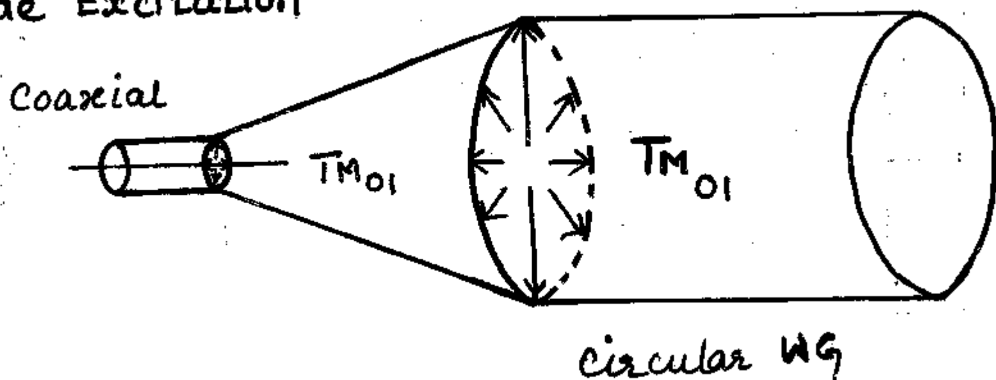




TE_{11} mode Excitation



TM_{01} Mode Excitation



EXCITATIONS OF MODES IN CIRCULAR WAVEGUIDE :

The methods of excitation for various modes in circular waveguide is as shown below.

* In Figure (a) coaxial line probe excite the dominant mode TE_{10} in a rectangular waveguide which is converted to dominant mode TE_{11} in the circular waveguide through the transition length between them.

* In Figure (b) longitudinal coaxial line probe directly excites the symmetric mode TM_{01} in a circular waveguide.

* In Figure (c), TE_{01} mode is excited by means of two diametrically opposite placed longitudinal narrow slots parallel to the wall of the rectangular waveguide.

EC 8651 - Transmission lines & RF systems

Unit V

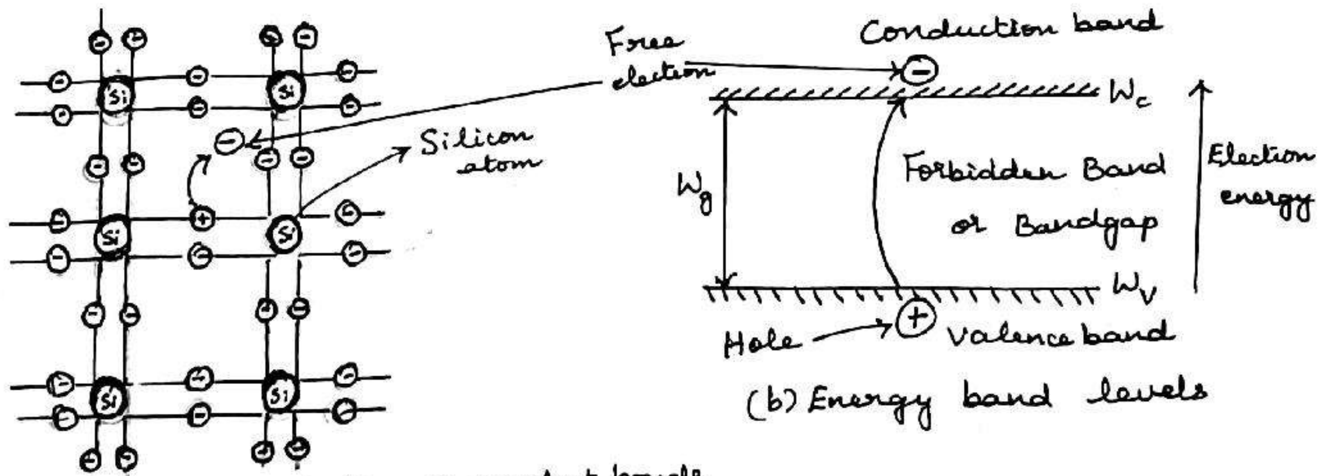
RF System Design concepts

Active RF components:

Semiconductor Physics:

The three most commonly used semiconductors are Germanium (Ge), Silicon (Si) and Gallium Arsenide (Ga As).

Bonding structure of Pure Silicon:



(a) Planar representation of covalent bonds

(b) Energy band levels

Each silicon atom shares its four valence elections with the four neighbouring atoms, forming four covalent bonds.

When temperature is equal to zero degree Kelvin, all elections are bonded to the atoms and therefore the Semiconductor is not conductive. But when temperature increases, some of

Scanned with CamScanner

the elections obtain sufficient energy to break the covalent bond (called bandgap energy). These negatively charged free elections causes current conduction. The concentration of conduction elections is denoted as 'n'. The positively charged vacancy created due to break of covalent bond are called holes and their concentration is denoted as 'p'.

At T > 0K, recombinations of elections & holes may occur. In thermal equilibrium, the number of recombinations & generations of holes & elections are equal.

<u>Effective carrier concentrations:</u>

According to Fermi statistics,

$$n = N_c \exp\left(-\frac{W_c - W_F}{kT}\right) \quad \&$$

$$p = N_v \exp\left(-\frac{W_F - W_v}{kT}\right)$$

where $N_{c,v} = 2\left(2 m_{n,p} \pi k T / h^2\right)^{3/2}$

Scanned with CamScanner

$N_c \rightarrow$ Conduction band

$N_v \rightarrow$ Valence band

$W_c \rightarrow$ Energy associated with conduction band

$W_v \rightarrow$ Energy associated with valence band

$W_F \rightarrow$ Fermi energy level

$m_{n,p} \rightarrow$ Effective mass of electrons / holes

$k \rightarrow$ Boltzmann's constant

$h \rightarrow$ Planck's constant

$T \rightarrow$ Absolute temperature in Kelvin.

Electrical conductivity of Intrinsic Semiconductor:

In an Intrinsic semiconductor, the number of free electrons is equal to the number of holes.

(ie. $n = p = n_i$).

According to concentration law,

$$np = n_i^2 \qquad \text{——— (1)}$$

Substituting the expression of n & p in the above equation, we get

$$n_i^2 = N_c \cdot \exp\left(-\frac{W_c - W_F}{kT}\right) \cdot N_v \exp\left(-\frac{W_F - W_v}{kT}\right)$$

$$= N_c \cdot N_v \exp\left(\frac{-W_c + W_F - W_F + W_v}{2kT}\right)$$

$$= N_c \cdot N_v \exp\left(-\frac{W_c - W_v}{2kT}\right)$$

or $n_i = \sqrt{N_c N_v} \exp\left(-\frac{W_c - W_v}{2kT}\right)$

Let $W_c - W_v = W_g$.

$$\therefore n_i = \sqrt{N_c N_v} \exp\left(-\frac{W_g}{2kT}\right) \quad \text{--- (2)}$$

The expression of electrical conductivity in a material due to applied electric field E is given as,

$$\sigma = qN\frac{V_d}{E} \quad \text{--- (3)}$$

where $V_d \rightarrow$ Drift velocity

$V_d \propto E$

$\Rightarrow V_d = \mu E$ [where μ is called carrier mobility]

$\therefore$ Equation (3) can be written as,

$$\sigma = qN\mu$$

or $\sigma = qn\mu_n + qp\mu_p$ [$\mu_n, \mu_p \rightarrow$ Mobilities of electrons & holes]

$$\sigma = qn_i(\mu_n + \mu_p) \quad [\because n = p = n_i] \quad \text{--- (4)}$$

Substituting equation (2) in (4), we get

$$\sigma = q\sqrt{N_c N_v} \exp\left(-\frac{W_g}{2kT}\right)(\mu_n + \mu_p)$$

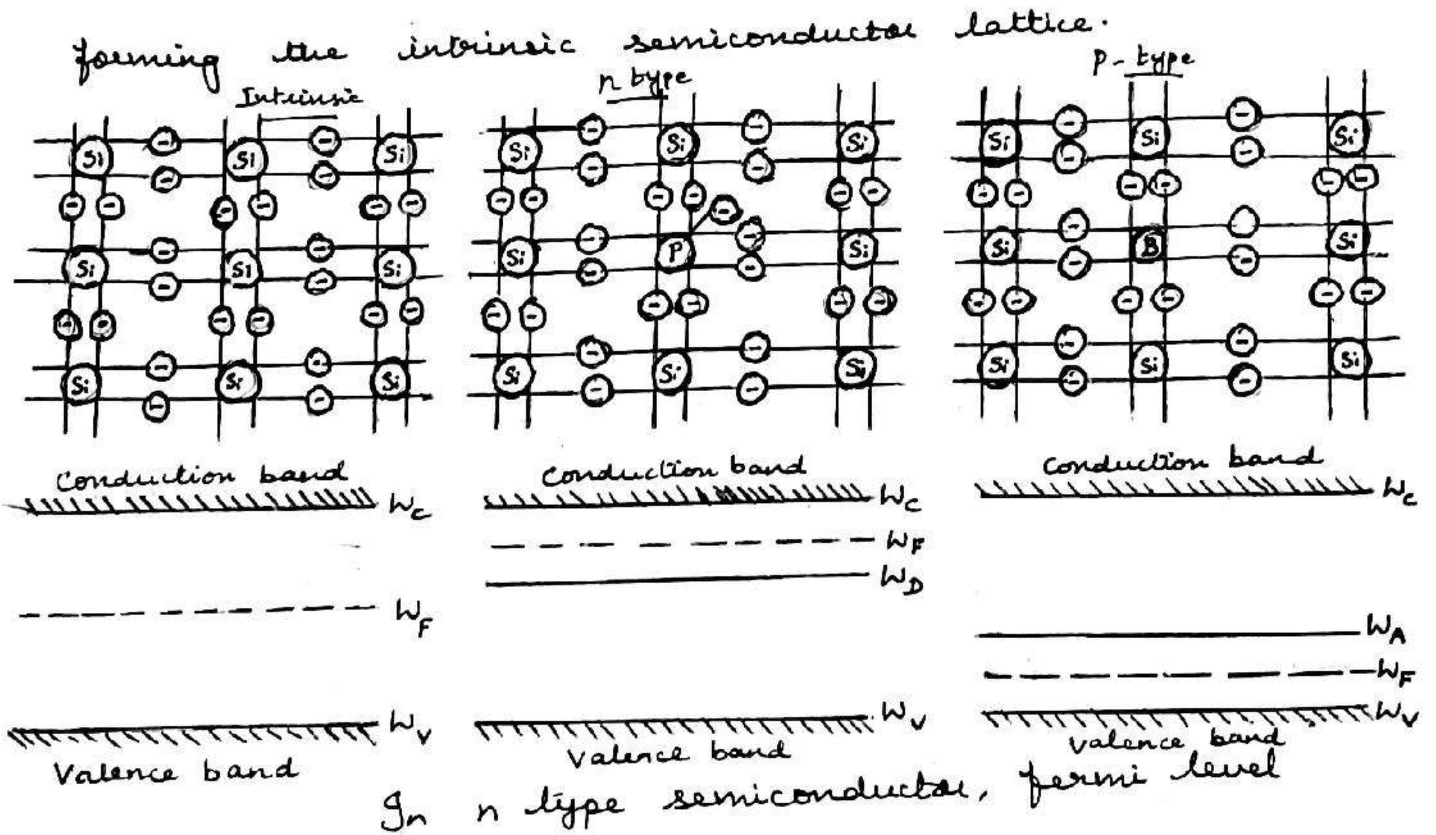
Scanned with CamScanner

Doping:

The process of introducing impurity atoms to change the electrical properties of a Semiconductor is called Doping.

n-type Semiconductor:

This type of semiconductor is formed by adding pentavalent impurities (such as Phosphorous) with large number of Valence electrons than the atoms forming the intrinsic semiconductor lattice.



In n type semiconductor, fermi level is increased as more electrons are located in the conduction band. The electron concentration is related to the hole concentration as

$$n_n = N_D + p_n$$

$N_D \rightarrow$ Donor concentration

Where,
$$n_n = \frac{N_D + \sqrt{N_D^2 + 4n_i^2}}{2}$$

&
$$p_n = \frac{-N_D + \sqrt{N_D^2 + 4n_i^2}}{2}$$

If $N_D >> n_i$

then $n_n \simeq N_D$

&
$$p_n = \frac{-N_D + N_D\sqrt{1 + \frac{4n_i^2}{N_D^2}}}{2} = \frac{-N_D + N_D\left(1 + \frac{2n_i^2}{N_D^2}\right)}{2}$$

$$\therefore p_n \simeq \frac{n_i^2}{N_D}$$

p - type Semiconductor:

This type of semiconductor is formed by adding trivalent impurities (such as Boron) with fewer valence electrons than the atoms forming the intrinsic semiconductor lattice. The hole concentration in the p type semiconductor is given as,

$$p_p = N_A + n_p$$

where $N_A \rightarrow$ Acceptor electron concentration

$n_p \rightarrow$ electron concentration in p type semi-conductor.

$$p_p = \frac{N_A + \sqrt{N_A^2 + 4n_i^2}}{2}$$

Scanned with CamScanner

$$\& \; n_p = \frac{-N_A + \sqrt{N_A^2 + 4n_i^2}}{2}$$

For high doping level, $N_A >> n_i$

$$\therefore \; p_p \simeq N_A$$

$$\& \quad n_p \simeq \frac{-N_A + N_A\left(1 + 2n_i^2/N_A^2\right)}{2} \simeq \frac{n_i^2}{N_A}$$

Where,

$N_A \rightarrow$ Acceptor concentration

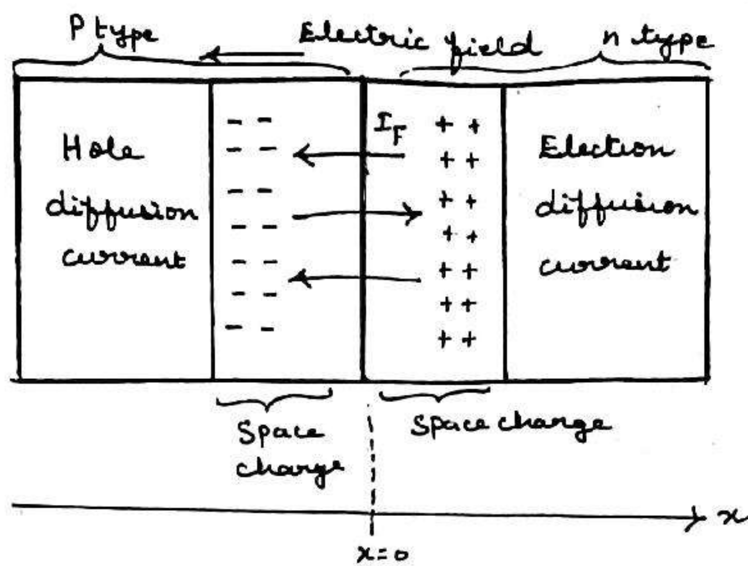
$n_i \rightarrow$ Intrinsic electron concentration

The pn Junction:

The physical contact of a p type & n type semiconductor is called pn junction.

The difference in the carrier concentration between the two types of semiconductor causes current flow across the interface and this current is called diffusion current. This current is composed of electrons & holes.

Scanned with CamScanner



The diffusion current is composed of $I_{n\,diff}$ & $I_{p\,diff}$ components:

$$I_{diff} = I_{n\,diff} + I_{p\,diff}$$

$$= qA\left(D_n \frac{dn}{dx} + D_p \frac{dp}{dx}\right)$$

where

$A \rightarrow$ Semiconductor cross sectional area

$D_n \rightarrow$ Diffusion constants for elections

$D_p \rightarrow$ Diffusion constant for holes

$$D_{n,p} = \mu_{n,p} \frac{kT}{q} = \mu_{n,p} V_T$$

$V_T \rightarrow$ Thermal potential

$$V_T = \frac{kT}{q}$$

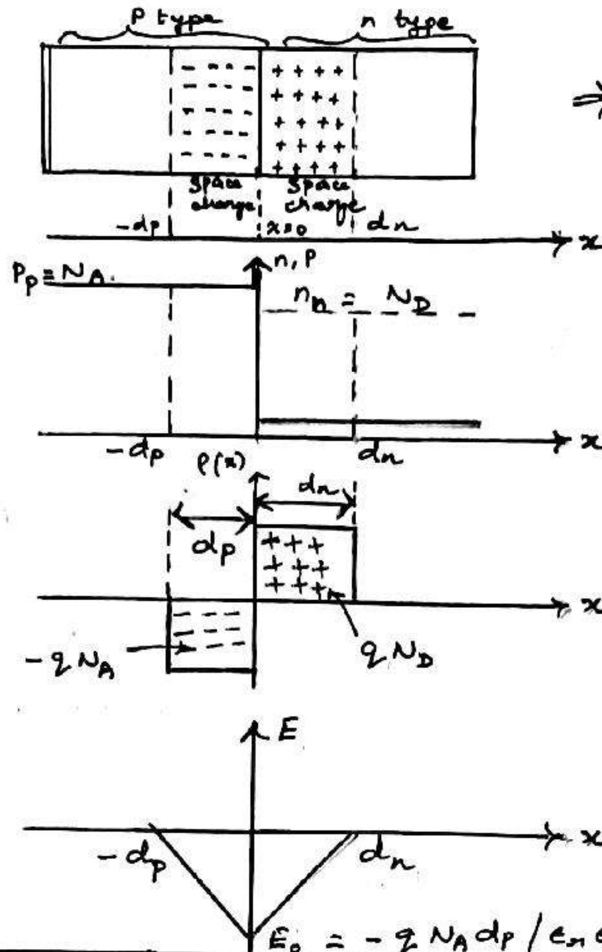
<u>Diffusion barrier voltage or built in potential:</u>

$$V_{diff} = V_T \ln\left(\frac{p_p}{p_n}\right) = V_T \ln\left(\frac{n_n}{n_p}\right)$$

If $N_D >> n_i$, then $n_n \simeq N_D$ & $n_p = \frac{n_i^2}{N_A}$

$$\therefore\ V_{diff} \simeq V_T \ln\left(\frac{N_A N_D}{n_i^2}\right)$$

Pn junction in the absence of external applied voltage:



⇒ Pn junction with space charge extent

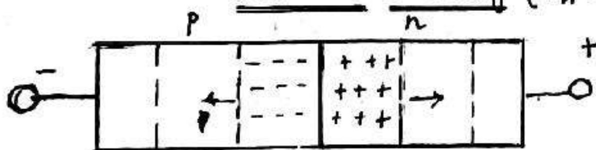
⇒ Acceptor & donor concentrations

⇒ Polarity of charge density distribution

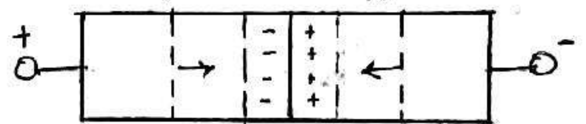
⇒ Electric field distribution

Pn junction in the presence of external applied voltage:

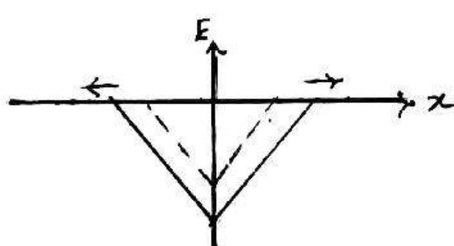
Reverse biasing ($V_A < 0$)

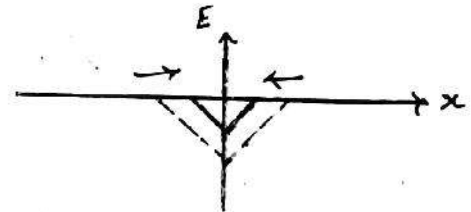


Forward biasing ($V_A > 0$)

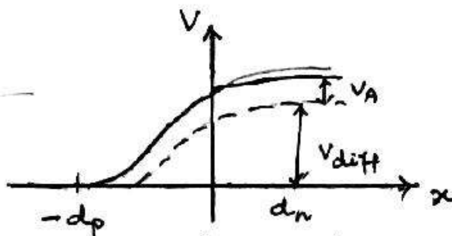


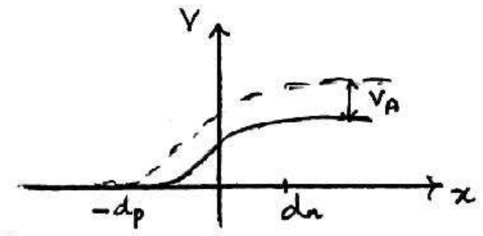
(a) Space charge distribution in pn junction





(b) Electric field distribution in pn junction





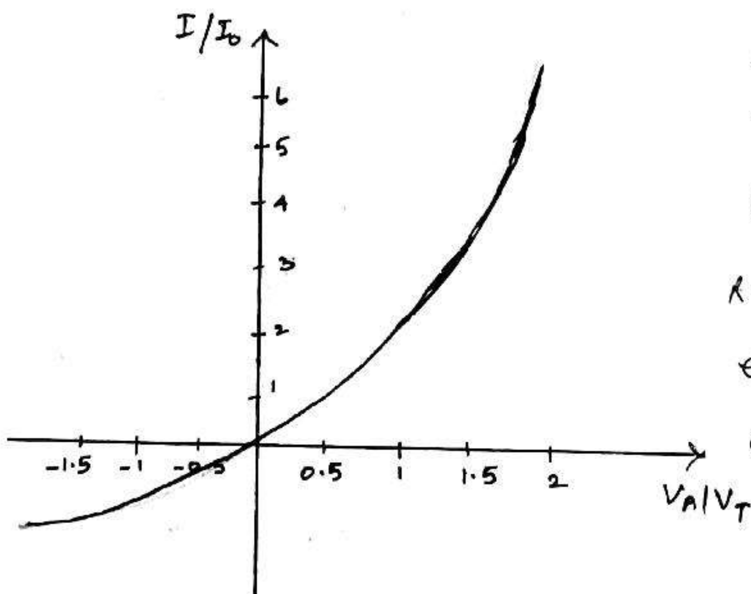
(c) Voltage distribution in pn junction

Forward bias	Reverse bias
1. Forward polarity decreases space charge domain	1. Reverse polarity increases the space charge domain
2. Increase in flow of current	2. Prevents the flow of current
3. Additional diffusion capacitance is encountered due to the presence of diffusion charges. $C_d = \frac{I_0 \tau_T}{V_T} e^{V_A/V_T}$	3. Leakage current occurs due to the movement of minority charge carriers. $I = I_0 \left(e^{V_A/V_T} - 1 \right)$ $I_0 \rightarrow$ Reverse saturation or leakage current.

Depletion layer capacitance:

$$C_J = C_{J0} \left(1 - \frac{V_A}{V_{diff}} \right)^{-1/2}$$

I-V characteristics:



* For -ve voltage, a small voltage independent current ($-I_0$) will flow.
* For +ve voltages, an exponentially increasing current is observed.

In general, the total capacitance C of a pn diode can be roughly divided into 3 regions.

(1) $V_A < 0$: only the depletion capacitance is significant : $C = C_J$

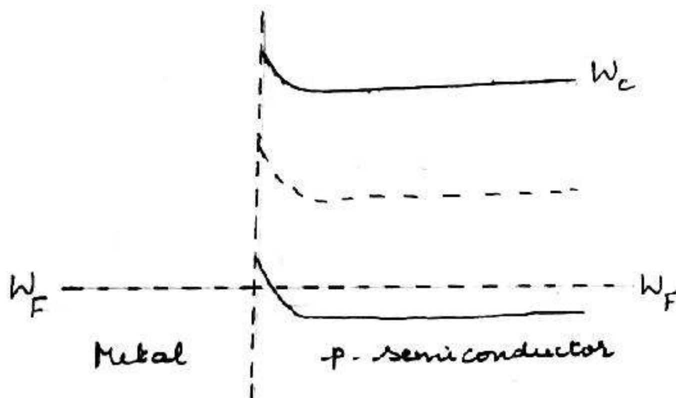
(2) $0 < V_A < V_{diff}$: depletion & diffusion capacitance combine : $C = C_J + C_d$

(3) $V_A > V_{diff}$: Only the diffusion capacitance is significant : $C = C_d$

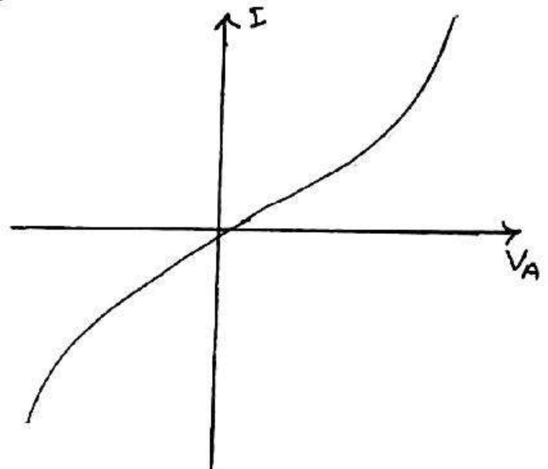
<u>Schottky</u> <u>contact</u>:

This refers to the contact between metallic electrode and semiconductor.

If a p semiconductor is in contact with a copper or alumunium electrode, there is a tendency for the electrons to diffuse into the metal, leading to increase in hole concentrations in the semi-conductor. This effect modifies the valence and conduction band energy levels near the interface.



Energy band model



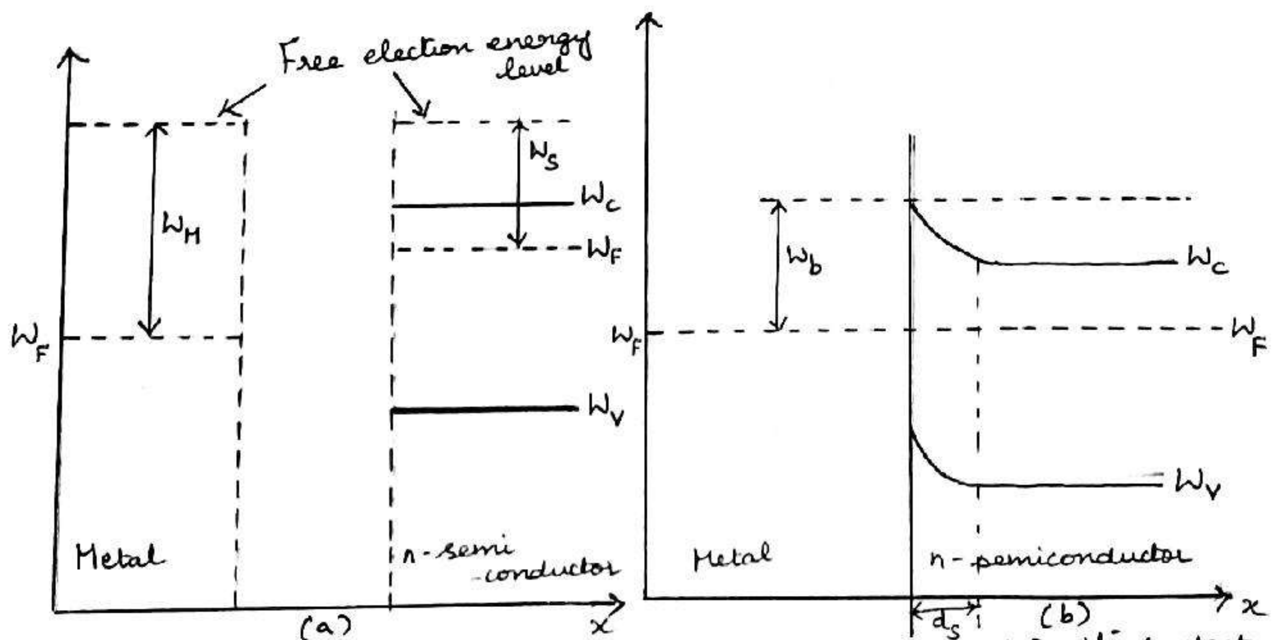
Voltage current characteristic

Metal electrode in contact with p-semiconductor

Scanned with CamScanner

Because of the higher concentration of holes, the valence band bends toward the fermi level. The conduction band bends away from the fermi level. ∴ Irrespective of the polarity of the applied voltage, a low resistance contact is obtained.

When n type semiconductor is in contact with metal, electrons diffuse from the n semiconductor & leave behind positive space charge. The depletion zone grows until the electrostatic repulsion of the space charges prevents further electron diffusion.



Energy band diagram of Schottky contact (a) before contact (b) after contact

The junction capacitance of the schottky contact is given as,

$$C_J = A\left[\frac{q\epsilon}{2(V_d - V_A)}N_D\right]^{1/2}$$

; $V_d \rightarrow$ Built in schottky barrier voltage

Bipolar Junction Transistor:

Bipolar Junction Transistor is a multi junction semiconductor device, where both the types of charge carriers take part in current carrying mechanism. two types of Bipolar Junction Transistors are n-p-n and p-n-p. The n-p-n Bipolar Junction Transistor is the complimentary structure of the p-n-p Bipolar Transistor.

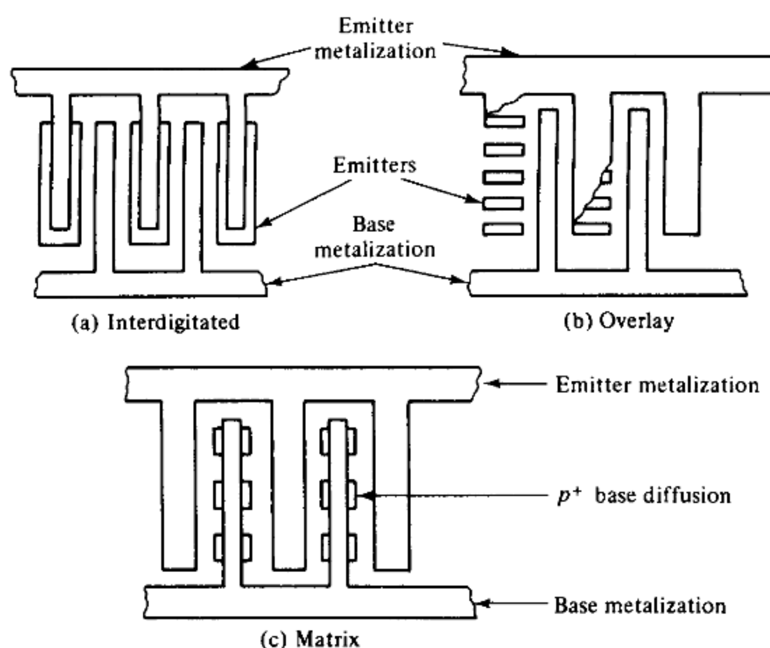
The principle of operation of microwave Bipolar Junction Transistor is similar to that of low frequency device bipolar transistor. All microwave Bipolar Transistor are planar in form and is of n-p-n type.

The majority of Bipolar Junction Transistors are fabricated from silicon because of low cost, more reliable integrative, offers higher gain and moderate noise figure when used as a microwave amplifier.

Microwave Bipolar Junction Transistors are capable of generating power upto a frequency of 22GHz.

Physical Structure:

The physical structure of microwave power transistor is as shown in figure below. the physical structure can be classified as a) inter-digitated b) Overlay c) Matrix type (also called as mesh or emitter grid)



Inter-digitated structure consists of large number of emitter strips alternating with base strips. Both of these are metallized. Overlay structure has a large number of segmented emitters overlaid through a number of wide metal strips. Matrix or mesh structure has emitter that forms the grid, the base filling the meshes of this grid with a p+ contact area in the middle of each mesh.

Inter-digitated structure is suitable for small signal applications in the L,S, and C bands whereas overlay and mesh structures are useful as power devices in the VHF and UHF regions.

Bipolar Transistor Configurations:

In general, there are two types of Bipolar transistors: n-p-n and p-n-p. A transistor can be connected as 3 different configurations: Common Base (CB), Common Emitter (CE) and Common Collector (CC).

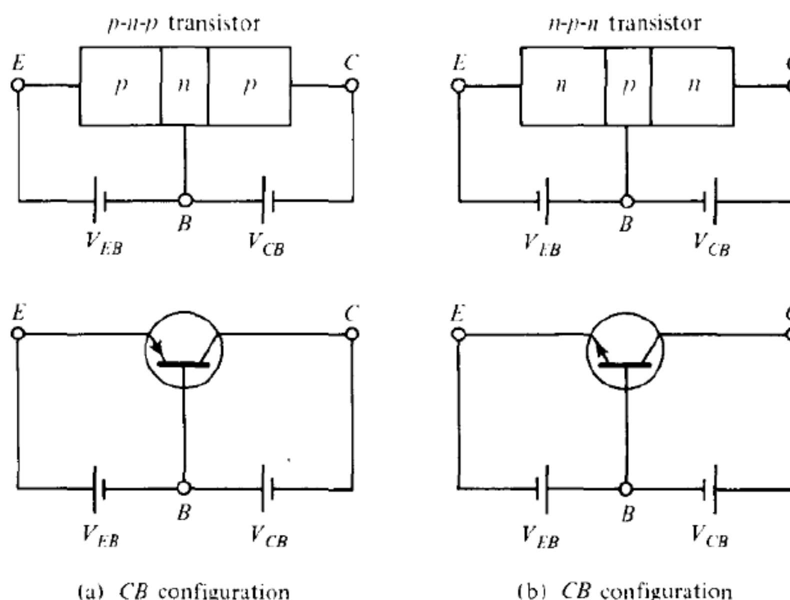
Common Base Configuration:

In common base configuration, the base terminal is common for both input circuit (Emitter) and output circuit(Collector). The common base configuration is also called as grounded base configuration.

common base configuration's input voltage V_{EB} and output current I_C can be expressed in terms of the output voltage V_{CB} and input current I_E as,

V_{EB} = some function (V_{CB}, I_E)

I_C = some function (V_{CB}, I_E)



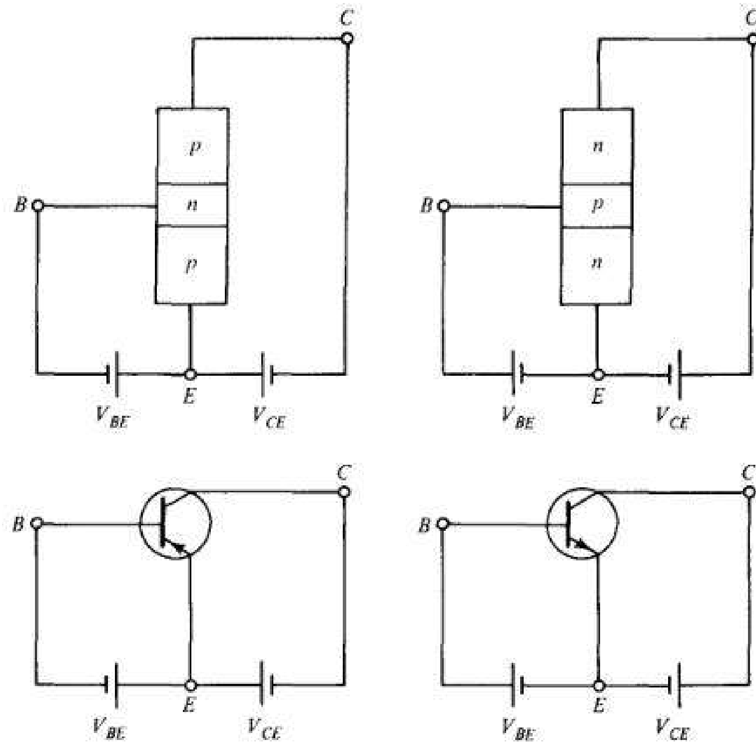
Common Emitter Configuration:

In common Emitter configuration, the emitter terminal is common for both input circuit (Base) and output circuit(Collector). The common emitter base configuration is also called as grounded emitter configuration.

common emitter configuration's input voltage V_{EB} and output current I_C can be expressed in terms of the output voltage V_{CB} and input current I_B as,

V_{EB} = some function (V_{CE}, I_B)

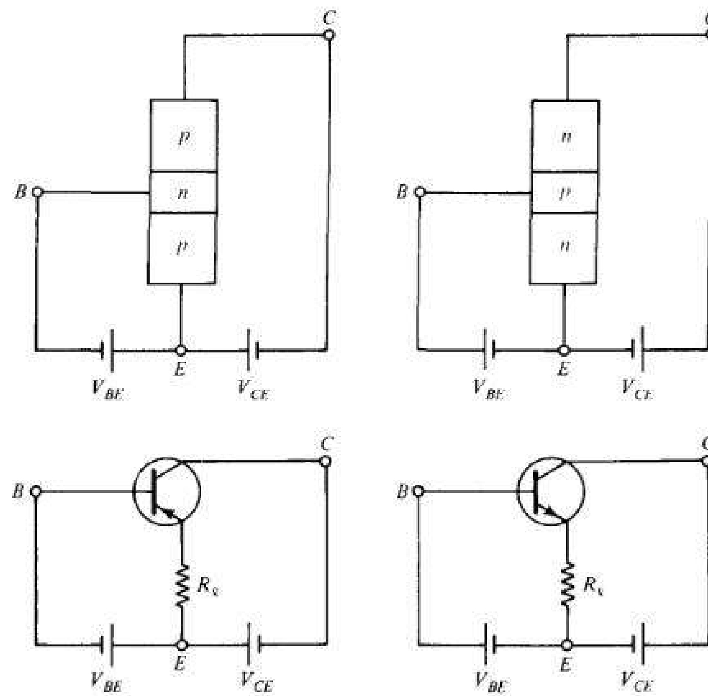
I_C = some function (V_{CE}, I_B)



Common Collector Configuration:

In common collector configuration, the collector terminal is common for both input circuit and output circuit. In a common collector configuration, the output voltage of the load is taken from the emitter terminal instead of the collector as in the common base and common emitter configuration.

The common collector configuration transistor can be used as a switch or pulse amplifier. The common collector amplifier has no voltage gain.



Principles of Operation:

The bipolar junction transistor is an active device which is commonly used as an amplifier or switch. A BJT can operate in four different modes depending on the voltage polarities across the two junctions.

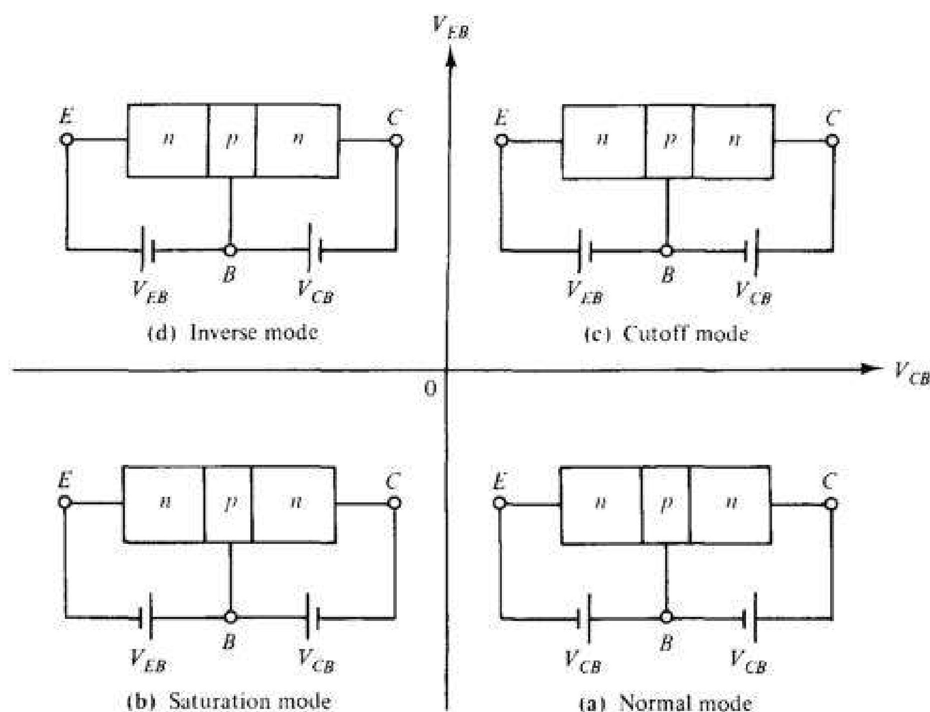
1. Normal Mode: In this mode, emitter junction of npn transistor is forward biased and collector junction is reverse biased . Generally at ON state a transistor remains in the normal mode.

2. Saturation Mode: When both the junctions are forward biased, the transistor is in its saturation mode with very low resistance and acts like a short circuit.

3. Cut-Off Mode: If both transistor junctions are reverse biased, the transistor is operated in cut-off mode and the transistor acts like an open circuit.

Thus saturation and cut-off modes are equivalent to the ON and OFF state of a switch.

4. Inverse Mode: A transistor is said to be in inverse mode when the emitter is reverse biased and collector is forward biased. In practice transistor is not commonly used in inverse mode.

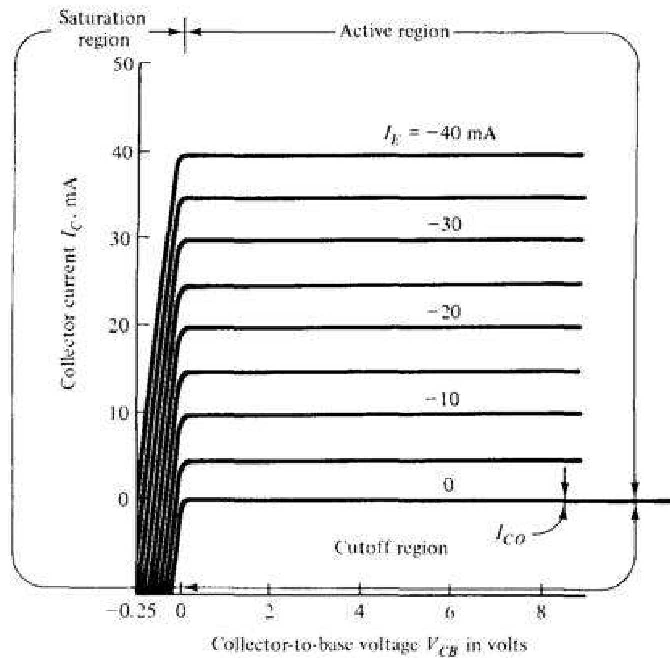


There are three regions for the 1-V characteristics of an *n-p-n* bipolar transistor:

1. Active Region: In this region the emitter junction is forward-biased and the collector junction is reverse-biased. The collector current I_C is essentially independent of collector voltage and depends only on the emitter current I_E. When the emitter current is zero, the collector current is equal to the reverse saturation current I_{CO}.

2. Saturation Region: In this region, as shown on the left side of figure, both emitter and collector junctions are forward-biased. The electron current flows from the *n* side across the collector junction to the p-type base. As a result, the collector current increases sharply.

3.Cutoff Region: In this region the emitter and collector junctions are both reverse-biased. Consequently, the emitter current is cut off to zero, as shown in the lower right side of figure.



Performance Parameter:

In high frequency operation, the performance of a microwave transistor depends on the cut-off frequency 'f_c' and maximum frequency of oscillation (f_{max}) rather than the two current gains α and β.

Now the cut-off frequency depends on the delay of the carrier results due to their movement from emitter to collector.

$$f_c = \frac{1}{2\pi\tau_{ec}} \qquad (1)$$

$$\tau_{ec} = \tau_e + \tau_b + \tau_d + \tau_c \qquad (2)$$

where, τ_e **-** Emitter base junction transit time

τ_c - Collector depletion layer charging time

τ_b - Base transit time

τ_d - Collector depletion layer transit time

Maximum frequency of operation is higher than f_c because although 'β' falls to unity at this frequency, power gain does not.

$$f_{max} = \sqrt{\frac{f_c}{8\pi r'_b C_C}} \qquad (3)$$

where, rb' - Base resistance

C_C - Collector Capacitance

RF FIELD EFFECT TRANSISTOR:

- ❖ Field effect transistor is a multi junction monopolar device, where only one carrier type either holes or electrons contribute to the current flow through the channel.
- ❖ Based on the contribution there are two types, n-Channel (electron) and p-channel (hole).
- ❖ FET is a voltage controlled device.
- ❖ RF field effect transistors has the capability of amplifying small signals up to the frequency range of X band with low noise figures.
- ❖ The RF field effect transistors has several advantages over the Bipolar junction transistor.
    1. Its Efficiency is higher than BJT.
    2. Its noise figure is low.
    3. Its operating frequency is up to X band
    4. Its input resistance is very high up to several mega ohms.

CONSTRUCTION:

FETs are classified according to how the gate is connected to the conducting channel. Specifically, there are four types. They are,

1. MISFET - Metal Insulator Semiconductor FET:

Here the gate is separated from the channel through an insulation layer. One of the most widely used type is MOSFET (Metal Oxide Semiconductor FET)

2. JFET - Junction FET:

This type relies on a revere biased pn-junction that isolates the gate from the channel.

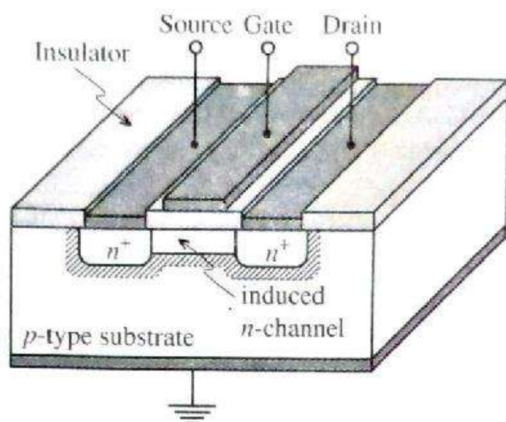
3.MESFET - Metal Semiconductor FET:

If the reverse biased pn-junction is replaced by a schottky contact, the channel can be controlled just as in the JFET case.

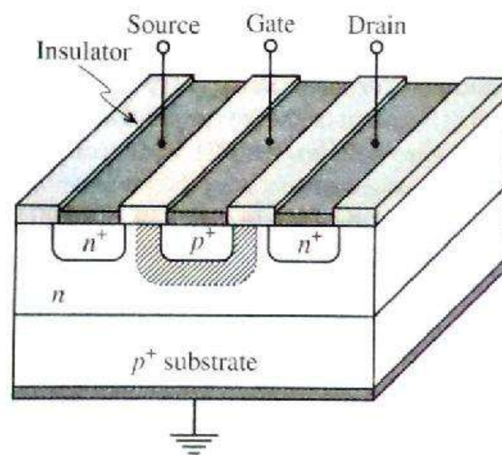
4. Hetro FET:

As the name implies, the transitions takes place between different layer of semiconductor materials. Examples: GaAlAs to GaAs or GaInAs to GaAlAs interfaces. High Electron Mobility Transistor(HEMT) belongs to this class.

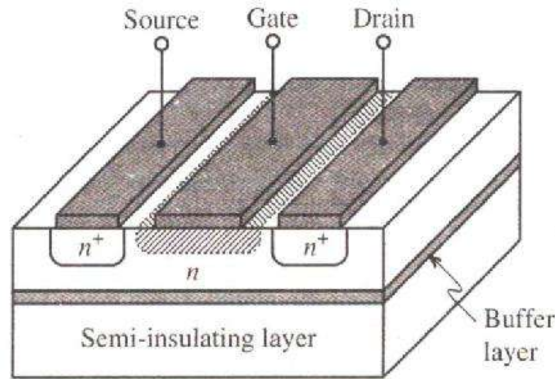
The construction of MISFET, JFET, and MESFET is as shown in the figure below.



(a) Metal insulator semiconductor FET (MISFET)



(b) Junction field effect transistor (JFET)



(c) Metal semiconductor FET (MESFET)

- In the above shown FETs, the current flows from source to drain and the gate controls the current flow. Due to the presence of a large capacitance formed by the gate electrode and the reverse biased pn-junction, MISFETs and JFETs have a relatively low cut-off frequency and are usually operated in low and medium frequency ranges of typically upto 1 GHz.
- GaAs MESFETs find applications upto 60-70 GHz and HEMT can operate beyond 100GHz.
- Electrically FETs can be classified into two types, 1) Enhancement and 2) Depletion type based on increase in carriers or depletion in carriers when the gate voltage is increased.

Functionality:

The functionality of MESFET for different drain-source voltages are shown in figure below. The transistor is operated in depletion mode. The schottky contact builds up channel space charge domain that affects the current flow from the source to drain. The space extent d_s can be controlled via the gate voltage.

$$d_s = \left(\frac{2\varepsilon V_d - V_{GS}}{qN_D} \right)^{\frac{1}{2}} \qquad (1)$$

where, d_s - Space extent or Space charge

N_D - Donor concentration

V_d - Barrier voltage 0.9v for GaAs-Au interface

q - Charge of an electron (1.602×10^{-19})

V_{GS} - Gate source voltage

The resistance 'R' between source and drain is predicted by,

$$R = \frac{L}{\sigma(d - d_s)W} \qquad (2)$$

where, W - Gate Width

L - Gate Length

σ - Conductivity

d - Channel depth

d_s - Space charge

$$\sigma = q.\mu_n.N_D \qquad (3)$$

Where, $.\mu_n$ - Mobility of electron

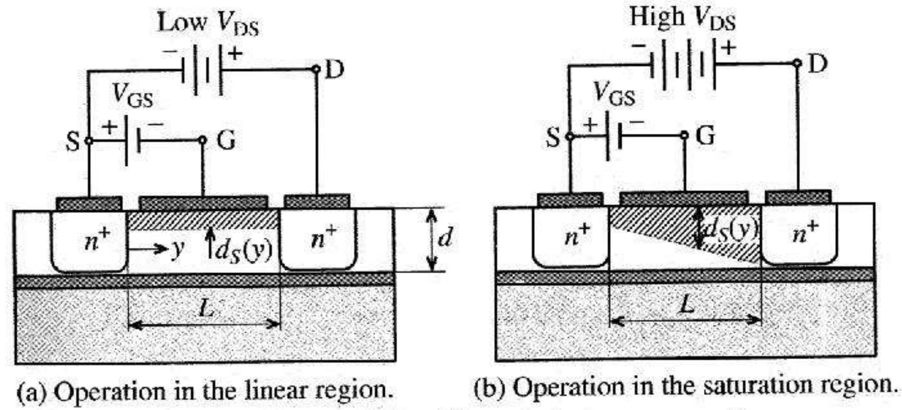
N_D - Donor Concentration

The drain current is given by,

$$I_D = \frac{V_{DS}}{R} = G_O \left[1 - \left(\frac{2\varepsilon}{qd^2} \frac{V_d - V_{GS}}{N_D} \right)^{\frac{1}{2}} \right] V_{DS} \quad (4)$$

Where, Conductance G_O is,

$$G_O = \frac{\sigma q N_D W d}{L} \quad (5)$$



(a) Operation in the linear region. (b) Operation in the saturation region.

Functionality of MESFET for different drain-source voltages.

The pinch-off voltage for the FET is independent of the gate-source voltage and is computed as,

$$V_P = \frac{qN_D d^2}{2\varepsilon} \quad (6)$$

where, V_P - pinch-off voltage

q - Charge of an electron (1.6×10^{-19})

ε - Permittivity

The threshold voltage for the FET is given as,

$$V_{TO} = V_d - V_P \quad (7)$$

The Drain saturation current is

$$I_{DSat} = G_O \left[\frac{V_P}{3} - (V_d - V_{GS}) + \frac{2}{3\sqrt{V_P}} (V_d - V_{GS})^{\frac{3}{2}} \right] \quad (8)$$

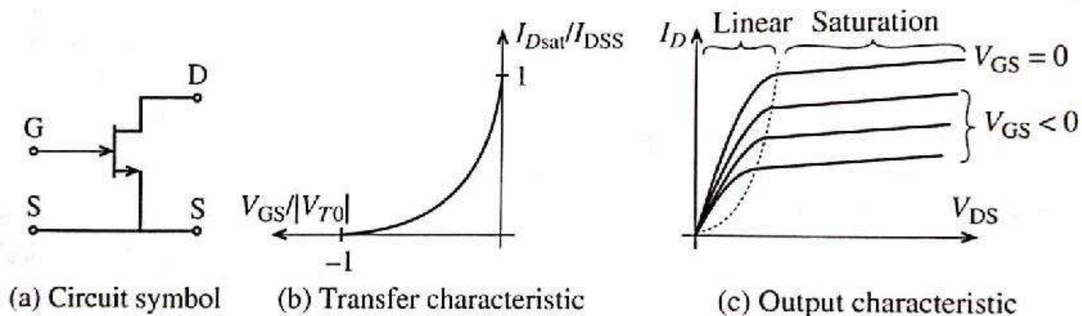
The maximum saturation current is obtained when $V_{GS} = 0$

$$I_{DSat} = G_O \left[\frac{V_P}{3} - (V_d) + \frac{2}{3\sqrt{V_P}} (V_d)^{\frac{3}{2}} \right] \quad (9)$$

The saturation drain current is often approximated by the simple relation

$$I_{DSat} = I_{DSS} \left(1 - \frac{V_{GS}}{V_{TO}}\right)^2 \quad (10)$$

The transfer and output characteristics of an n-channel MESFET is as shown below.



(a) Circuit symbol (b) Transfer characteristic (c) Output characteristic

Transfer and output characteristics of an *n*-channel MESFET

Problem:

1. A GaAs MESFET has the following parameters: $N_D = 10^{16}cm^{-3}$, d=0.75μm, W=10μm, L=2μm, ε_r=12.0, V_d=0.8v and μ_n=8500cm²/(Vs). Determine a) pinch-off voltage, b)Threshold Voltage, c) The maximum saturation current I_{DSS}.

Solution:

a) pinch-off voltage:

The pinch-off voltage for the FET is,

$$V_P = \frac{qN_D d^2}{2\varepsilon} = \frac{1.6x10^{-19}\, x\,(10^{16} x10^{6})\, x\,(0.75x10^{-6})^2}{2\, x\, 8.854\, x10^{-12}\, x\,12}$$ **= 4.235V**

b)Threshold Voltage,

V_{TO} = V_d-V_P = 0.8 - 4.235 = **-3.435v**

c) The maximum saturation current I_{DSS}

$$I_{DSat} = G_O\left[\frac{V_P}{3} - (V_d) + \frac{2}{3\sqrt{V_P}}(V_d)^{\frac{3}{2}}\right]$$

$$G_O = \frac{\sigma q N_D W d}{L} = \frac{q^2 \mu_n N_D{}^2 W d}{L} = \frac{(1.6x10^{-19})^2\, x\,(8500\, x10^{-4})x\,(10^{16} x10^{6})^2\, x(10x10^{-6})x(0.75x10^{-6})}{2x10^{-6}}$$ **=8.16**

$$I_{DSat} = 8.16\left[\frac{4.235}{3} - (0.8) + \frac{2}{3\sqrt{4.235}}(0.8)^{\frac{3}{2}}\right]$$ **=6.883A**

High Electron Mobility Transistor (HEMT)

High electron mobility transistor(HEMT) is a transistor that operates at higher frequencies, typically in the microwave range. They are used in applications that require high frequency, such as cell phones, RF applications, and some power applications.Essentially the device is a field-effect transistor that incorporates a junction between two materials with different band gaps (i.e. a heterojunction) as the channel instead of a doped region which is used in the standard MOSFET.

As a result of its structure, the HEMT may also be referred to as a heterojunction FET, HFET or modulation doped FET, MODFET on some occasions.

HEMTs are transistors that utilize the 2-dimensional electron gas(2DEG) created by a junction between two materials with different band gaps called a heterojunction. The two most commonly used materials to create the heterojunction are a highly doped n-type donor material, typically AlGaAs and an undoped material, typically GaAs.

HEMT's take advantage of 2DEG which is created at the AlGaAs/GaAsheterojunction. The 2DEG is confined at the heterojunction and free to move parallel to the channel. This results in a higher electron mobility which is good for large gain and high frequency characteristics.

HEMT structure & fabrication

The key element within a HEMT is the specialized PN junction that it uses. It is known as a hetero-junction and consists of a junction that uses different materials either side of the junction. The most common materials used aluminium gallium arsenide (AlGaAs) and gallium arsenide (GaAs).Gallium arsenide is generally used because it provides a high level of basic electron mobility which is crucial to the operation of the device. Silicon is not used because it has a much lower level of electron mobility.

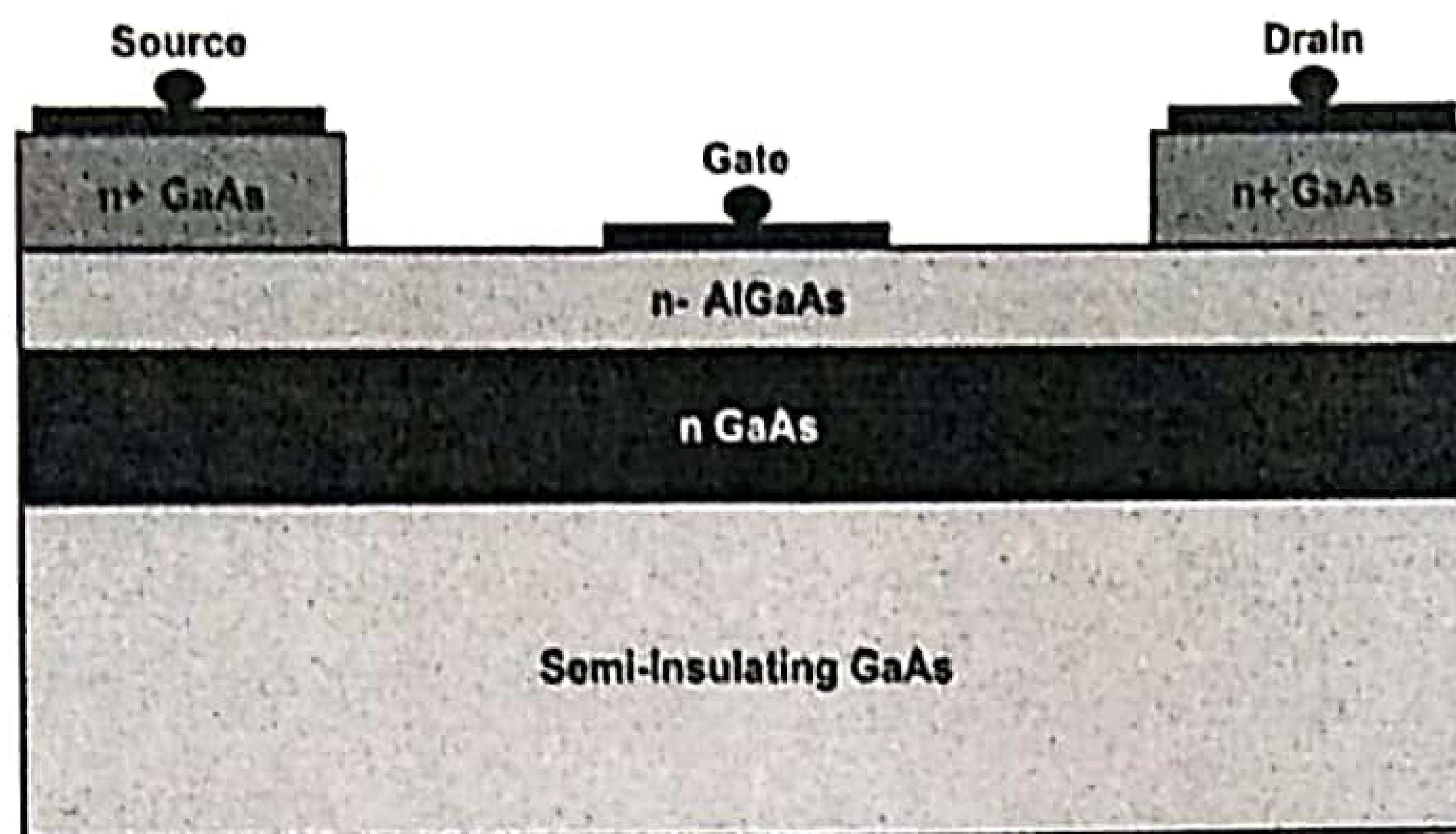
In the manufacture of a HEMT, first an intrinsic layer of gallium arsenide is set down on the semi-insulating gallium arsenide layer. This is only about one micron thick.Next a very thin layer (between 30 and 60 Angstroms) of intrinsic aluminium gallium arsenide is set down on top of this. Its purpose is to ensure the separation of the hetero-junction interface from the doped aluminium gallium arsenide region. This is critical if the high electron mobility is to be achieved.

The doped layer of aluminium gallium arsenide about 500 Angstroms thick is set down above this. Precise control of the thickness of this layer is required and special techniques are required for the control of this.

There are two main structures that are used. These are the self-aligned ion implanted structure and the recess gate structure. In the case of the self-aligned ion implanted structure the gate, drain and source are set down and are generally metallic contacts, although source and drain contacts may sometimes be made from germanium. The gate is generally made from titanium, and it forms a minute reverse biased junction similar to that of the GaAsFET.

Scanned by CamScanner

For the recess gate structure another layer of n-type gallium arsenide is set down to enable the drain and source contacts to be made. Areas are etched as shown in the diagram. The thickness under the gate is also very critical since the threshold voltage of the FET is determined by this. The size of the gate, and hence the channel is very small. Typically the gate is only 0.25 microns or less, enabling the device to have a very good high frequency performance.



HEMT operation

Electrons from the n-type region move through the crystal lattice and many remain close to the hetero-junction. These electrons form a layer that is only one electron thick forming what is known as a two dimensional electron gas. Within this region the electrons are able to move freely because there are no other donor electrons or other items with which electrons will collide and the mobility of the electrons in the gas is very high.

A bias applied to the gate formed as a Schottky barrier diode is used to modulate the number of electrons in the channel formed from the 2 D electron gas and in turn this controls the conductivity of the device. This can be compared to the more traditional types of FET where the width of the channel is changed by the gate bias.

Advantages of HEMT devices:

- ***High gain:*** HEMTs have a high gain at microwave frequencies because the charge carriers are almost exclusively the majority carriers and the minority carriers are not significantly involved.
- ***Low noise:*** HEMTs provide very low noise operation because the current variation in the devices is low when compared to other field effect devices.

Applications of HEMT

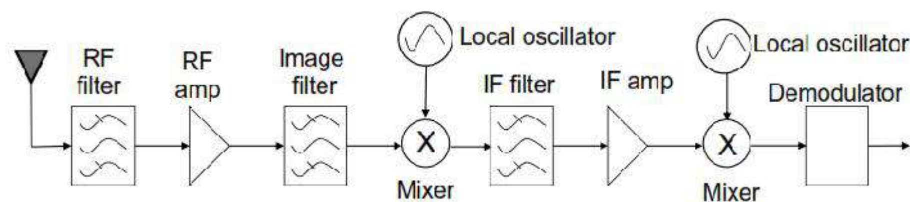
- Next generation wired/wireless communication
- Advanced radars
- Power electronics

Scanned by CamScanner

BASIC CONCEPTS OF RF DESIGN:

Radio-frequency (RF) engineering is about systems that operate at radio frequencies such as microwave frequencies. The RF portion of radio transmitters and receivers will be viewed as a subsystem of wireless systems. Thus, the relationships of the RF portion to other parts of the overall wireless system design will be pertinent. For example, radio receiver sensitivity depends on the RF design. RF generally includes other aspects, such as the device technologies and RF circuits (including active circuits and passive circuits). Real electronic components introduce noise and have other imperfections, such as nonlinearities. While the nature of noise, nonlinearities, and so on, is intimately related to the devices themselves, the results on the system can be studied and quantified at the systems level based on models of these effects.

The most popular wireless receiver architecture is known as the superheterodyne receiver. A block diagram of a superheterodyne radio receiver is as shown in figure below. In the figure shown below amplifiers, mixers, frequency synthesizers, and filters are the fundamental building blocks of the RF part of radios. Broadly speaking, an amplifier amplifies the power of a signal; a mixer is used to up-convert or down-convert a signal, by multiplying (also described as mixing) it with a periodic signal, such as would be produced by a frequency synthesizer. A frequency synthesizer may be as simple as an oscillator, or it may include an oscillator together with additional circuitry. A filter selects a band of frequencies to pass through, and attenuates signal components at other frequencies.



Superhetrodyne Receiver

The communications signals transmitted over wireless are at very high frequencies and so are often referred to as being "at RF." To demodulate the signals and detect what was transmitted, the RF section of the receiver often needs to bring the signal down to around baseband. The superheterodyne receiver brings the signal from RF down to around baseband in two stages; first, it down-converts from RF to an intermediate frequency (IF), and second, it down-converts from IF to around baseband. Having two stages of down-conversion introduces some challenges.

Noise and distortion are the limiting factors in the RF circuit performance. Quantifying noise and distortion is necessary to quantify the performance of a transceiver.

1. NOISE:

Noise is always being picked up by a receiver from the rest of the universe when a desired signal is being fed into a system. Noises can be introduced into a circuit during a radio signal transformation. There is one kind of noise, which is called thermal energy, generated due to the temperature related motion of charged particles. Thermal energy is caused by atoms and electrons move in a random way resulting in random currents in a circuit itself. On the other hands, there are also many other man-made noises coming from outside the circuit system, such as microwave, cell phones, and even power chargers. In order to check out how much noise has been added to a source signal, a ratio of the signal to noise power is defined for a receiver. The sum of thermal noise power and circuit generated noise presented at the receiver front-end is defined as the noise floor. To detect a reliable signal, the minimum detectable signal level must typically be larger than its noise floor.

Thermal Noise:

Resistors are the most possible components that will cause noise in a circuit. Due to thermal energy, noise will be generated in resistors causing random currents in the circuit. The formula of thermal noise in spectral density from resistors can be expressed as follows:

$$N_{resistor} = 4kTBR \quad (1)$$

where, k - Boltzman constant (1.38×10^{-23} J/K)

T - Kelvin temperature of resistor (300K)

B - Bandwidth

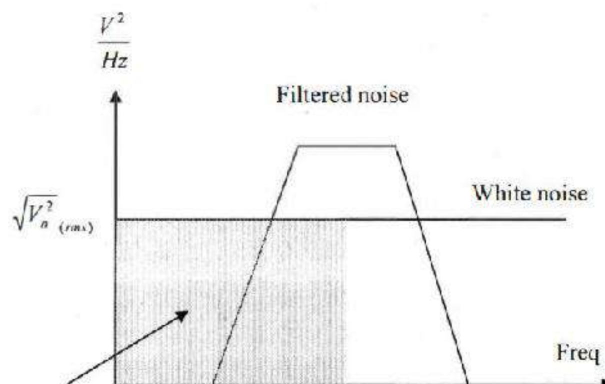
R - Value of Resistor

In additional, thermal noise is also white noise. This means that the thermal noise involves a constant power spectral density with respect to frequency. Therefore, to find out how much power is generated in a finite bandwidth in a resistor, the formula is presented as follows:

$$V_n^2 = 4kTR\Delta f \ (v^2/Hz) \quad (2)$$

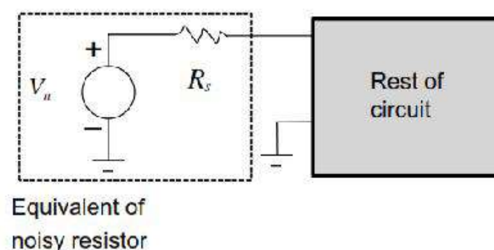
where, Δf - bandwidth

V_n^2 - noise voltage in rms value.



Noise power in spectral density respect to frequency.

Usually, the mean value of noise will be zero when noise is random. Therefore, in order to measure the dissipated noise power, it is needed to use mean square values. The below figure shows the spectral noise power density with respect to frequency. A model of resistor noise with a voltage source is as shown below.



Noise Factor and Noise Figure:

The Noise Figure (NF) describes how much noise is added to a signal by elements of a radio's receiver chain. There are many different ways to define NF, but the most common definition is,

$$NF = \frac{SNR_{in}}{SNR_{out}} \quad (1)$$

where SNR_{in} is the input SNR due to thermal noise and SNR_{out} is the device output SNR. The NF provides an indication of how the device degrades the SNR. The manufacturer of the device usually supplies the NF. In the case of passive components, the NF equals the loss of the passive components. Thus if a passive RF filter provides a 3 dB loss of signal, the NF is defined to be 3 dB.

Once the NF is determined, it is possible to provide an equivalent RF receiver NF, NF_{total}, that relates the noise back to the antenna. The Friis equation allows for the NF of all the devices in the RF chain to be combined.

$$NF_{total} = 1 + (NF_1 - 1) + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1 G_2} + \frac{NF_4 - 1}{G_1 G_2 G_3} + \frac{NF_5 - 1}{G_1 G_2 G_3 G_4} + \ldots\ldots \quad (2)$$

Here NF_i represents the NF at the ith stage and G_i represents the gain at the ith stage. This equation assumes a linear scale, although NF is usually used with a dB scale. Given a component with a noisy input having noise power P_{i-1} dBm, gain G_i dB, and NF NF_i dB, the output noise power P_i dBm is given by

$$P_i \text{ dBm} = P_{i-1} \text{ dBm} + NF_i \text{ dB} + G_i \text{ dB} \quad (3)$$

Once the overall NF is determined, it is possible to determine the minimum input signal level discernible by the receiver or the sensitivity of the receiver to achieve a minimal SNR, SNR_{min}. Sensitivity, S dBm is calculated with the basic definition,

$$S \text{ dBm} = \text{Noise floor dBm} + SNR_{min} \text{ dB} \quad (4)$$

The NF is calculated from the source thermal noise, its magnification due to the total NF, and the bandwidth over which this noise exists.

Noise floor dBm = 10 log(kT_eB) + NF_{total} dB

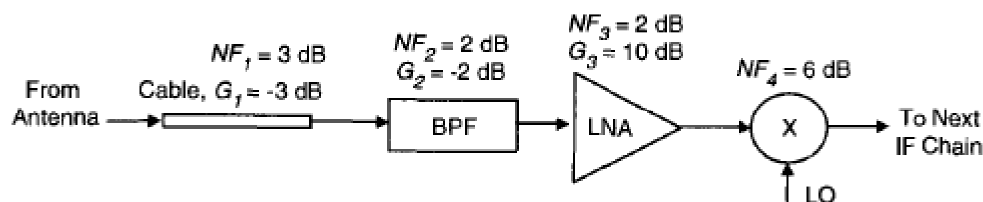
= 10 log(kT_e) dBm + 10 log (B) dB + NF_{total} dB (5)

Where, B is the end of the system bandwidth in Hz and T_e is taken as its usual value of room temperature, 290K. For room temperature, the sensitivity becomes

S = -174 dBm/Hz + NF dB + 10log(B) + SNR_{min} dB. (6)

Problem:

1. The block diagram of an RF stage of a receiver is as shown in figure. The transmission line is connected to the antenna, and the output of the mixer goes to the IF stage. Calculate the Noise factor and sensitivity of the receiver for a bandwidth of 1 MHz and minimal SNR of 12 dB.



Solution:

The linear values of the gain and individual NFs are

NF_1 = 2 NF_2 = 1.585 NF_3 = 1.585 NF_4 = 4

G_1 = 0.5 G_2 = 0.631 G_3 = 10

The total NF is given by

$$NF_{total} = 1 + (NF_1 - 1) + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1 G_2} + \frac{NF_4 - 1}{G_1 G_2 G_3} + \frac{NF_5 - 1}{G_1 G_2 G_3 G_4} + \ldots\ldots$$

$$NF_{total} = 1 + (2-1) + \frac{1.585 - 1}{0.5} + \frac{1.585 - 1}{0.5x0.631} + \frac{4 - 1}{0.5x0.631x10} = 5.98 \text{ Or } 7.76dB$$

So for a bandwidth of 1 MHz and minimal SNR of 12 dB, the sensitivity is

S = -174 dBm/Hz + 60 dB + 7.76 dB + 12 dB

S = -94.24 dB

Flicker noise:

Flicker noise is a type of electronic noise with a 1/f power spectral density. It is therefore often referred to as 1/f noise or pink noise. device. It is basically due to variation in the conduction mechanism. There is no outstanding solution to decreasing it yet, but techniques do exist to minimize the effect. The power in spectral density of 1/f noise is inversely proportional to frequency.

2. Distortion Characterization:

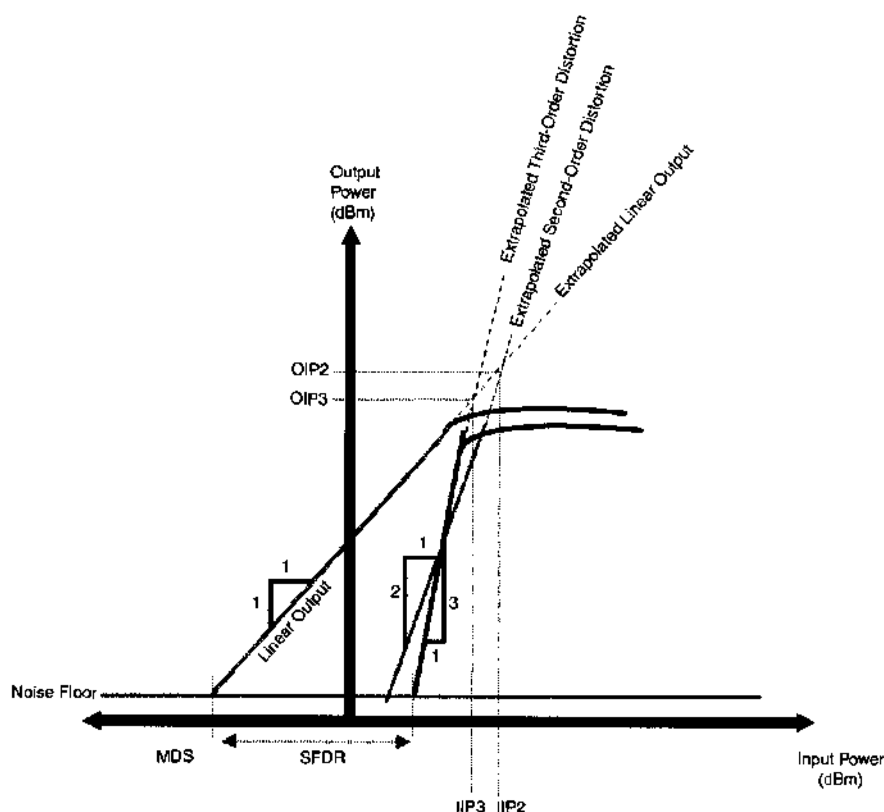
Distortion occurs in the RF chain because of the non-linearities in the system. The distortion takes the form of harmonics, i.e., sinusoidal terms that occur at multiples of the frequency of an input sinusoid. Distortion can take the form of cross modulation when a weak signal and a strong interferer enter a non-linearity and the amplitude of the interferer modifies the amplitude of the weak signal and vice versa.

In an ideal system, linear time-invariant (LTI) operations is expected which allows the outputs to be expressed as a linear combination of responses to inputs. For example, if there are two input signals, $x_1(t)$ and $x_2(t)$, the outputs of these signals can be: $x_1(t) \rightarrow y_1(t)$, $x_2(t) \rightarrow y_2(t)$

Therefore, a linear system has to be satisfied in the following condition.

$$a.x_1(t) + b.x_2(t) \rightarrow a.y_1(t) + b.y_2(t)$$

The relationship between increased input signal power and the output power of the desired signal and distortion is shown in Figure below. In a linear circuit, a linear relationship exists between input power and output power, and this linear relationship is represented as a line with a slope of one in below figure.



MDS = Minimum Detectable Signal (Output Noise Floor)
IIP3 = Third-Order Intercept Point
SFDR =1/3 IIP3 - MDS
OIP3 = Output Referred Third-Order Intercept Point
OIP2 = Output Referred Second-Order Intercept Point

Non-Linear Output Distortion Characterization Using Input Versus Output Power Characteristics.

However, in any real RF component, the transfer function is much more complicated. These complexities can be due to active or passive components in a RF circuit. It is common to have nonlinearity and time variance present in a system. Mathematically, any nonlinearity function can be written as a series expansion of power terms. Assume a nonlinear system $y(t) = \alpha_1.x(t) + \alpha_2.x_2(t) + \alpha_3.x_3(t)$ is memoryless and has an input signal $x(t) = A\cos(\omega t)$. where α_1, α_2, α_3 are functions of time. then, the result of the system is,

$$y(t) = \alpha_1.\ A\cos(\omega t) + \alpha_2.\ A^2\cos^2(\omega t) + \alpha_3.\ A^3\cos^3(\omega t).$$

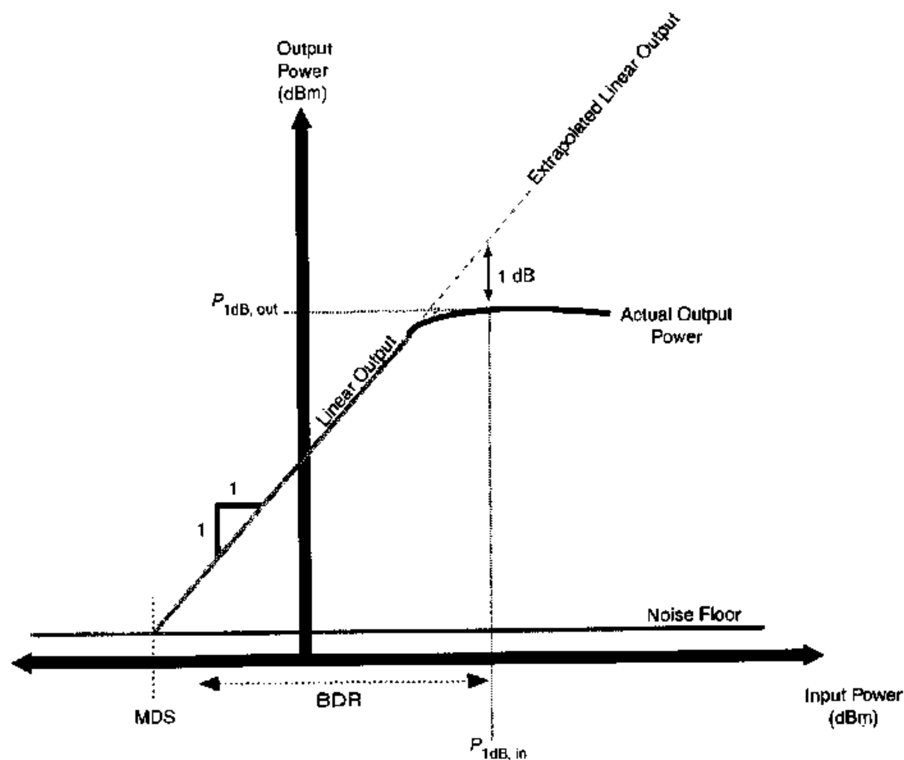
Typically a non-linear device output can be modeled as

$$v_o = a_0 + a_1 v_{in} + a_2 v^2_{in} + a_3 v^3_{in} + \ldots$$

where v_{in} is the input voltage and a_is are constant terms. The square term produces second-order products, and the cubic term represents third-order products.

For non-linear devices that exhibit a pure cubic characteristic, the third-order distortion power grows at three times the rate of the desired signal. Eventually, there is a practical limitation on how much the power of the desired signal can be raised with a corresponding linear increase in the output signal level. Eventually the device begins to saturate, and when the actual desired signal's output power level differs by 1 dB with the desired signal's ideal output value, the 1 dB compression, point P_1dB, is reached. This amplitude compression characteristic, shown in Figure below, tends to block the detection of lower level signals in the presence of stronger signals, and the blocking dynamic range (BDR), i.e., the difference between the minimum detectable signal (MDS) level and the input that produces a 1 dB compression, quantifies this effect:

$$BDR = P_1dB - MDS$$



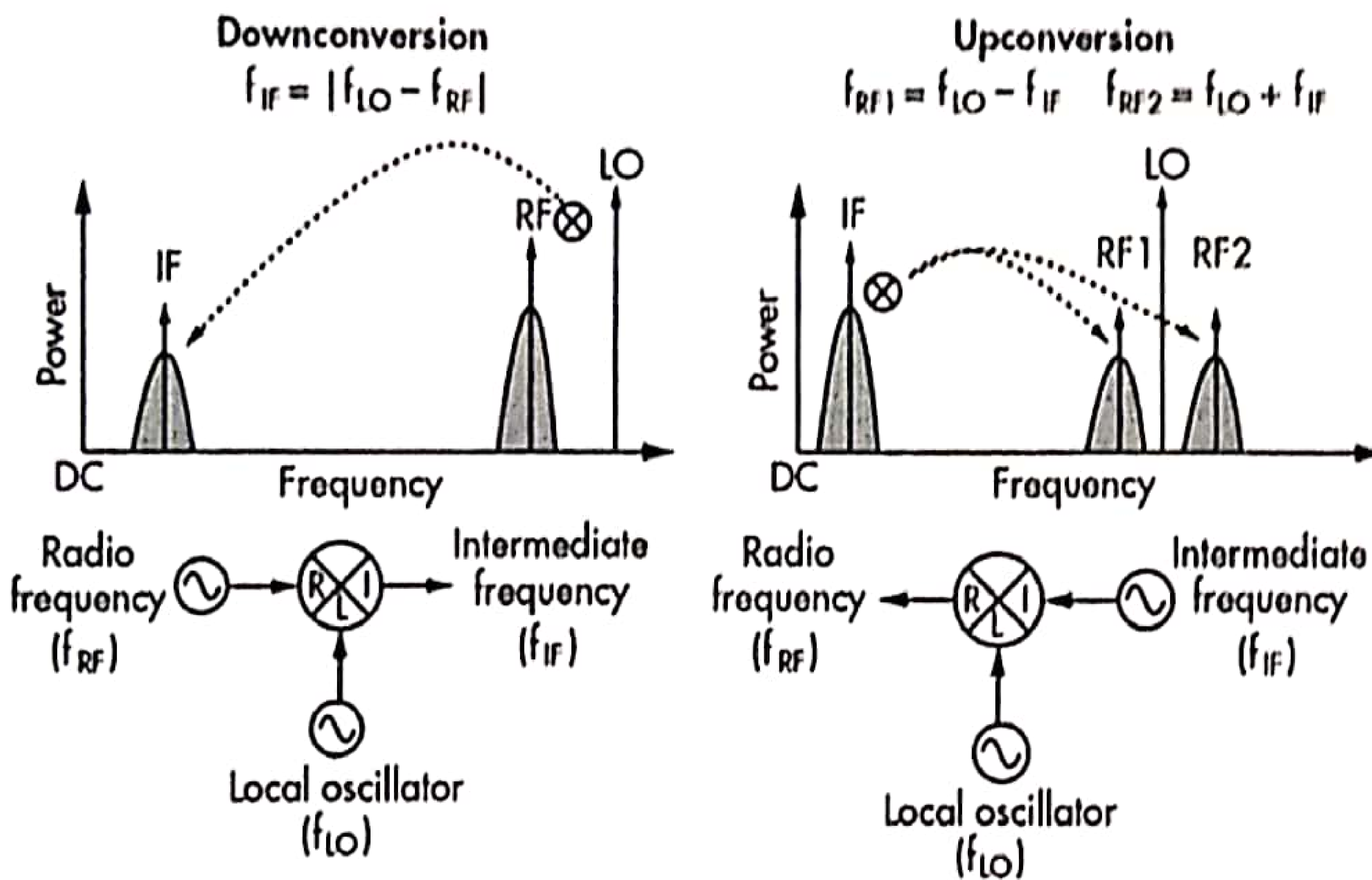
1 dB Compression Point

ADC and DAC Distortion:

The ADC and DAC introduce both noise and distortion. The main sources of noise are quantization noise, internal thermal noise, and sampling aperture jitter. Quantization noise occurs because of the limited number of states that can be represented by the ADC and DAC. The round-off or truncation is often modeled as an additive noise process onto a true signal representation. Thermal noise is a problem for all components. Aperture jitter (aperture uncertainty), the result of sampling at unevenly spaced intervals, produces a modulation on the phase of the signal that is typically modeled as background noise. This type of distortion is especially apparent for signals that have high frequency content. if the signal level exceeds the maximum range of the ADC, nonlinear distortion results.

Mixers

A mixer is a three-port component, which performs the task of frequency conversion. Mixers translate the frequency of an input signal to a different frequency. This functionality is vital for a wide range of applications, including military radar, satellite-communications (satcom), cellular base stations, and more. Mixers are used to perform both frequency upconversion and downconversion.



1.

These simple diagrams provide an illustration of frequency conversion. "Two of a mixer's three ports serve as inputs, while the other port serves as an output port. An ideal mixer produces an output that consists of the sum and difference frequencies of its two input signals. In other words:

$$f_{out} = f_{in1} \pm f_{in2}$$

The three ports of a mixer are known as the intermediate-frequency (IF), radio-frequency (RF), and local-oscillator (LO) ports. The LO port is usually an input port."

The RF and IF ports can be used interchangeably, depending on whether the mixer is being used to perform upconversion or downconversion. The LO signal is typically the strongest signal injected into a mixer. The required LO drive level is dependent on several factors, including the mixer's configuration and device technology.

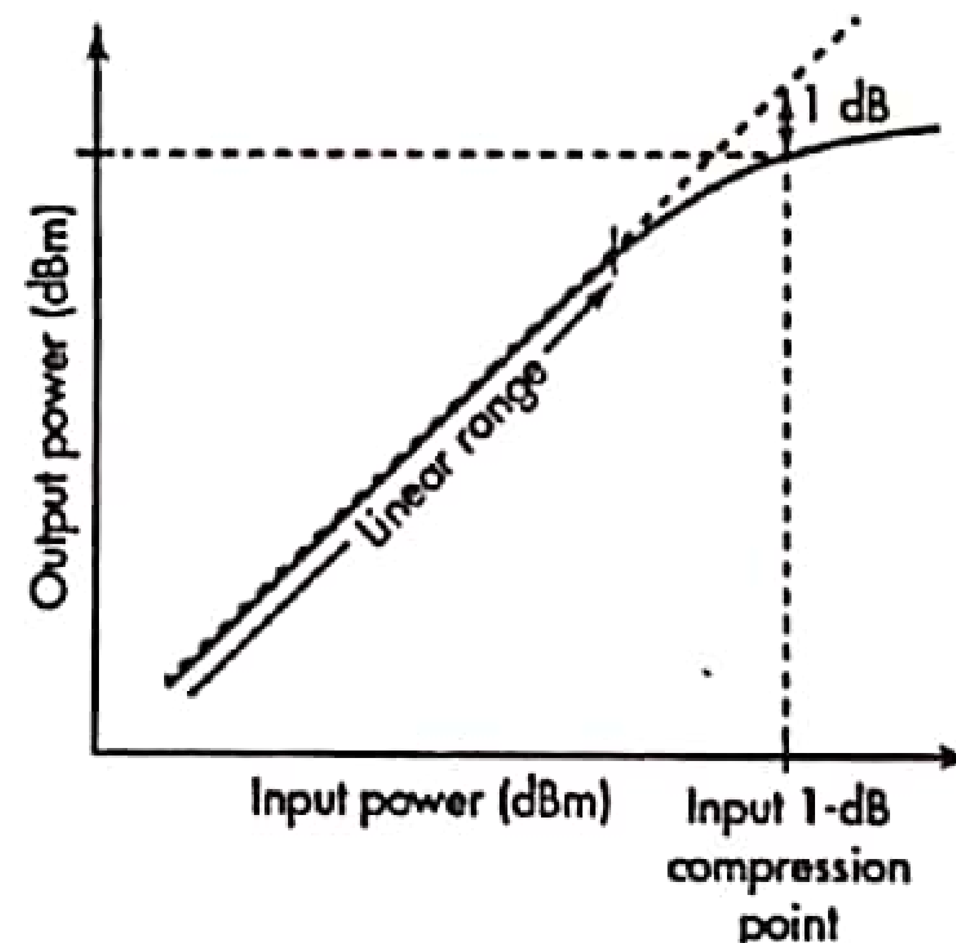
Scanned by CamScanner

When a mixer is used to perform downconversion, an input signal enters the RF port and an LO signal enters the LO port. These two input signals produce an output signal at the IF port. The frequency of this output signal is equal to the difference of the RF input signal's frequency and the LO signal's frequency.

When a mixer is used to perform upconversion, an input signal enters the IF port and an LO signal enters the LO port. These two input signals produce an output signal at the RF port. The frequency of this output signal is equal to the sum of the IF input signal's frequency and the LO signal's frequency. Both downconversion and upconversion are shown graphically in *Fig. 1.* Upconversion is normally part of a transmitter, while downconversion is typically used in a receiver.

Mixer Performance Parameters

Conversion Loss: In passive mixers, conversion loss is defined as the difference in signal level between the amplitude of the input signal and the amplitude of the desired output signal. In a mixer used for downconversion, the conversion loss is the difference between the RF input signal's amplitude and the IF output signal's amplitude. In a mixer used for upconversion, the conversion loss is the difference between the IF input signal's amplitude and the RF output signal's amplitude.



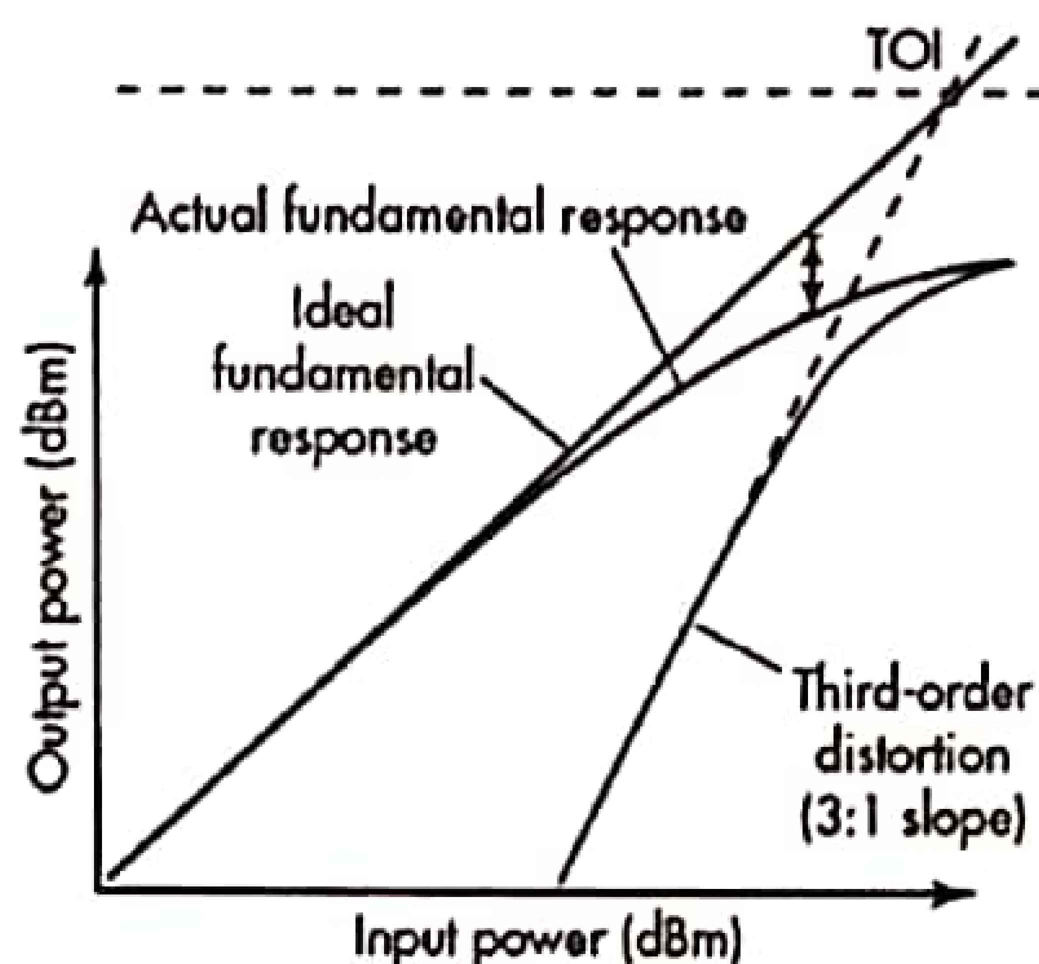
The above fig. is a graphical representation of 1-dB compression point.

Isolation: "Isolation is a measurement of the amount of power that leaks from one port to another. Isolation is defined as the difference in signal level between the amplitude of an input signal and the amplitude of the leaked power from that input signal to another port." When isolation is high, the amount of power leaked from one port to a different port is small.

Scanned by CamScanner

"Three types of isolation are commonly quoted in microwave mixers: LO-RF isolation, LO-IF isolation, and RF-IF isolation."

1-dB Compression Point: A mixer's conversion loss remains constant when the mixer is in linear operation. As the amplitude of the input signal increases, the amplitude of the output signal rises by the same amount. However, once the input signal's amplitude reaches a certain level, the amplitude of the output signal ceases to exactly follow the input signal. The mixer deviates from linear behavior and its conversion loss begins to increase.



Intermodulation Distortion: Two-tone third-order intermodulation distortion (IMD) occurs when two signals simultaneously enter the mixer's IF or RF input port. In practice, this could happen in a multi-carrier signal environment. These two signals interact with each other and with the LO signal, which creates distortion. In a receiver, two-tone third-order IMD is a serious problem because it can generate third-order distortion products that fall within the IF bandwidth.

If f_{RF1} and f_{RF2} represent two separate RF input signals and f_{LO} represents the LO signal, the third-order distortion products generated at the mixer's IF port are:

$$Interferer_1 = 2f_{RF1} - f_{RF2} - f_{LO}$$
$$Interferer_2 = 2f_{RF2} - f_{RF1} - f_{LO}$$

These third-order distortion products are extremely close to the desired IF output frequency. No amount of filtering can remove these unwanted distortion products. Thus, the signal-to-noise ratio of the received signal is degraded, highlighting the need to suppress these distortion products.

The third-order input intercept point (TOI or IP3) is a widely accepted figure of merit used to describe a mixer's capability to suppress third-order distortion products. TOI is used in predicting the nonlinear behavior of a mixer as the amplitude of its input signal increases, which

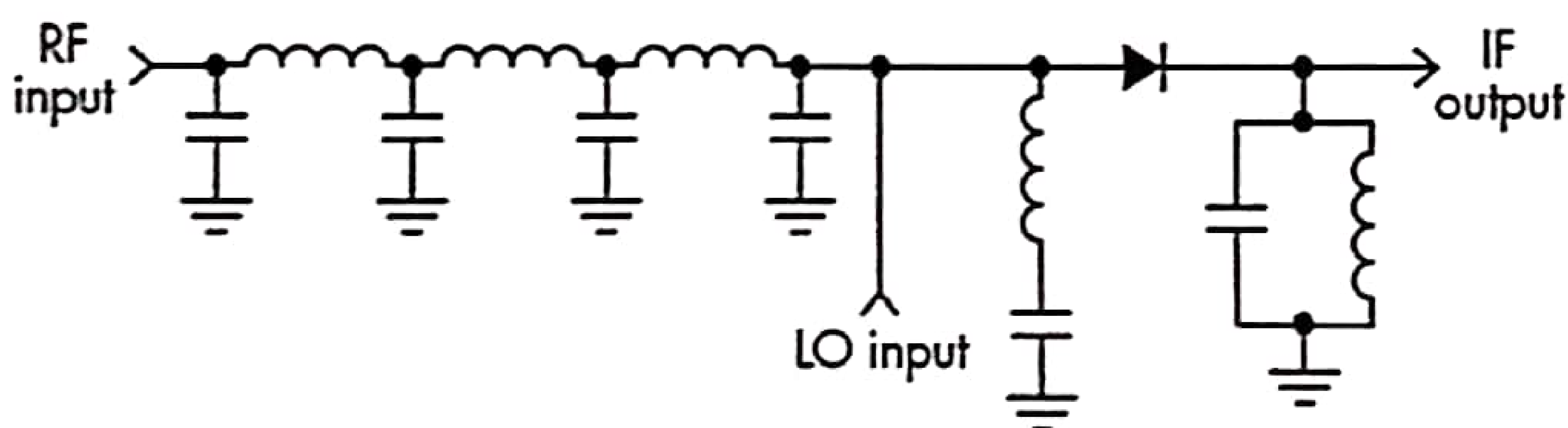
Scanned by CamScanner

causes the third-order products to increase by a 3:1 ratio. For any 1-dB increase in the input signal's amplitude, the third-order products increase by 3 dB

Mixer Design Techniques

In theory, any nonlinear device can be used to create a mixer circuit. However, only a few devices satisfy the requirements needed to design mixers with acceptable performance. Devices that are commonly used to design modern mixers include Schottky diodes, gallium-arsenide (GaAs) field-effect transistors (FETs), and CMOS transistors. Various topologies can be used to design mixers. Mixers can be designed as either passive or active components.

Single diode mixer



Passive mixers primarily use Schottky diodes, although the FET resistive mixer has recently become another popular passive mixer. Active mixers use either FETs or bipolar devices. Schottky diodes, in comparison with FETs and bipolar devices, have the advantage of possessing an inherently wide bandwidth. This is a major reason why diodes are still widely used to design mixers.

Mixers can be designed with just a single diode, which is the simplest mixer topology. Balanced mixers, which consist of two, four, or even eight diodes in a balanced structure, build upon the single-diode mixer. The majority of mixers available today incorporate some form of mixer balancing.A single diode can be used to create a mixer . Here, the RF and LO signals combine at the anode of the diode. The LO signal needs to be large enough to switch the diode on and off, which causes the actual mixing process. The frequency components generated by single-diode mixers are:

$f_{IF} = nf_{LO} \pm mf_{RF}$ (m and n are all integers)

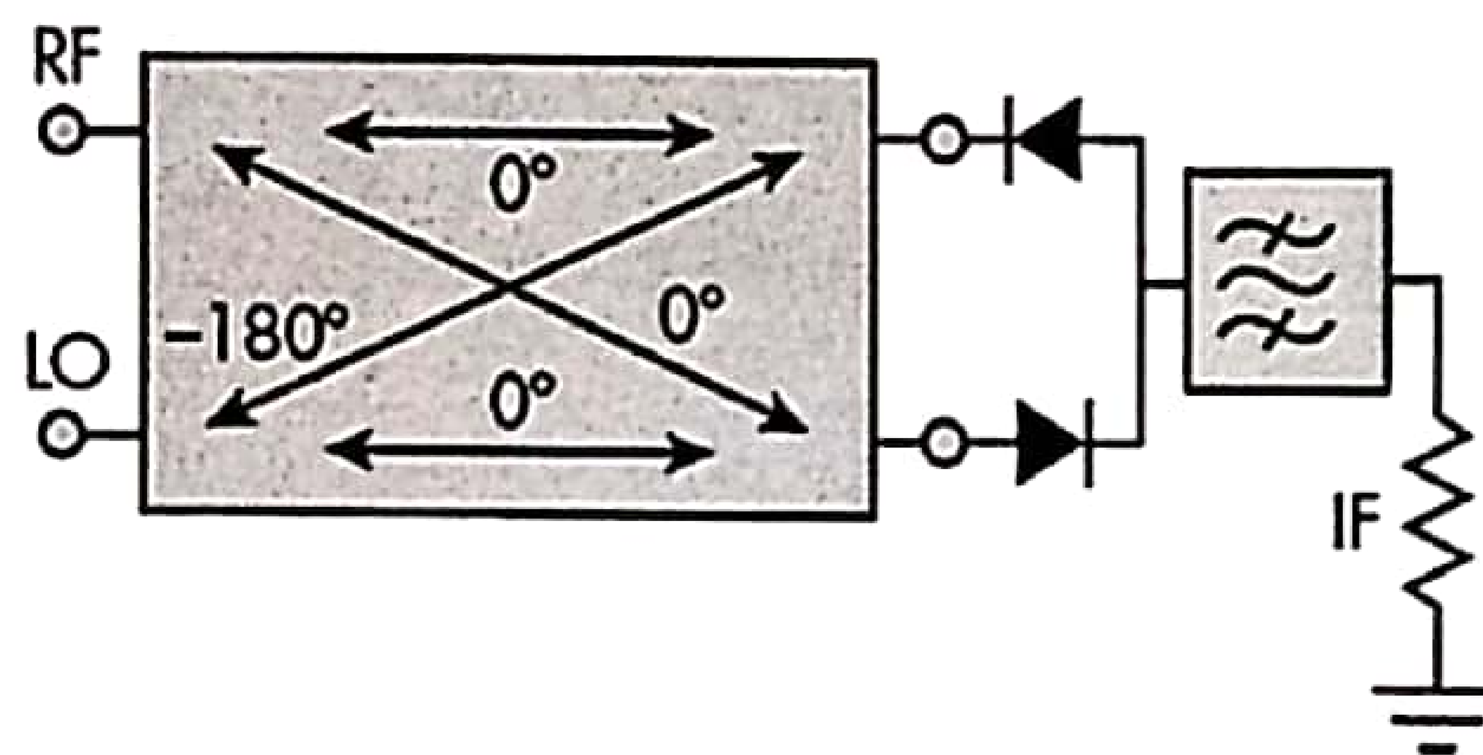
where:

Scanned by CamScanner

f_{LO} = the LO input signal frequency
f_{RF} = the RF input signal frequency
f_{IF} = the IF output signal frequency

Although only one output frequency is desired (when n = 1 and m = 1), additional unwanted harmonics are generated by the diode's current-voltage (I-V) characteristics and the transconductance modulation caused by the RF signal. Because the single-diode mixer has no inherent isolation between the RF and LO ports, external filters also are needed to achieve isolation between ports. This need for external filtering makes it difficult to achieve wideband mixers with just a single diode.

A single-balanced mixer consists of two diodes and a hybrid.



Balanced mixers overcome some of the limitations of single-diode mixers. They do require baluns or hybrids, which largely determine the bandwidth and overall performance of the mixer. Inherent isolation between ports is achieved by balanced mixers as well as increased cancellation of intermodulation products. Common-mode noise cancellation is another advantage gained by balanced mixers. However, balanced mixers do require a higher LO drive level.

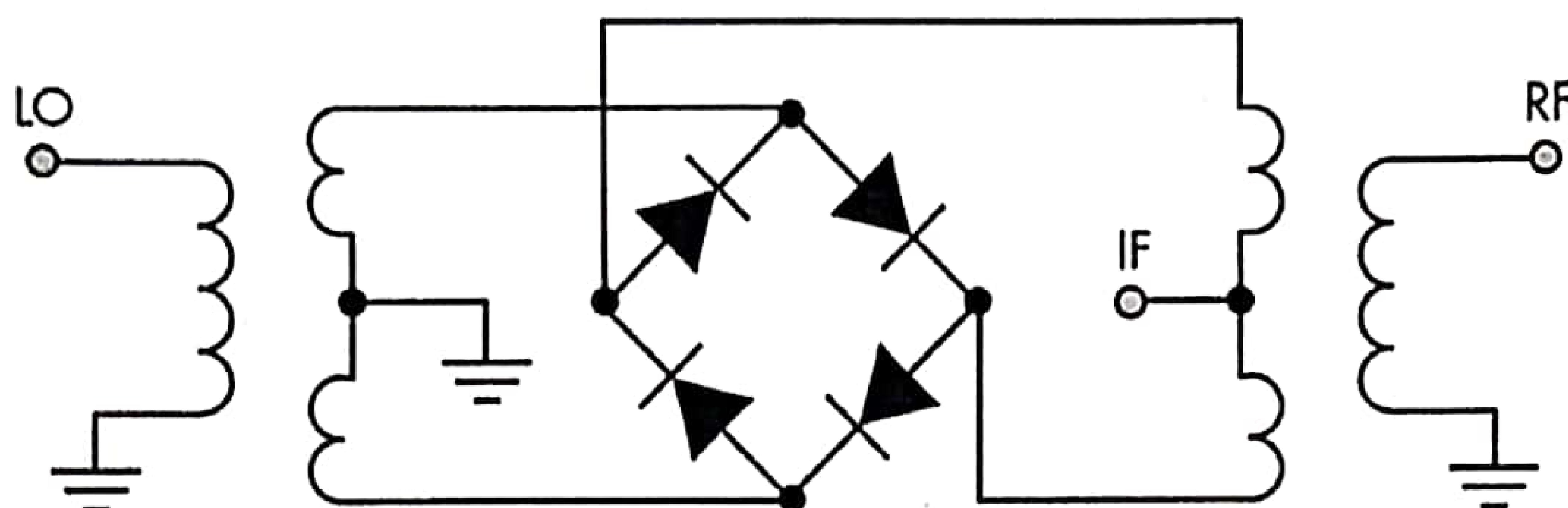
Single-balanced mixers consist of two diodes along with a hybrid (*Fig. 5*). Although 90-deg. and 180-deg. hybrids can both be used to design single-balanced mixers, the majority of single-balanced mixers incorporate a 180-deg. hybrid. The 180-deg. hybrid's input ports are mutually isolated, enabling the LO port to be isolated from the RF port. This provides frequency-band independence and equal power division to the load. In comparison with single-diode mixers, single-balanced mixers also have 50% fewer intermodulation products.

Two single-balanced mixers can be combined to form a double-balanced mixer. Traditional double-balanced mixers are typically based on four Schottky diodes in a quad ring configuration. Baluns are placed at both the RF and LO ports, while the IF signal is tapped off from the RF balun. The IF signal can also be tapped off from the LO balun, but this would worsen the LO-IF isolation.

Scanned by CamScanner

For this reason, it is usually preferred to tap off the IF signal from the RF balun instead of the LO balun. An example of a double-balanced mixer is shown in Fig. 6. This mixer has high LO-RF isolation and LO-IF isolation along with moderate RF-IF isolation. Double-balanced mixers also have the benefit of reducing intermodulation products by as much as 75% in comparison with single-diode mixers.

Double balanced mixer



An even more complex mixer circuit is the triple-balanced mixer. Triple-balanced mixers have separate baluns for the LO, RF, and IF ports, which enables them to achieve high LO-RF isolation, LO-IF isolation, and RF-IF isolation. Triple-balanced mixers also offer higher suppression of intermodulation products than double-balanced mixers. The downside of triple-balanced mixers is that they need a higher LO drive level. They also are greater in both size and complexity.

Applications

Mixer circuits can be used to shift the frequency of an input signal like as in a receiver. They can also be used as a product detector, modulator, frequency multiplier or phase detector.

Scanned by CamScanner

VCO

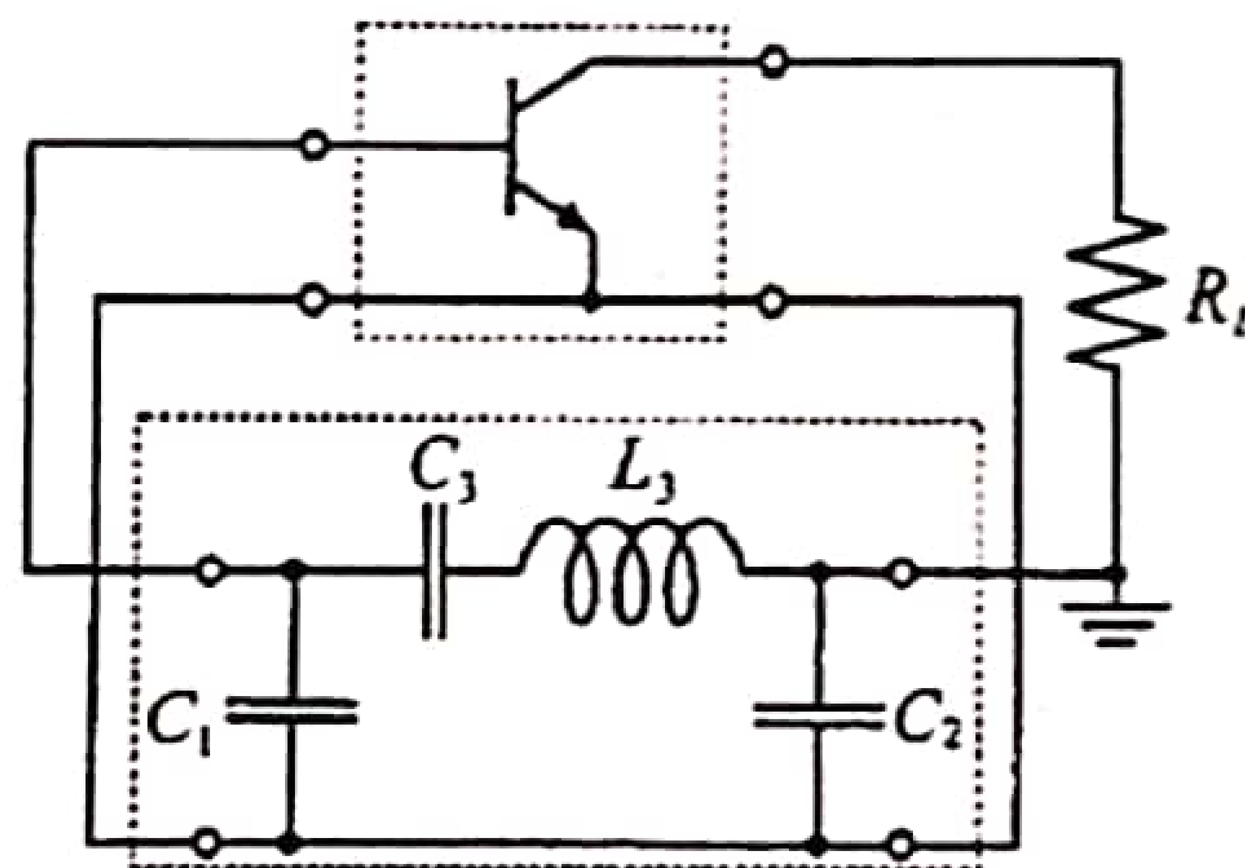
A voltage-controlled oscillator (VCO) is an electronic oscillator whose oscillation frequency is controlled by a voltage input. The applied input voltage determines the instantaneous oscillation frequency. Consequently, a VCO can be used for frequency modulation (FM) or phase modulation (PM) by applying a modulating signal to the control input. A VCO is also an integral part of a phase-locked loop.

Types of Voltage Controlled Oscillators

- Harmonic Oscillators: The output is a signal with sinusoidal waveform. Examples are crystal oscillators and tank oscillators
- Relaxation Oscillators: The output is a signal with saw tooth or triangular waveform and provides a wide range of operational frequencies. The output frequency depends on the time of charging and discharging of the capacitor.

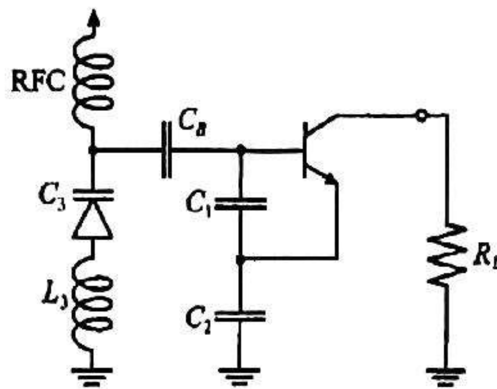
Applications of VCO

- Tone Generators
- Function generators
- Phase-Locked Loops
- In synthesizers to generate variable tones for the production of electronic music
- In communication equipment these are used as frequency synthesizers
- Clock generators
- Frequency Shift Keying



(a) Pi-type feedback loop

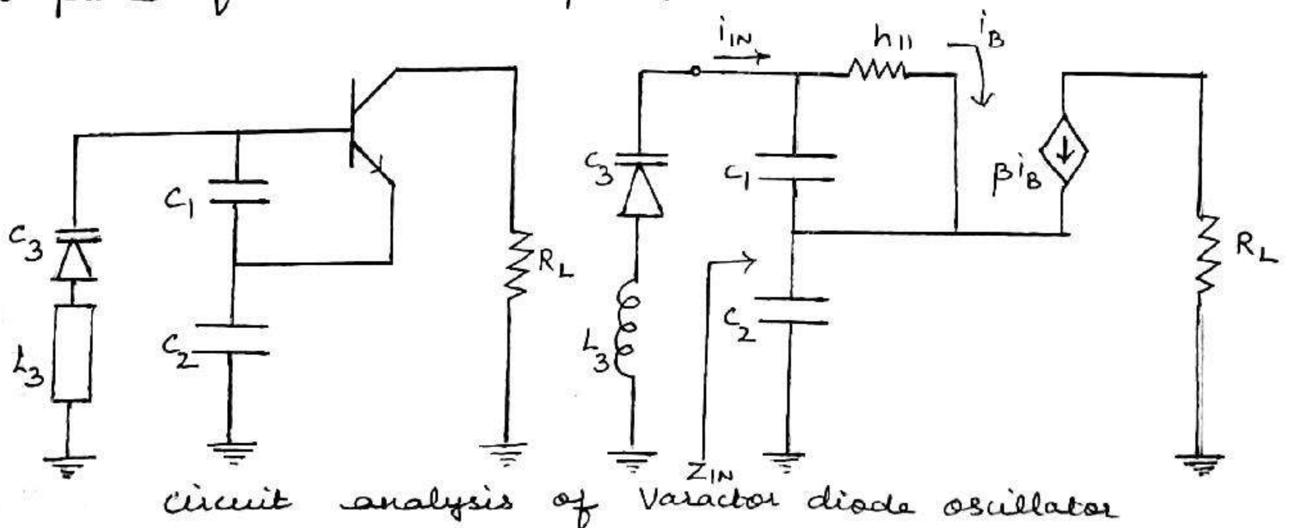
Scanned by CamScanner



(b) Redrawn circuit with DC isolation

The feedback loop for the clapp oscillator shown in Fig. 1(a) can be modified by replacing C_3 with a Varactor diode as shown in Fig. 1. (b).

In Fig. 1. (c) the Varactor diode and a transmission line element, whose length is adjusted to be inductive, form the termination circuit connected to the input of the oscillator. If the Varactor diode & transmission line segment is disconnected, the input impedance Z_{IN} can be computed from two loop equations:



Circuit analysis of Varactor diode oscillator

$$V_{IN} - i_{IN} X_{c1} - i_{IN} X_{c2} + i_B X_{c1} - \beta i_B X_{c2} = 0 \quad \text{——} (1)$$

$$h_{11} i_B + i_B X_{c1} - i_{IN} X_{c1} = 0 \quad \text{——} (2)$$

From equation (2),

$$i_B (h_{11} + X_{c1}) = i_{IN} X_{c1}$$

$$\therefore \quad i_B = \frac{i_{IN} X_{c1}}{h_{11} + X_{c1}} \quad \text{——} (3)$$

Sub. (3) in (1)

$$V_{IN} + i_B (X_{c1} - \beta X_{c2}) = i_{IN} X_{c1} + i_{IN} X_{c2}$$

$$V_{IN} + \frac{i_{IN} X_{c1}}{h_{11} + X_{c1}} (X_{c1} - \beta X_{c2}) = i_{IN} (X_{c1} + X_{c2})$$

$$V_{IN} = i_{IN} (X_{c1} + X_{c2}) - \frac{i_{IN} X_{c1}}{h_{11} + X_{c1}} (X_{c1} - \beta X_{c2})$$

$$= i_{IN} \left[(X_{c1} + X_{c2}) - \frac{X_{c1}}{h_{11} + X_{c1}} (X_{c1} - \beta X_{c2}) \right]$$

$$= i_{IN} \left[\frac{(h_{11} + X_{c1}) (X_{c1} + X_{c2}) - X_{c1} (X_{c1} - \beta X_{c2})}{h_{11} + X_{c1}} \right]$$

$$V_{IN} = \frac{i_{IN}}{h_{11} + X_{c1}} \left[h_{11} (X_{c1} + X_{c2}) + X_{c1} X_{c2} (1 + \beta) \right]$$

$$\therefore \quad Z_{IN} = \frac{V_{IN}}{i_{IN}} = \frac{1}{h_{11} + X_{c1}} \left[h_{11} (X_{c1} + X_{c2}) + X_{c1} X_{c2} (1 + \beta) \right] \quad \text{——} (4)$$

Equation (4) can further be simplified by considering

$(1+\beta) \simeq \beta$ and $h_{11} >> X_{c1}$

$$\therefore Z_{IN} = \frac{1}{h_{11}}\left[h_{11}(X_{c1}+X_{c2}) + X_{c1}X_{c2}\beta\right]$$

$$= (X_{c1}+X_{c2}) + \frac{\beta}{h_{11}} X_{c1}X_{c2}$$

$X_{c1} = \frac{1}{j\omega C_1}$ & $X_{c2} = \frac{1}{j\omega C_2}$

$$= \frac{1}{j\omega C_1} + \frac{1}{j\omega C_2} + \frac{\beta}{h_{11}}\cdot\frac{1}{j\omega C_1}\cdot\frac{1}{j\omega C_2}$$

$$\therefore Z_{IN} = \frac{1}{j\omega}\left(\frac{1}{C_1}+\frac{1}{C_2}\right) \text{ \& } -\frac{\beta}{h_{11}}\left(\frac{1}{\omega^2 C_1 C_2}\right) \quad \text{—⑤}$$

From equation ⑤ & by considering $g_m = \frac{\beta}{h_{11}}$

$$R_{IN} = -\frac{g_m}{\omega^2 C_1 C_2}$$

$$\text{\&} \quad X_{IN} = \frac{1}{j\omega C_{IN}} \quad \left[\text{Where } C_{IN} = \frac{C_1 C_2}{C_1 + C_2}\right]$$

The resonant frequency can be obtained by considering the condition $X_1 + X_2 + X_3 = 0$.

$$\text{ie. } j\left(\omega_0 L_3 - \frac{1}{\omega_0 C_3}\right) - \frac{1}{j\omega_0}\left[\frac{1}{C_1}+\frac{1}{C_2}\right] = 0$$

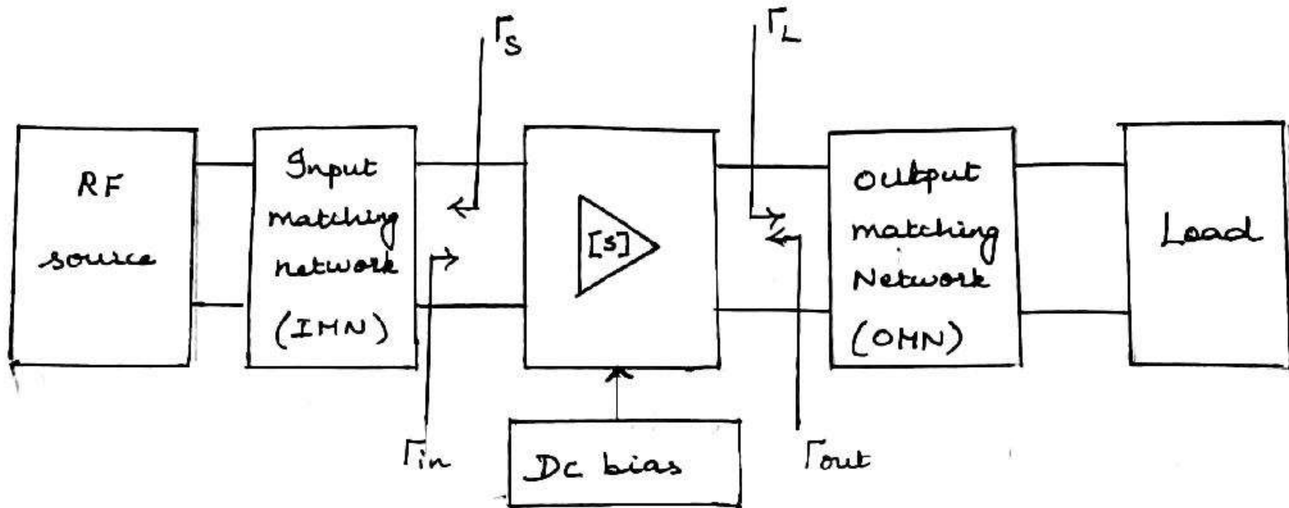
By simplifying the above equation we get,

$$\boxed{f_0 = \frac{1}{2\pi}\sqrt{\frac{1}{L_3}\left(\frac{1}{C_1}+\frac{1}{C_2}+\frac{1}{C_3}\right)}}$$

Scanned with CamScanner

Amplifier

A generic single stage amplifier configuration embedded between input and output matching networks is shown in the following figure; The amplifier is characterized through its S-parameter matrix at a particular DC bias point.



Generic amplifier system

S-Parameter matrix:

S-parameters (Scattering parameters) are power wave descriptors that permit us to define the input-output relations of a network in terms of incident and reflected power waves.

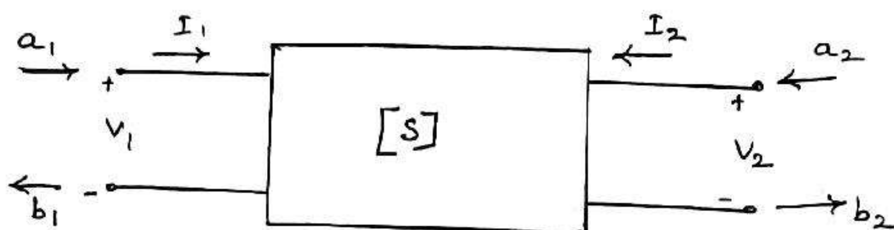
a_n refers to incident normalized power wave

b_n refers to reflected normalized power wave

Where index n refers to port number 1 or 2.

Scanned with CamScanner

Consider a 2 port network. a_1 & a_2 represent incident power wave at port 1 and port 2 respectively. b_1 and b_2 represent reflected / transmitted power wave at port 1 and port 2 respectively.



Two port network

The S parameter matrix for the above network is given as;

$$\begin{Bmatrix} b_1 \\ b_2 \end{Bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{Bmatrix} a_1 \\ a_2 \end{Bmatrix}$$

where $S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0} = \frac{\text{reflected power wave at port 1}}{\text{incident power wave at port 1}}$

$$S_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0} = \frac{\text{transmitted power wave at port 2}}{\text{incident power wave at port 1}}$$

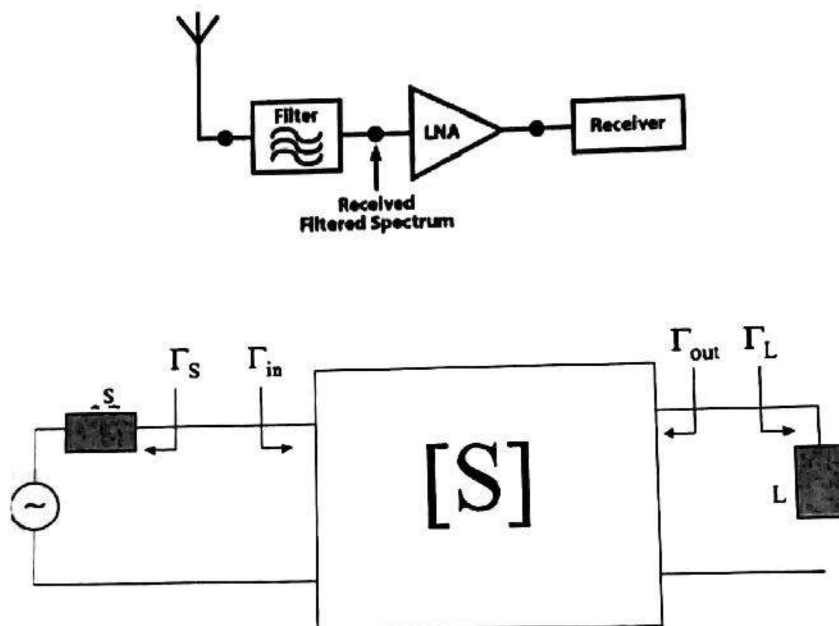
$$S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0} = \frac{\text{reflected power wave at port 2}}{\text{incident power wave at port 2}}$$

$$S_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0} = \frac{\text{transmitted power wave at port 1}}{\text{incident power wave at port 2}}$$

Scanned with CamScanner

Low noise Amplifier

A **low-noise amplifier** (LNA) is an electronic amplifier that amplifies a very low-power signal without significantly degrading its signal-to-noise ratio. An amplifier will increase the power of both the signal and the noise present at its input, but the amplifier will also introduce some additional noise. LNAs are designed to minimize that additional noise. Designers can minimize additional noise by choosing low-noise components, operating points, and circuit topologies. Minimizing additional noise must balance with other design goals such as power gain and impedance matching.



LNAs are found in radio communications systems, medical instruments and electronic test equipment. A typical LNA may supply a power gain of 100 (20 decibels (dB)) while decreasing the signal-to-noise ratio by less than a factor of two (a 3 dB noise figure (NF)). Although LNAs are primarily concerned with weak signals that are just above the noise floor, they must also consider the presence of larger signals that cause intermodulation distortion.

Low noise amplifiers are the building blocks of communication systems and instruments. The four important parameters in LNA design are: gain, noise figure, non-linearity and impedance matching.

Applications

LNAs are used in communications receivers such as in cellular telephones, GPS receivers, wireless LANs (WiFi), and satellite communications.

In a satellite communications system, the ground station receiving antenna uses an LNA because the received signal is weak since satellites have limited power and therefore use low-power transmitters. The satellites are also distant and suffer path loss: low earth orbit satellites might be 200 km (120 miles) away; a geosynchronous satellite is 35,786 miles (57,592 km) away. The

Scanned with CamScanner

LNA boosts the antenna signal to overcome feed line losses between the antenna and the receiver.

LNAs are becoming increasingly popular for enhancing the performance of software-defined radio (SDR) receiver systems. SDRs are typically designed to be general purpose and therefore the noise figure is not optimized for any one particular application. With a LNA and appropriate filter, the receive sensitivity and performance can be greatly enhanced at any particular frequency or range of frequencies.

RF Power amplifier

A **radio frequency power amplifier (RF power amplifier)** is a type of electronic amplifier that converts a low-power radio-frequency signal into a higher power signal. Typically, RF power amplifiers drive the antenna of a transmitter. Design goals often include gain, power output, bandwidth, power efficiency, linearity (low signal compression at rated output), input and output impedance matching, and heat dissipation.

Amplifier classes

Many modern RF amplifiers operate in different modes, called "classes", to help achieve different design goals. Some classes are class A, class AB, class B, class C, which are considered the linear amplifier classes. In these classes the active device is used as a controlled current source. The bias at the input determines the class of the amplifier.

A common trade-off in power amplifier design is the trade-off between efficiency and linearity. The previously named classes become more efficient, but less linear, in the order they are listed. Operating the active device as a switch results in higher efficiency, theoretically up to 100%, but lower linearity.Among the switch-mode classes are Class D, Class F and class E.[2] The Class D amplifier is not often used in RF applications because the finite switching speed of the active devices and possible charge storage in saturation could lead to a large I-V product, which deteriorates efficiency.

Applications

The basic applications of the RF power amplifier include driving to another high power source, driving a transmitting antenna and exciting microwave cavity resonators. Among these applications, driving transmitter antennas is most well known. The transmitter–receivers are used not only for voice and data communication but also for weather sensing (in the form of a radar)

RF power amplifiers using LDMOS (laterally diffused MOSFET) are the most widely used power semiconductor devices in wireless telecommunication networks, particularly mobile networks. LDMOS-based RF power amplifiers are widely used in digital mobile networks such as 2G, 3G, and 4G.

Scanned with CamScanner

Transducer power gain:

The transducer power gain (G_T) quantifies the gain of the amplifier placed between source and load.

$$G_T = \frac{\text{Power delivered to the load}}{\text{available power from the source}} = \frac{P_L}{P_A}$$

Where,

$$P_L = \frac{1}{2} |b_2|^2 \cdot (1 - |\Gamma_L|^2)$$

where,

* Γ_L & Γ_S refers to reflection coefficient at load & source respectively.
* $b_S \rightarrow$ Source transmitted power

$$\& \quad P_A = \frac{1}{2} |b_S|^2 (1 - |\Gamma_S|^2)$$

$$b_2 = \frac{S_{21} a_1}{1 - S_{22}\Gamma_L}$$

$$b_S = \left[1 - \left(S_{11} + \frac{S_{21} S_{12} \Gamma_L}{1 - S_{22}\Gamma_L}\right)\Gamma_S\right] a_1$$

$$\therefore G_T = \frac{P_L}{P_A} = \frac{|b_2|^2}{|b_S|^2} (1 - |\Gamma_L|^2)(1 - |\Gamma_S|^2)$$

To find $\frac{b_2}{b_S}$:

$$\frac{b_2}{b_S} = \frac{\dfrac{S_{21} a_1}{1 - S_{22}\Gamma_L}}{\left[1 - \left(S_{11} + \dfrac{S_{21} S_{12} \Gamma_L}{1 - S_{22}\Gamma_L}\right)\Gamma_S\right] a_1}$$

Scanned with CamScanner

$$= \frac{\dfrac{S_{21}}{1 - S_{22}\Gamma_L}}{\left\{ (1 - S_{22}\Gamma_L) - \dfrac{(S_{11}(1 - S_{22}\Gamma_L) + S_{21}S_{12}\Gamma_L)\Gamma_S}{1 - S_{22}\Gamma_L} \right\}}$$

$$= \frac{S_{21}}{(1 - S_{22}\Gamma_L) - S_{11}\Gamma_S(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_S}$$

$$= \frac{S_{21}}{(1 - S_{22}\Gamma_L)(1 - S_{11}\Gamma_S) - S_{21}S_{12}\Gamma_L\Gamma_S}$$

$\therefore$ G_T is given as,

$$G_T = \frac{|S_{21}|^2 (1 - |\Gamma_L|^2)(1 - |\Gamma_S|^2)}{|(1 - S_{11}\Gamma_S)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_S|^2}$$

The input & output reflection coefficient is given as

$$\Gamma_{in} = S_{11} + \frac{S_{21}S_{12}\Gamma_L}{1 - S_{22}\Gamma_L}$$

$$\& \quad \Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S}$$

Therefore, with respect to Γ_{in}, G_T can be expressed as,

$$G_T = \frac{|S_{21}|^2 (1 - |\Gamma_L|^2)(1 - |\Gamma_S|^2)}{\left|\left(1 - \left(\Gamma_{in} - \dfrac{S_{21}S_{12}\Gamma_L}{1 - S_{22}\Gamma_L}\right)\Gamma_S\right)(1 - S_{22}\Gamma_L) - S_{21}S_{12}\Gamma_L\Gamma_S\right|^2}$$

Scanned with CamScanner

$$= \frac{|S_{21}|^2 (1-|\Gamma_L|^2)(1-|\Gamma_S|^2)}{\left|(1-S_{22}\Gamma_L) - (\Gamma_{IN}\Gamma_S(1-S_{22}\Gamma_L)) - S_{21}S_{12}\Gamma_L\Gamma_S + S_{21}S_{12}\Gamma_L\Gamma_S\right|^2}$$

$$\therefore G_T = \frac{|S_{21}|^2 (1-|\Gamma_L|^2)(1-|\Gamma_S|^2)}{|1-\Gamma_{in}\Gamma_S|^2 \, |1-S_{22}\Gamma_L|^2}$$

Similarly, in terms of output reflection coefficient, G_T can be expressed as,

$$G_T = \frac{|S_{21}|^2 (1-|\Gamma_L|^2)(1-|\Gamma_S|^2)}{|1-\Gamma_L\Gamma_{out}|^2 \, |1-S_{11}\Gamma_S|^2}$$

Unilateral power gain G_{TU} can be obtained by neglecting S_{12}. (ie. $S_{12}=0$).

$$\therefore G_{TU} = \frac{(1-|\Gamma_L|^2)\,|S_{21}|^2\,(1-|\Gamma_S|^2)}{|1-\Gamma_L S_{22}|^2 \, |1-S_{11}\Gamma_S|^2}$$

Scanned with CamScanner

Stability considerations:

Stability circles:

The main requirement of an amplifier circuit is to possess stable performance over the entire frequency range. If $|\Gamma| > 1$, then the return voltage increases in magnitude (positive feedback) causing instability. Conversely, if $|\Gamma| < 1$, the return voltage decreases in magnitude (negative feedback) causing stability.

Let us consider amplifier as a two port network characterized through its S-parameters, with external terminations described by Γ_L and Γ_S. Stability implies that the magnitude of the reflection coefficients are less than unity.

ie. $|\Gamma_L| < 1$, $|\Gamma_S| < 1$

$$|\Gamma_{in}| = \left| \frac{S_{11} - \Gamma_L \Delta}{1 - S_{22}\Gamma_L} \right| < 1$$

$$|\Gamma_{out}| = \left| \frac{S_{22} - \Gamma_S \Delta}{1 - S_{11}\Gamma_S} \right| < 1$$

Where $\Delta = S_{11} S_{22} - S_{12} S_{21}$

Scanned with CamScanner

Since the S parameters are fixed for a particular frequency, the only factors that have a parametric effect on stability are Γ_L and Γ_S.

The output stability circle equation is given as,

$$\left(\Gamma_L^R - C_{out}^R\right)^2 + \left(\Gamma_L^I - C_{out}^I\right)^2 = r_{out}^2$$

where the circle radius is given by,

$$r_{out} = \frac{|S_{12}\,S_{21}|}{\left|\,|S_{22}|^2 - |\Delta|^2\right|}$$

& the center of this circle is located at

$$C_{out} = \frac{(S_{22} - S_{11}^{*}\,\Delta)^{*}}{|S_{22}|^2 - |\Delta|^2}$$

Similarly, the input stability circle equation is given as,

$$\left(\Gamma_S^R - C_{in}^R\right)^2 + \left(\Gamma_S^I - C_{in}^I\right)^2 = r_{in}^2$$

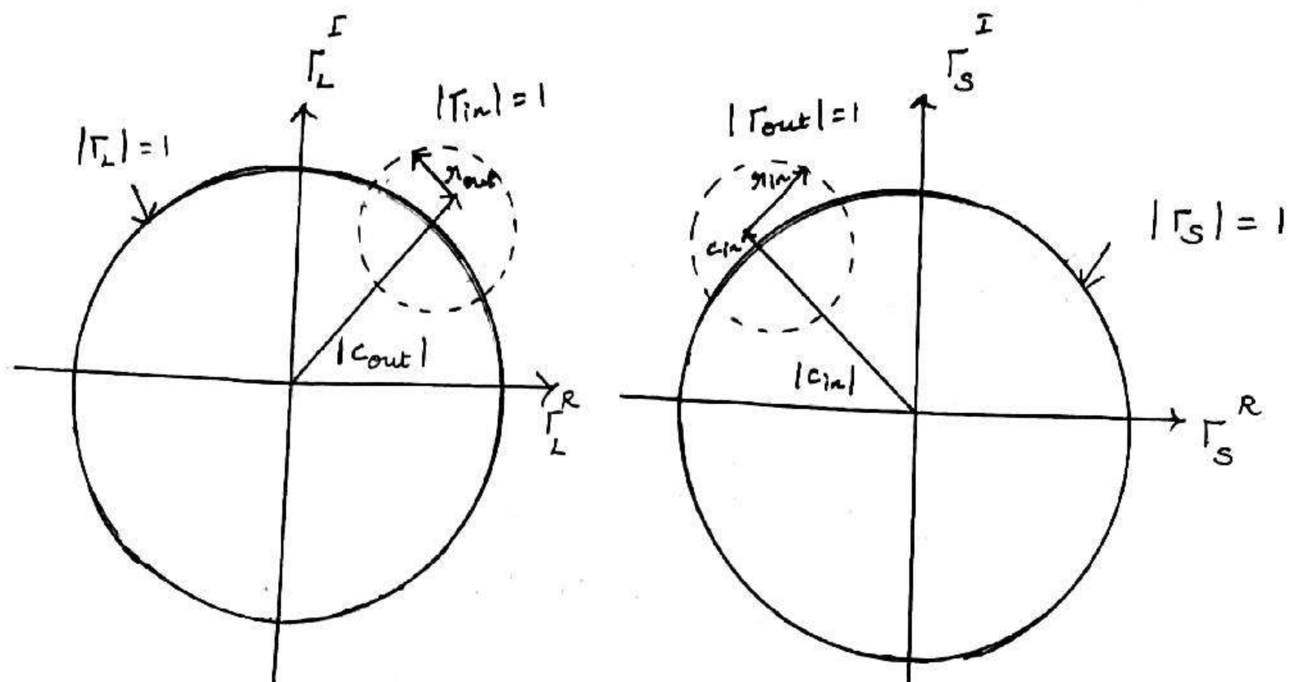
Where,

$$r_{in} = \frac{|S_{12}\,S_{21}|}{\left|\,|S_{11}|^2 - |\Delta|^2\right|}$$

& $$C_{in} = \frac{(S_{11} - S_{22}^{*}\,\Delta)^{*}}{|S_{11}|^2 - |\Delta|^2}$$

Scanned with CamScanner

The output & input stability circle is shown in the following figure;



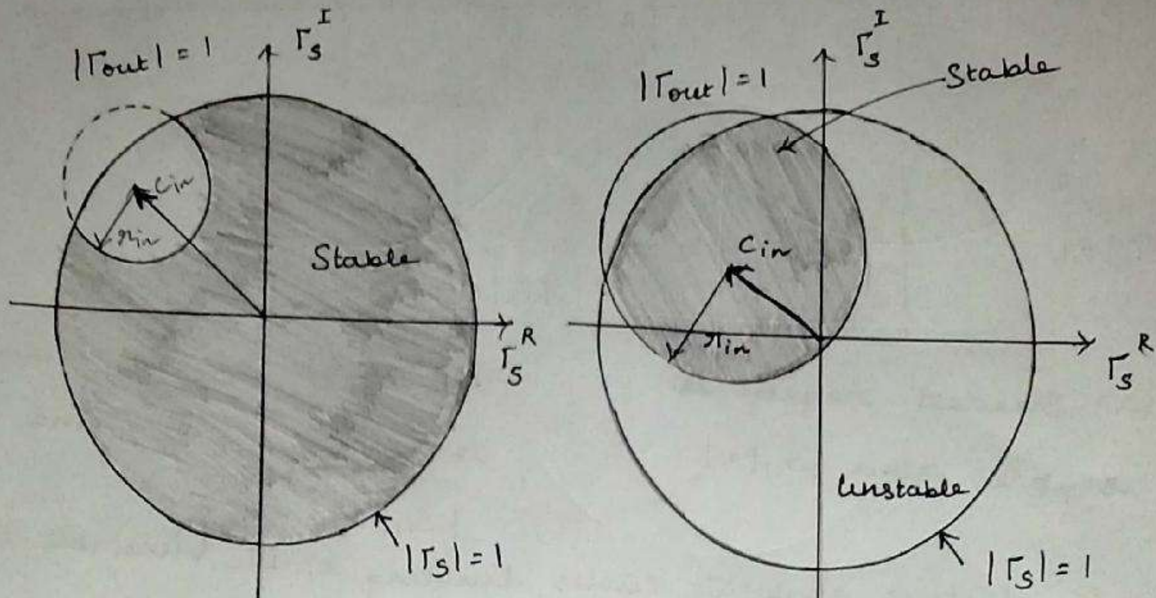
(a) Output stability circle (b) Input stability circle

Fig. (1)

For output stability circle, the rule is that if $|S_{11}| < 1$, the center ($\Gamma_L = 0$) must be stable; otherwise the center becomes unstable for $|S_{22}| > 1$.

Output stability circles denoting stable & unstable region is shown in Fig. 2.

The following diagram depicts the input stability circles for $|S_{22}| < 1$ & the two possible stability domains depending on $r_{in} < |c_{in}|$ or $r_{in} > |c_{in}|$.



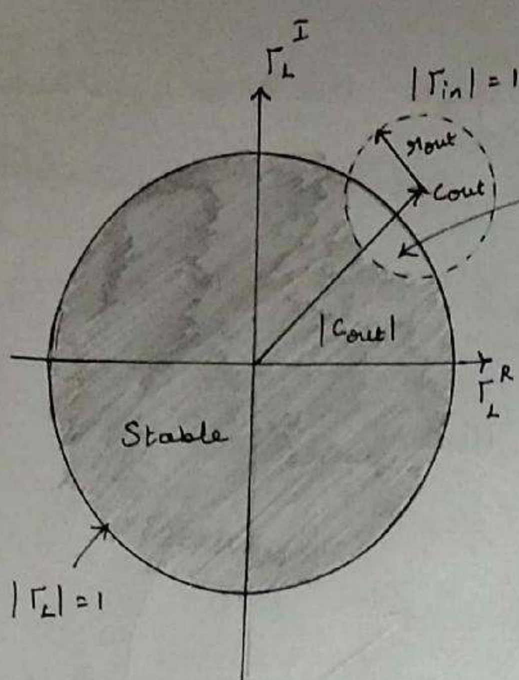
Different input stability regions for $|S_{22}| < 1$ depending on ratio between r_s and $|c_{in}|$.

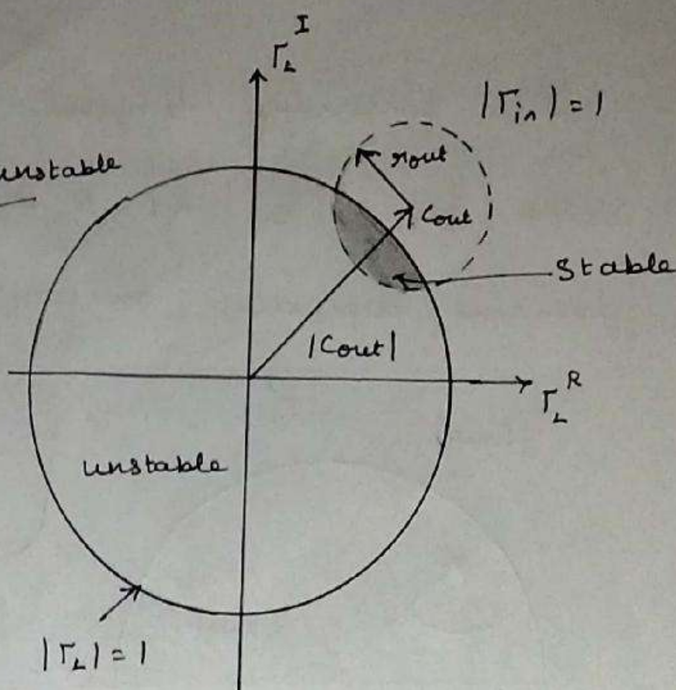
<u>Unconditional</u> <u>stability</u>:

Unconditional stability refers to the situation where the amplifier remains stable for any passive source & load at the selected frequency & bias conditions. For $|S_{11}| < 1$ & $|S_{22}| < 1$, it is stated as

$$||c_{in}| - r_{in}| > 1$$

$$||c_{out}| - r_{out}| > 1$$



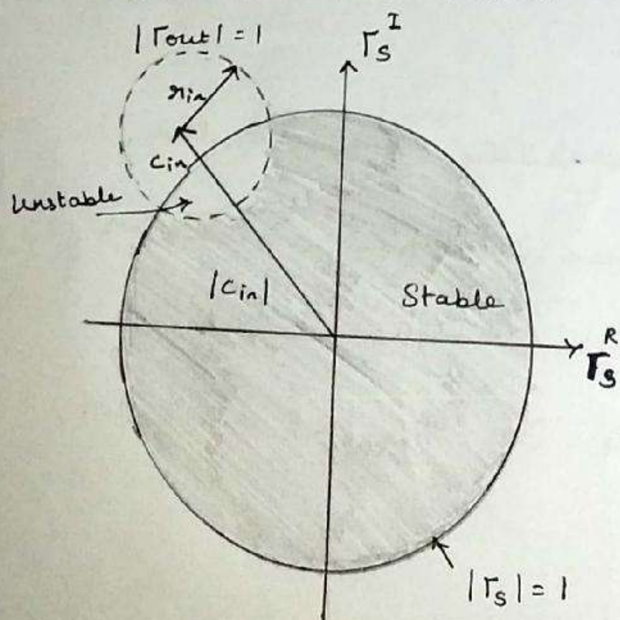


(a) Shaded region is stable, since $|S_{11}| < 1$

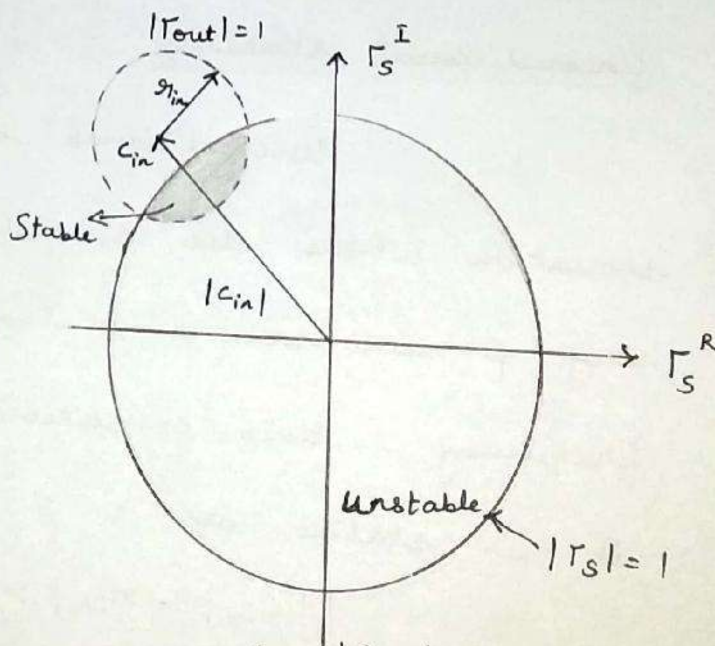
(b) Stable region excludes the origin, $\Gamma_L = 0$, sine $|S_{11}| > 1$.

Fig (2) Output stability circles denoting stable & unstable regions

For input stability circle, the rule is that if $|S_{22}| < 1$, the center ($\Gamma_S = 0$) must be stable; otherwise the center becomes unstable for $|S_{22}| > 1$.



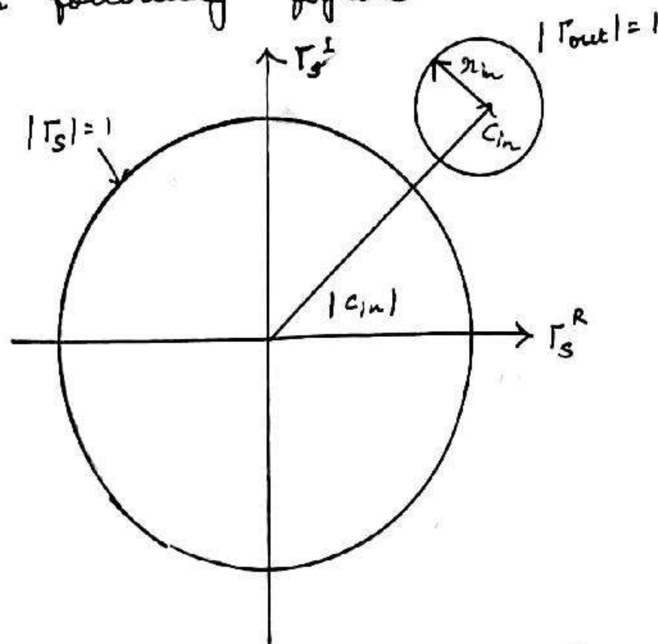
(a) $|S_{22}| < 1$



(b) $|S_{22}| > 1$

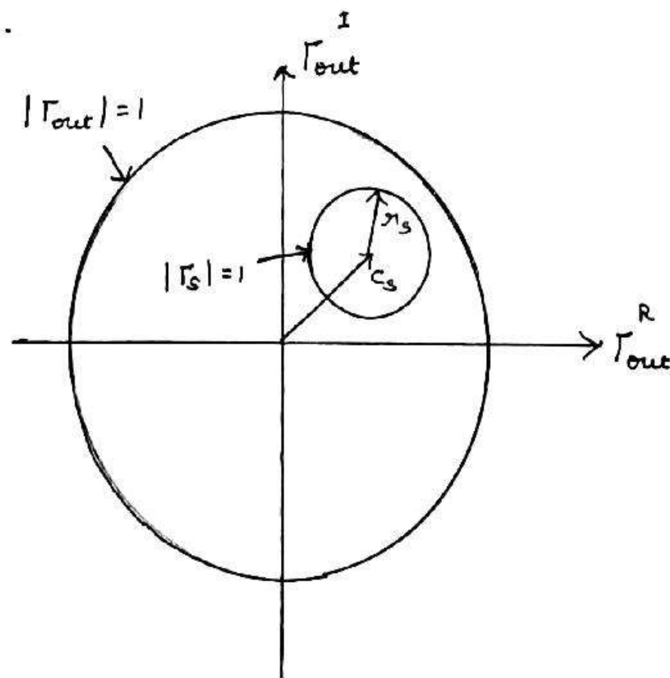
Input stability circles denoting stable & unstable regions

In other words, stability circles have to reside completely outside the $|\Gamma_S|=1$ and $|\Gamma_L|=1$ circles as shown in the following figure.



$|\Gamma_{out}|=1$ circle must reside outside

Alternatively, unconditional stability can also be viewed in terms of Γ_S behavior in Γ_{out} plane. Here, the $|\Gamma_S|\le 1$ domain must reside completely within the $|\Gamma_{out}|=1$ circle.



$|\Gamma_S|=1$ circle must reside inside

Scanned with CamScanner

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

GE8077 TOTAL QUALITY MANAGEMENT

Semester - 05

Notes

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

Vision

To excel in providing value based education in the field of Electronics and Communication Engineering, keeping in pace with the latest technical developments through commendable research, to raise the intellectual competence to match global standards and to make significant contributions to the society upholding the ethical standards.

Mission

- ✓ To deliver Quality Technical Education, with an equal emphasis on theoretical and practical aspects.
- ✓ To provide state of the art infrastructure for the students and faculty to upgrade their skills and knowledge.
- ✓ To create an open and conducive environment for faculty and students to carry out research and excel in their field of specialization.
- ✓ To focus especially on innovation and development of technologies that is sustainable and inclusive, and thus benefits all sections of the society.
- ✓ To establish a strong Industry Academic Collaboration for teaching and research, that could foster entrepreneurship and innovation in knowledge exchange.
- ✓ To produce quality Engineers who uphold and advance the integrity, honour and dignity of the engineering.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

1. To provide the students with a strong foundation in the required sciences in order to pursue studies in Electronics and Communication Engineering.
2. To gain adequate knowledge to become good professional in electronic and communication engineering associated industries, higher education and research.
3. To develop attitude in lifelong learning, applying and adapting new ideas and technologies as their field evolves.
4. To prepare students to critically analyze existing literature in an area of specialization and ethically develop innovative and research oriented methodologies to solve the problems identified.
5. To inculcate in the students a professional and ethical attitude and an ability to visualize the engineering issues in a broader social context.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO1: Design, develop and analyze electronic systems through application of relevant electronics, mathematics and engineering principles.

PSO2: Design, develop and analyze communication systems through application of fundamentals from communication principles, signal processing, and RF System Design & Electromagnetics.

PSO3: Adapt to emerging electronics and communication technologies and develop innovative solutions for existing and newer problems.

GE 6757
TOTAL QUALITY MANAGEMENT

UNIT I - INTRODUCTION

Introduction – Need for Quality – Evolution of quality- Definition of quality – Dimensions of manufacturing and service quality – Basic concepts of TQM – Definition of TQM – TQM Frame work – Contributions of Deming, Juran and Crosby – Barriers to TQM - Quality statements – Customer focus – Customer orientation, Customer satisfaction, Customer complaints, Customer retention – Costs of Quality.

INTRODUCTION TO QUALITY

Total Quality Management (TQM) is an enhancement to the traditional way of doing business.

Total - Made up of the whole

Quality - Degree of Excellence a Product or Service provides.

Management - Art of handling, controlling, directing etc.

TQM is the application of quantitative methods and human resources to improve all the processes within an organization and exceed CUSTOMER NEEDS now and in the future.

One of the important issues that business has focused on in the last two decades is "quality". The other issues are cost and delivery. Quality has been widely considered as a key element for success in business in the present competitive market. Quality refers to meeting the needs and expectations of customers. It is important to understand that quality is about more than a product simply working properly. Quality refers to certain standards and the ways and means by which those standards are achieved, maintained and improved. Quality is not just confined to products and services. It is a homogeneous element of any aspect of doing things with

high degree of perfection. For example Business success depends on the quality decision making.

EVOLUTION OF QUALITY

PERIOD	EVENTS
Prior to the 20th century	Quality is an art Demands overcome potential production An era of workmanship
F.Taylor 1900s	The scientific approach to management resulting in rationalization of work and its break down leads to greater need for standardization, inspection and supervision
Shewart 1930s	Statistical beginnings and study of quality control. In parallel, studies by R A Fisher on experimental design; the beginning of control charts at western Electric in USA
Late 1930s	Quality standards and approaches are introduced in France and Japan. Beginning of SQC, reliability and maintenance engineering
1942	Seminal work by Deming at the ministry of war in USA on quality control and sampling Working group setup by Juran and Dodge on SQC in US army Concepts of acceptance sampling devised
1944	Daodge and Deming carried out seminal research on acceptance sampling
1945	Founding of the Japan standard association
1946	Founding of the ASQC
1950	Visit of Deming in Japan at the invitation of K Ishikawa
1951	Quality assurance increasingly accepted
1954	TQC in Japan ; Book published 1956
1957	Founding of European organization for the control of quality
1961	The Martin Co in USA introduces the zero defects

	approach while developing and producing Pershing Missiles. Quality motivation is starting in the US and integrated programmes begun
1962	Quality circles are started in Japan
1964	Ishikawa publishes book on Quality management
1970	Iskiawa publishes the book on the basics of quality circles and the concept of Total Quality is affirmed and devised in Japanese industries
1970 to 1980	Just – in –Time and quality become crucial for competitiveness. A large number of US and European corporations are beginning to appreciate the advance of Japan's industries. Taguchi popularizes the use of environmental design to design robust systems and products
1980+	Facing the rising sun challenge in quality management Development and introduction of FMSs and greater dependence on supplier contracts. Growth of economic based on quality control, information software packages
1990+	The management of quality has become a necessity that is recognized at all levels of management Increasing importance is given to off line quality management for the design of robust manufacturing processes and products. The growth of process optimization

QUALITY – DEFINITION

1. Predictable degree of uniformity and dependability at low cost and suited to the market -Deming
2. Fitness for use-Juran
3. Conformance to requirements - Crosby
4. Minimum loss imparted by a product to society from the time the product is shipped - Taguchi
5. A way of managing tile organization -Feigenbaum
6. Correcting and preventing loss, not living with loss - Hosffin .
7. The totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs – ISO

QUANITIFICATION OF QUALITY:

$$Q = P/E$$

P = Performance E = Expectations

DIMENSIONS OF QUALITY:

Dimension Meaning and Example

Performance - Primary product characteristics, such as the brightness of the picture

Features - Secondary characteristics, added features, such as remote control

Conformance - Meeting specifications or industry standards, workmanship

Reliability - Consistency of performance over time, average time of the unit to fail

Durability - Useful life, includes repair

Service - Resolution of problems and complaints, ease of repair

Response - Human – to – human interface, such as the courtesy of the dealer

Aesthetics - Sensory characteristics, such as exterior finish

Reputation - Past performance and other intangibles, such as being ranked first

----

QUALITY PLANNING

The following are the important steps for quality planning.

1. Establishing quality goals.
2. Identifying customers.
3. Discovering customer needs.
4. Developing product features.
5. Developing process features.
6. Establishing process controls and transferring to operations.

IMPORTANT POINTS TO BE NOTED WHILE QUALITY PLANNING

1. Business, having larger market share and better quality, earn returns much higher than their competitors.
2. Quality and Market share each has a strong separate relationship to profitably.

3. Planning for product quality must be based on meeting customer needs, not just meeting product specifications.
4. For same products. We need to plan for perfection. For other products, we need to plan for value.

QUALITY COSTS

1. PREVENTION COST

¬Marketing / Customer / User.

¬Product / Service / Design Development.

¬Purchasing

¬Operations (Manufacturing or Service)

¬Quality Administration.

2. APPRAISAL COST

¬Purchasing Appraisal Costs.

¬Operations Appraisal Costs

¬External Appraisal Costs

¬Review of Test and Inspection Data

¬Miscellaneous Quality Evaluations

3. INTERNAL FAILURE COST

¬Product or Service Design Failure Costs (Internal)

¬Purchasing Failure Costs

¬Operations (Product or Service) Failure Costs

4. EXTERNAL FAILURE COST

¬Complaint Investigations of Customer or User Service

¬Returned Goods

¬Retrofit and Recall Costs

¬Warranty Claims

¬Liability Costs

¬Penalties

¬Customer or User Goodwill

¬Lost Sales

Why TQM:

1. A question of survival in the intense competitive environment
2. Increasing customer consciousness

DEFINITIONS OF TQM:

1. TQM is the management approach of an organization, centered on quality, based on me participation of all its members and aiming at long-term success through customer satisfaction. and benefits to all members of me organization and to society.- **ISO**

2. TQM is an integrated organizational approach in delighting customers (both internal and external) by meeting their expectations on a continuous basis through every one involved with the organization working on continuous improvement in all products, services, and processes along with proper problem solving methodology - **INDIAN STATISTICAL INSTITUTE (ISI)**

3. TQM is a. people - focused management system that aims at continual increase in customer satisfaction at continually lower cost. TQM is a total system approach (not a separate area of program), and an integral part of high level strategy. It works horizontally across functions and departments, involving all employees, top to bottom, and exceeds backwards and forward to include the supply chain and the customer chain – **TOTAL QUALITY FORUM OF USA**

CHARACTERISTICS

1. Customer Oriented
2. Long term commitment for continuous improvement of all process
3. Team work
4. Continuous involvement of top management
5. Continuous improving at all levels and all areas of responsibility

BASIC CONCEPTS OF TQM:

1. Top management commitment
2. Focus on the customer – Both internal and external
3. Effective involvement and utilization of entire work force
4. Continuous improvement
5. Treating suppliers as partners
6. Establishing performance measures for the processes

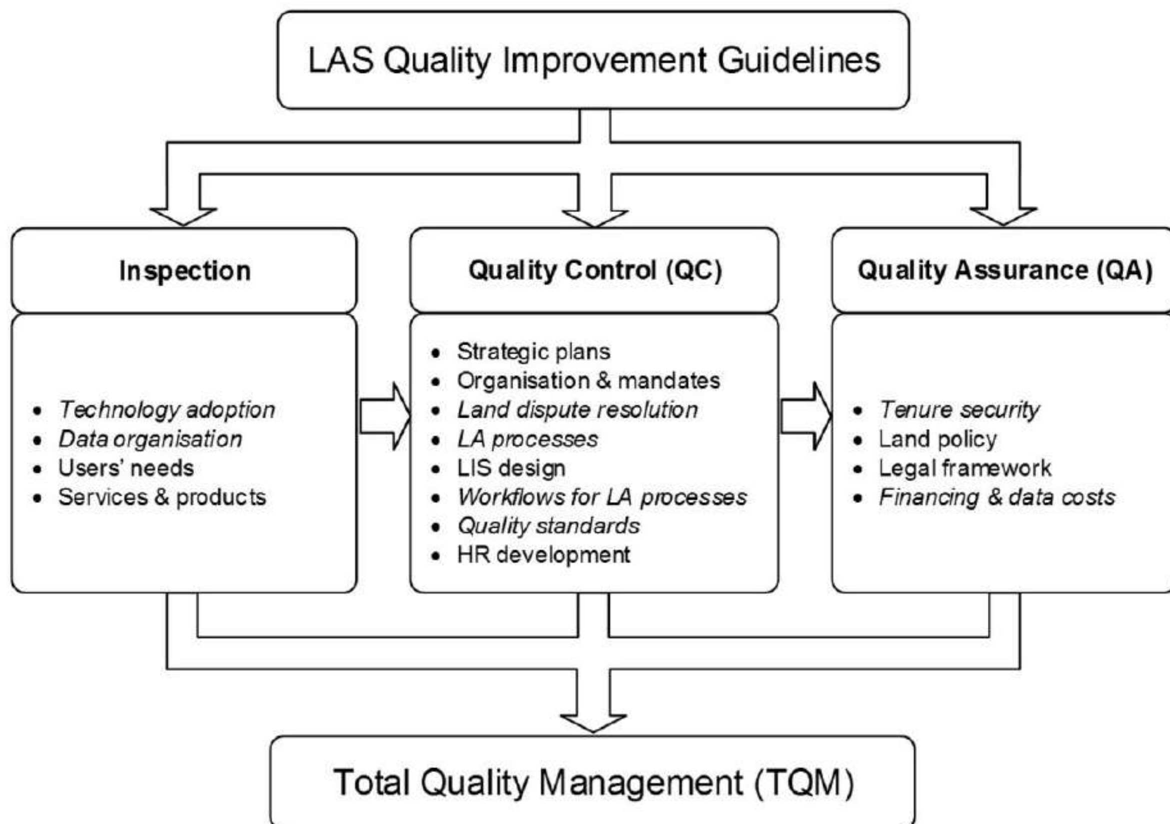
PRINCIPLES OF TQM:

1. Customers requirements - (both internal & external) must be met first time & every time
2. Everybody must be involved
3. Regular two way communication must be promoted I
4. Identify the training needs and supply it to the employees
5. Top management commitment is must
6. Every job must add value
7. Eliminate waste & reduce total cost
8. Promote creativity
9. Focus on team work.

NEW AND OLD CULTURES :

Quality Element	Previous State	TQM
Definition	Product Oriented	Customer Oriented
Priorities and cost	Second to service	First among equals of service and cost
Decisions	Short term	Long term
Emphasis	Detection	Prevention
Errors	Operations	System
Responsibility	Quality control	Everyone
Problem Solving	Managers	Teams
Procurement	Price	Life cycle costs, Partnership
Manager‟s Role	Plan, assign, control and enforce	Delegate, coach, facilitate and mentor

TQM FRAME WORK:



GURUS OF TQM :

SHEWHART	- Control chart theory
	- PDCA Cycle
DEMING	- Statistical Process Control
JURAN	- Concepts of SHEWHART - Return on Investment (ROI)
FEIGANBAUM	- Total Quality Control
	- Management involvement
	- Employee involvement
	- Company wide quality control
ISHIKAWA	- Cause and Effect Diagram
	- Quality Circle concept
CROSBY	- "Quality is Free"
	- Conformance to requirements
TAGUCHI	- Loss Function concept
	- Design of Experiments

OBSTACLES IN IMPLEMENTING TQM :

¬Lack of Management Commitment

¬Inability to change Organizational culture

¬Improper planning

¬Lack of continuous training and education

¬Incompatible organizational structure and isolated individuals and departments

¬Ineffective measurement techniques and lack of access to data and results

¬Paying inadequate attention to internal and external customers

¬Inadequate use of empowerment and teamwork

¬Failure to continually improve

BENEFITS OF TQM :

¬Improved quality

¬Employee participation

¬Team work

¬Working relationships

¬Customer satisfaction

¬Employee satisfaction

¬Productivity

¬Communication

¬Profitability

¬Market share

CONTRIBUTIONS OF QUALITY GURUS:

I. W.EDWARDS DEMING

Deming, an American, was the senior quality guru.

1928 – awardede doctorate in mathematical physics.

1946 – after sharing his expertise in statistical quality control to help the US war effort during world war II, the war department sent Deming to Japan to help nation recover from its wartime losses.

1951 – after having impressed by his contributions, the Japanese established the Deming prize. Deming prize is awarded annually to firms that distinguish themselves

with quality management programs.
1956 – awarded the shewhart medal by the American society of quality control.
1960 – awarded by the Japanese emperor with the second order of the Sacred Treasure for his teachings.
Deming was a prominent consultant, teacher and author on the subject of quality. He has published more than 200 works, including well-known books 'Quality' , 'Productivity and Competitive Position' , and 'Out of the Crisis'.

DEMING'S CONTRIBUTIONS:

Deming's contributions can be grouped under the following four topics:

1. Deming's 14 points on route to Quality.
2. Deming cycle (or PDCA Cycle)
3. Seven deadly diseases of Management
4. System of profound knowledge.

1. DEMING'S 14 POINTS:

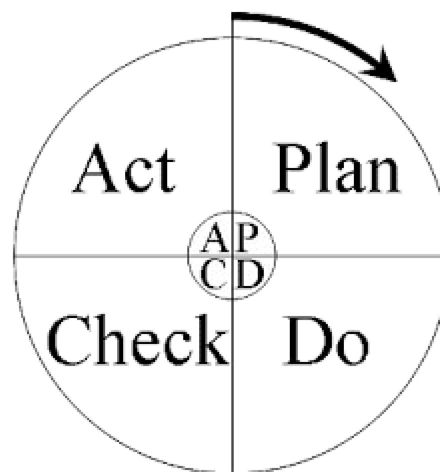
Deming Compiled a famous list of 14 points, which he believed were the prescription needed to achieve quality in an organisation.

1. Create constancy of purpose toward improvement of product and service.
2. Adopt the new philosopphy.
3. Cease dependence on inspection to achieve quality.
4. End the practice of awarding business on the basis of prize tag.
5. Improve constantly and forever the system of production and service.
6. Institute Training.
7. Institute Leadership.
8. Drive out fear.
9. Break down barriers between departments.
10. Eliminate slogans,exhortations, and targets for the work force.
11. Eliminate work standards (quotas) on the factory floor.
12. Remove barriers to pride of workmanship.

13. Institute a vigorous program of education and self-improvement.
14. Ccomplish the transformation.

2. DEMING CYCLE (or PDCA CYCLE):

Deming encouraged a systematic approach to problem solving and promoted the widely known plan. Do, check, Act (PDCA) cycle. The PDCA cycle is also known as the Deming cycle or Deming wheel, although it was developed by a colleague of Deming, Dr.Walter A.Shewhart.



It is an universal improvement methodology, the idea is to constantly improve, and thereby reduce the difference between the requirements of the customer and the performance of the process.

The cycle is about learning and ongoing improvement, learning what works and what does not in a systematic way: and the cycle repeats: after one cycle is complete, another is started.

3. SEVEN DEADLY DISEASES OF WESTERN MANAGEMENT:

The implementation of Demng's 14 points can transform the western style of management. This transformation can fully materialize only when certain bad practices, called by Deming unforgivable sins od deadly diseases (DD), are eliminated.

1. Lack of consistency of purpose.

2. Emphasis on short term profits.
3. Reliance on performance appraisal and merits.
4. Staff mobility.
5. Reliance on financial figures.
6. Excessive medical costs.
7. Excessive legal costs.

4. SYSTEM OF PROFOUND KNOWLEDGE:

The system of profound knowledge, or management by positive co-operation, is described by Deming. The four ingredients (or elements) of the system of profound knowledge that is necessary to learn and practice are given below.

1. Appreciation for a system.
2. Knowledge of statistical theory.
3. Theory of knowledge.
4. Knowledge of psychology.

II. JOSEPH M.JURAN

Juran born in Romania (1904) and emigrated to America in 1912.

1951 – published "Quality control Handbook".

Mid 1950's – like Deming, travelled to Japan to conduct top and middle level executive seminars on planning, organisational issues, management responsibilities

for quality and the need to set and monitor improvement target goals.
Juran has authored hundreds of papers and 12 books, including "Juran's Quality control Handbook", "Quality planning and Analysis", and "Juran on leadership for quality".
Juran has been awarded over 30 medals and fellowships world wide, including Japan's highest honour for a non-Japaneses citizen, the second order of the Sacred treasure presented by the Emperor.

JURAN'S CONTRIBUTIONS:

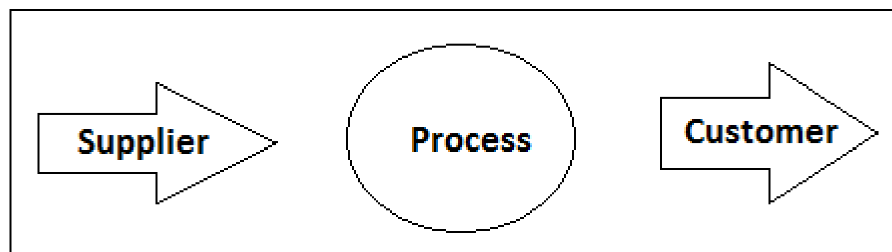
Juran's contributions can be studied under the following six topics:

1. Internal Customer
2. Cost of quality
3. Quality trilogy
4. Juran's 10 steps for quality improvement
5. The breakthrough concept.

1.INTERNAL CUSTOMER:

Juran realised that the customer was not just the end customer and that each person along the chain has an internal customer. Each person along the chain, fro product designer to final user, is a supplier and a customer.

In addition, the person will be a process, carrying out some transformation or activity. Therefore Juran maintained that at each stage was a "three role model". Supplier, Process and Customer.



2.COST OF QUALITY:

Juran classified the cost of quality into three classes as:

(i) **Failuer costs:** Scrap, rework, corrective actions, warranty claims, customer complaints and loss of customer.

(ii) **Appraisl Costs:** Inspection, compilance auditing and investigations.

(iii) **Prevention costs:** Training, Preventive auditing and process improvement implementation

Juran demonstrated the potential for increased profits that would result if the costs of poor quality could be reduced.

3.JURAN'S QUALITY TRILOGY:

Juran views quality as fitness-for-use. He also believes that roughly 80% of quality defects are managemnet controllable. Thus management has the responsibility to correct this deficiency.

Juran divides quality management into three parts. They are given below.

1. Quality Planning
2. Quality Control
3. Quality improvement

According to Juran, quality planning is necessary to establish process that are capable oof meeting quality standards. Quality control is necessary to know when the corrective action is needed and the quality improvement will help to find better ways of doing things.

The entire concepts of Juran's quality trilogy can be summarised as below.

1.Quality Planning:

- Identify the customers
- Determinr the customer's needs
- Develop product features
- Establish quality goals
- Develop a process
- Prove Process capability

2.Quality Control:

- Choose control subjects
- Choose units of measurement
- Establish measurement
- Establish standards of performance
- Measure actual performance
- Interpret the difference
- Take action on the difference

3.Quality Improvement:

- Prove need for improvement
- Identify specific projects for improvement
- Organise to guide the projects
- Organise for diagnosis – for discovery of causes
- Diagnose to find the causes
- Provide remedies
- Prove that remedies are effective under the operating conditions
- Provide for control to hold gains

4. JURAN'S 10 STEPS FOR QUALITY IMPROVEMENT:

1. Build awareness of the need and opportunity for improvement.
2. Set goals for improvement.
3. Organise to reach the goals.
4. Provide training.
5. Carryout projects to solve problems.
6. Report Progress.
7. Give Recognition.
8. Communicate results.
9. Keep score.

10. Maintain momentum by making annual improvement part of the regular systems and processes of the company.

5. THE BREAKTHROUGH CONCEPT:

Like the Deming cycle, Juran's breakthrough concerns itself with the product/service life cycle. In essence, it splits it up into two areas: the "journey from symptom to cause" and the "journey from cause to remedy".

III. PHILIP CROSBY

Crosby is another of the American quality gurus who rose to international frame mainly thanks to his teachings on quality management. He is best known for the concepts of 'Zero Defects' and 'Do it right first time'.

He has authored many books, including "Quality is free", "Quality without Tears", and "Let's talk Quality and Leading: The art of being an executive".
He was the founder and chairman of the board of career IV, an executive management consulting firm. He also founded Philip Crosby associates Inc. And the Quality College.

CROSBY CONTRIBUTIONS:

Crosby is known for his following contributions:

1. Four absolutes of quality.
2. Fourteen steps to quality management.
3. Croby's Quality vaccine.

1. Crosby Absolutes for Quality Management

First Absolute: The definition of quality is conformance to requirements, not goodness.

Second Absolute: The system for causing quality is preventive, not appraisal.

Third Absolute: The performance measure must be zero defect, not "that's close enough".

Fourth Absolute: The measurement of quality is the price of non-conformance, not indexes.

2. CROSBY'S FOURTEEN STEPS FOR QUALITY IMPROVEMENT:

Crosby has laid down 14 steps for implementing the quality improvement process in an organisation.

Step 1: Establish and ensure management commitment.

Step 2: Form quality improvement teams (QITs) for quality improvement process planning and administration.

Step 3: Establish quality measurements.

Step 4: Evaluate the cost of quality and explain its use as a management tool to measure waste.

Step 5: Raise quality awareness among all employees.

Step 6: Take actions to correct problems identified through previous steps.

Step 7: Establish a zero defects committee and programme.

Step 8: Train supervisors and managers on their role and responsibilities in the quality improvement process.

Step 9: Hold a zero defects day to reaffirm management commitment.

Step 10: Encourage individuals and groups to set improvement goals.

Step 11: Obstacle Reporting.

Step 12: Recognise and appreciate all participants..

Step 13: Establish quality councils to discuss quality matters on a regular basis.

Step 14: Do it all over again to demonstrate that the improvement process never ends.

3.Crosby's Quality Vaccine:

In the crosby style, the "vaccine" is explained as medicine for management to prevent poor quality. The five sections of vaccine that cover the reqiurements of total quality management are as follows.

1. Integrity
2. Systems
3. Communication
4. Operations
5. Politics

BARRIERS TO TQM IMPLEMENTATION:

1. Lack of management commitment
2. Lack of faith in and support to TQM activities among management personnel
3. Failure to appreciate TQM as a cultural revolution. In other words, inability to change organizational culture
4. Misunderstanding about the concept of TQM
5. Improper planning
6. Lack of employees commitment
7. Lack of effective communication
8. Lack of continuous training and education
9. Lack of interest or incompetence of leaders
10. Ineffective measurement techniques and lack of access to data and results
11. Non-application of proper tools and techniques
12. Inadequate use of empowerment and team work

BENEFITS OF TQM

Tangible Benefits	Intangible Benefits
· Improved product quality · Improved productivity · Reduced quality costs · Increased market and customers	· Improved employee participation · Improved team work · Improved working relationships · Improved customer satisfaction

· Increased profitability · Reduced employee grievances	· Improved communication · Enhancement of job interest · Enhanced problem solving capacity · Better company image

QUALITY STATEMENTS

VISION STATEMENT:

It is a short declaration of what an organization aspires to be tomorrow. It is an ideal state which may never be achieved.

Example:

"To continuously enrich knowledge base of practioners in mobility industry and institutions in the service of humanity" - **SAE**

MISSION STATEMENT:

Describes the function of the organization. It provides the clear statement of purpose for the employees, customers and suppliers.

Example:

" Facilitating world class technical education through high quality institutions, academic excellence and innovative research and development programmes, technology forecasting and global manpower planning, promoting industry institute interaction, inculcating entrepreneurship" - **AICTE**

QUALITY POLICY STATEMENT:

It is a guide for everyone in the organization as to how they provide products and services to the customer. Written by the CEO feedback from workforce and approved by quality council.

Example:

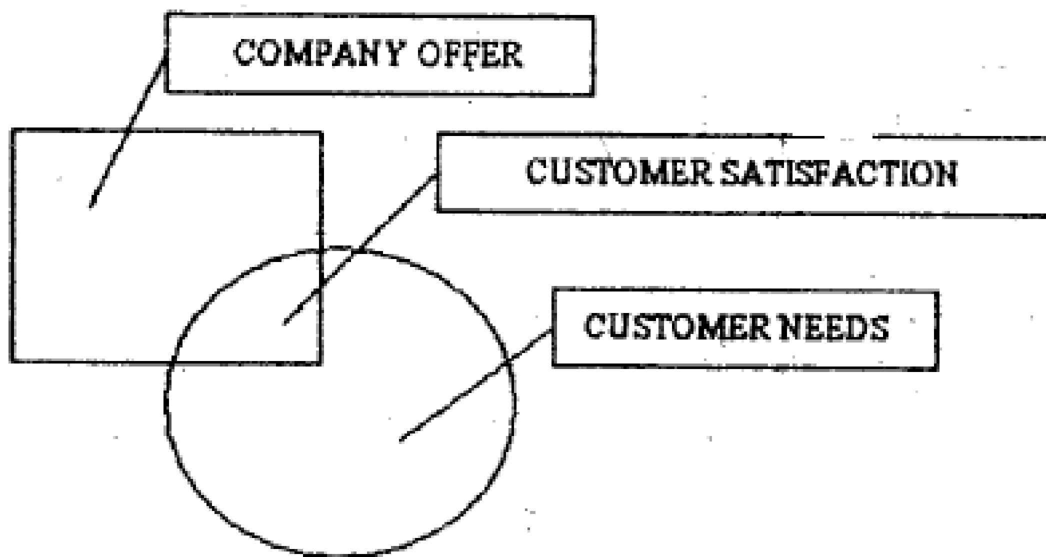
"Xerox is a quality company. Quality is the basic business principle for Xerox. Quality means providing our external and internal customers with innovative products and service that fully satisfy their requirements. Quality is the job of every employee" – **Xerox Corporation**

CUSTOMER SATISFACTION

The Customer is the King - Emphasized by Today's Buyers Market. TQM's Purpose is meeting or exceeding customer expectations, so that the customers are delighted. The customer satisfactions must be the primary goal of any organization.

CUSTOMER SATISFACTION MODEL

Teboul's Model of customer satisfaction as shown in figure



From the above diagram it is understood that the company should strive for increasing the intersection portion i.e. Customer Satisfaction.

THE CUSTOMERS ARE

· The most important people in the business
· Not dependent on the organization, but the organization depends on them.
· Not an interruption to work but are the purpose of it.
· Doing a favor when they seek business and not vice-versa.
· A part of business, not outsiders and they are life blood of the business
· People who come with their needs and jobs
· Deserve the most courteous and attentive treatment.

TYPES OF CUSTOMERS

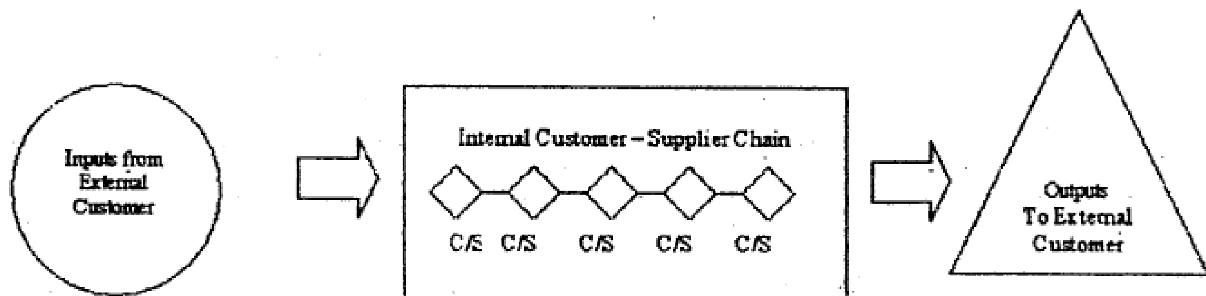
Internal Customer:

The customer inside the company are called internal customers

External Customers:

An external customer is the one who used the product or service or who purchase the products or service or who influences the sale of the product or service.

CUSTOMER SUPPLIER CHAIN



CUSTOMER COMPLAINTS (FEEDBACK)

Customer feedback must be continuously solicited and monitored to reduce the dissatisfied customers as much as possible.

CUSTOMER FEEDBACK OR CUSTOMER COMPLAINT IS REQUIRED

· To discover customer dissatisfaction

· To identify customer's needs

· To discover relative priorities of quality

· To compare performance with the competition

· To determine opportunities, for improvement

TOOLS USED FOR COLLECTING CUSTOMER COMPLAINTS

· **Comment card** - Low cost method, usually attached to warranty card

· **Questionnaire** - Popular tool, costly and time consuming - by mail or telephone preferably multiple choice questions or a point rating system (1 to 5) or (1 to 10)

Customer Focus groups - Meeting by a representative of the company with the group of customers. Imprint analysis is an emerging technique to obtain intrinsic feelings using customer meetings, word associations, discussion, relaxation techniques etc.

· **Phone -** Toll free Telephone numbers

· **Customer visits** - Visit customer's place of business.

· **Report cards** - Usually, send to customer on a quarterly basis.

· **The internet and computer** - It includes newsgroups, electronic bulletin board mailing lists,

· Employee feedback.

· **Mass Customization** - Capturing the voice of customers using data of what customer want instead of what customer is thinking about buying and manufacturing exact what they want.

STEPS TO SOLVE CUSTOMER COMPLAINTS

· Complaints can be collected from all sources (letters, phone -calls, meetings and verb inputs)

· Develop procedures for complaint resolution, that include empowering front-line personnel.

· Analyze complaints, but understand that complaints do not always fit into new categories

· Work to identify process and material variations and then eliminate the root cause.

· When a survey response is received, a senior manager should contact the customer and strive to resolve the concern.

· Establish customer satisfaction measures and constantly monitor them.

· Communicate complaint information, as well as the result of all investigation solution, to all people in the organization. .

· Provide a monthly complaint report to the quality council for their evaluation and

needed, the assignment of process improvement teams.
· Identify customer's expectations beforehand rather than afterward through complaint analysis.

CUSTOMER RETENTION

· More powerful and effective than customer satisfaction
· It is the process of retaining the existing customer
· Customer care can be defined as every activity which occurs within the organization that ensures that the customer is not only satisfied but also retained.

SIGNIFICANCE OF CUSTOMER RETENTION

· 60% of organizations future revenue will come from exiting customers
· 2% increase in customer retention has 10% decrease in operating cost.
· 96% of unhappy customers do not complain but 3 times likely to convey to other customers about their bad experience.
· 91% of unhappy customers never purchase goods and services from you.
· It costs 5 times more to attract the customer than retaining the existing customer.
· Customer retention creates customer loyalty and moves customer satisfaction to a next level called customer delight.

QUALITY COST

Quality costs are defined as those costs associated with the non-achievement of product or service quality as defined by the requirements established by the organization and its contracts with customer and society.

Quality costs:

Costs of control
Prevention costs
Appraisal costs
Costs of failure of control

Internal failure costs

External failure costs

Quality costs are broadly classified into:

1. Failure costs
2. Appraisal costs
3. Prevention costs

Failure costs:

Failure costs are the direct and indirect costs incurred on those products or services, which fail to comply with their prescribed specifications.

¬Internal failure costs – before delivered to customers

¬External failure costs – after delivered

Appraisal costs:

Appraisal costs are those costs involved in actual checking of the quality, viz., the cost of carrying out actual inspection.

Prevention costs:

Prevention costs are those, which are involved in ensuring that faulty or defective **work, or rejections are not produced in very first instance.**

UNIT II – TQM PRINCIPLES

Leadership – Strategic quality planning, Quality Councils - Employee involvement – Motivation, Empowerment, Team and Team work, Quality Circles - Recognition and Reward, Performance appraisal – Continuous process improvement – PDCA cycle, 5S, Kaizen – Supplier partnership – Partnering supplier selection – Supplier rating.

LEADERSHIP

A leader is one who instills purposes, not one who controls by brute force. He strengthens and inspires the followers to accomplish shared goals.

Leaders

¬Shape the Organization's value

¬Promote the Organization's value

¬Protect the Organization's value and

¬Exemplifies the Organization values

The process of influencing others towards the accomplishment of goals. He triggers tile will to do, shows the direction and guide the group members towards the accomplishment of goals.

CHARACTERISTICS OF QUALITY LEADERS :

1. They give priority attention to external and internal customers and their needs.
2. They empower, rather than control, subordinates.
3. They emphasis improvement rather than maintenance.
4. They emphasis prevention.
5. They emphasis collaboration rather than competition.
6. They train and coach, rather than direct and supervise.
7. They learn from the problems.
8. They continually try to improve communications.
9. They continually demonstrate their commitment to quality.

10.They choose suppliers on the basis of quality, not price.

11.They establish organizational systems to support the quality effort.

12.They encourage and recognize team effort.

LEADERSHIP CONCEPTS :

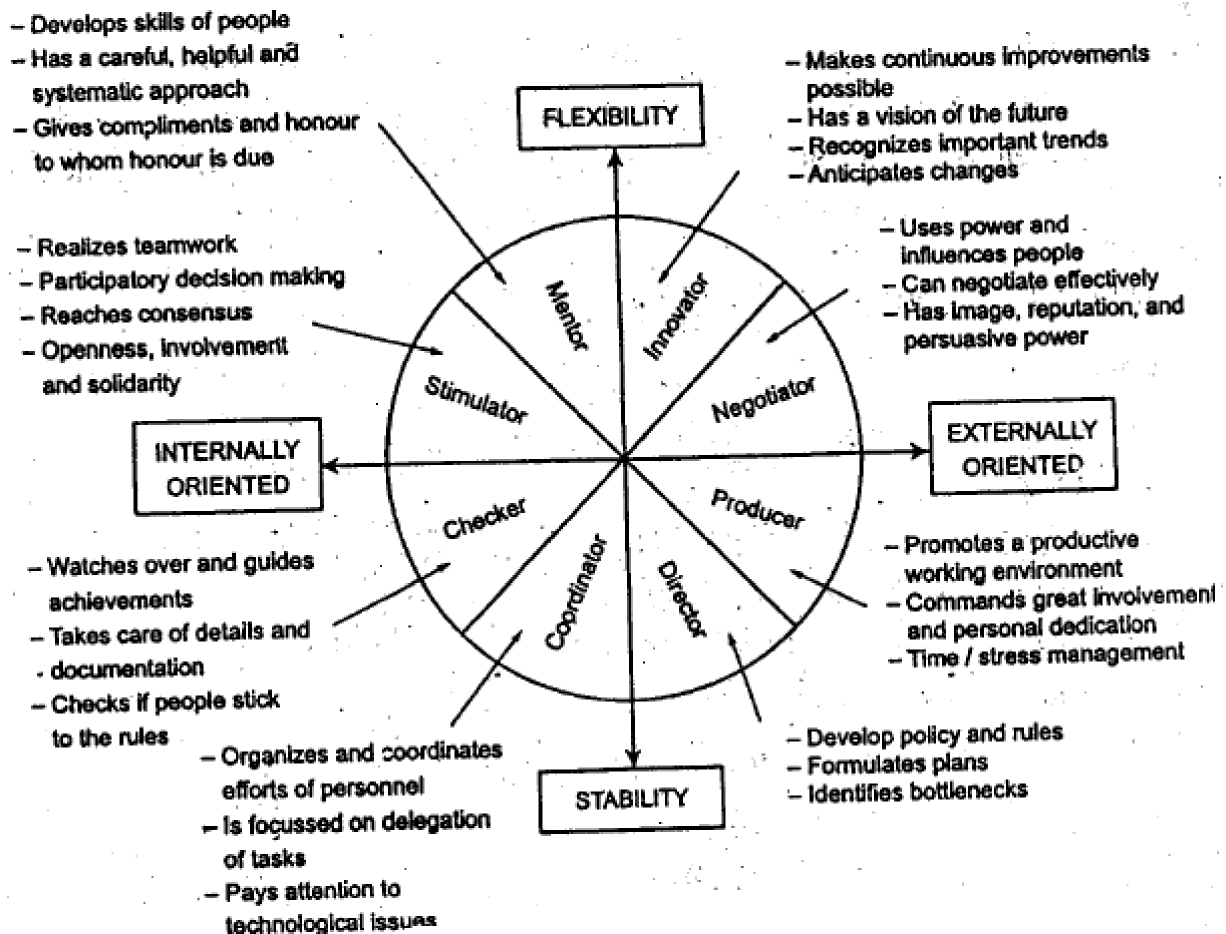
A leader should have the following concepts

1. People, Paradoxically, need security and independence at the same time.
2. People are sensitive to external and punishments and yet are also strongly self - motivated.
3. People like to hear a kind word of praise. Catch people doing something right, so you can pat them on the back.
4. People can process only a few facts at a time; thus, a leader needs to keep things simple.
5. People trust their gut reaction more than statistical data.
6. People distrust a leader's rhetoric if the words are inconsistent with the leader's actions.

THE 7 HABITS OF HIGHLY EFFECTIVE PEOPLE :

1. Be Proactive
2. Begin with the End in mind
3. Put First Things First
4. ThinkWin –Win
5. Seek First to Understand, then to Be Understood
6. Synergy
7. Sharpen the Saw (Renewal)

LEADERSHIP ROLES



ROLE OF SENIOR MANAGEMENT

1. Management byWandering Around (MBWA).
2. Strategy of problem solving and decision making.
3. Strong information base.
4. Recognition and Reward system.
5. Spending most of the time on Quality.
6. Communication.
7. Identify and encourage potential employee.
8. Accept the responsibility.
9. To play a role model.
10. Remove road blocks.
11. Study TQM and investigate how TQM is implemented elsewhere.
12. Establish policies related to TQM.

13.Establish 'priority of quality' and 'customer satisfaction' as the basic policy.

14.Assume leadership in bringing about a cultural change.

15.Check whether the quality improvement programmes are conducted as planned.

16.Become coaches and cheer leaders to implement TQM.

17.Generate enthusiasm for TQM activities.

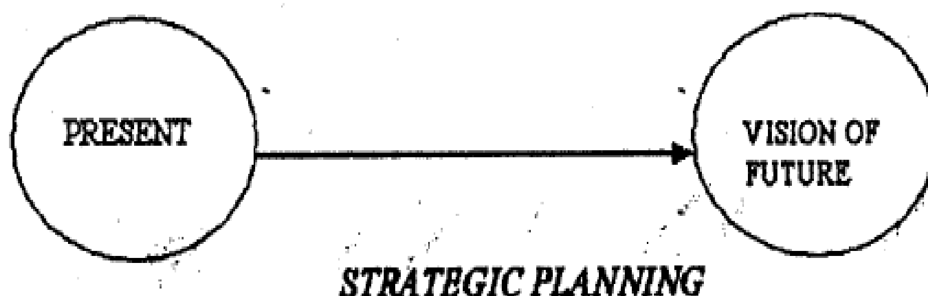
18.Visit other companies to observe TQM functioning.

19.Attend TQM training programme.

20.Teach others for the betterment of society and the surroundings.

STRATEGIC PLANNING

It sets the long term direction of the organization in which it wants to proceed in future. Can be defined "As the process of deciding on objectives of the organization, on changes on this objective, on the resource used to obtain these objectives and on the policies that are to govern the acquisition use and disposition of these resources"



STRATEGIC QUALITY PLANNING

Goals – Long term planning (Eg : Win the war)

Objectives – Short term planning (Eg : Capture the bridge)

Goals should

¬Improve customer satisfaction, employee satisfaction and process

¬Be based on statistical evidence

¬Be measurable

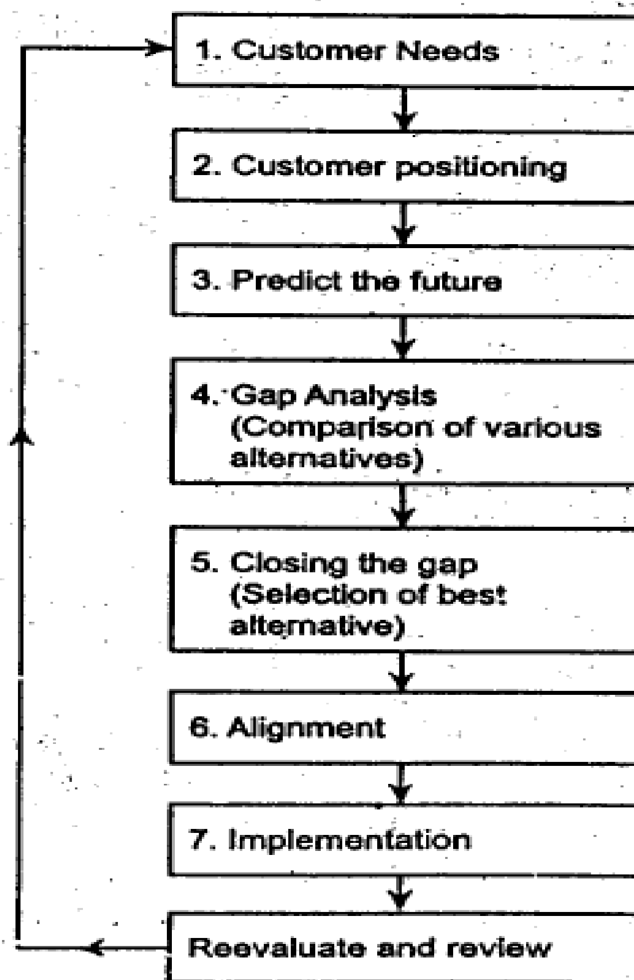
¬Have a plan or method for its achievement

¬Have a time frame for achieving the goal

¬Finally, it should be challenging yet achievable

SEVEN STEPS TO STRATEGIC QUALITY PLANNING :

1. Customer needs
2. Customer positioning
3. Predict the future
4. Gap analysis
5. Closing the gap
6. Alignment
7. Implementation



Strategic planning cycle

EMPLOYEE INVOLVEMENT

It is the total involvement from every person at all levels in the organization

ASPECTS OF EMPLOYEE INVOLVEMENT

1. Employee motivation
2. Employee Empowerment
3. Teams and Team work
4. Recognition and Reward Schemes
5. Performance Appraisal

1. EMPLOYEE MOTIVATION

It is the process of stimulating people or attempting to influence other to do your will or accomplish desire goals through the possibility of reward

· Improves employee involvement

· Reduces absenteeism and increases turn over

· Promotes job satisfaction

THEORIES OF MOTIVATION

HERZBERG'S TWO FACTOR THEORY

1. Motivation Factor: People are motivated by recognition, responsibility, achievement, advancement and the work itself. These are called as motivators
2. Dissatisfies or Hygiene Factor: Low salary, minimal fringe benefits, poor working conditions, ill defined organizational policy, mediocre technical supervision are dissatisfies which implies they are preventable.

EMPLOYEE WANTS

1. Good pay factor is normally in the middle of ranking.
2. Normal Wants are interesting work, appreciation, involvement job security, Good pay, Promotion/growth, Good working conditions, Loyalty to employees, Help with personal problems arid Tactful Discipline.

ACHIEVING A MOTIVATED WORK FORCE BY THE MANAGERS

Know thyself, Know employees, Establish a positive attitude, Share the goal, Monitor progress, Develop interesting work by job rotation, job enlargement(Horizontal) and job enrichment (Vertical), Communicate effectively, Celebrate success.

EMPLOYEE EMPOWERMENT

It is an environment in which people have the ability, the confidence and the commitment to take his responsibility and ownership to improve the process and initiate the necessary steps to satisfy customer requirements within well-defined boundaries in order to achieve organizational values and goals.

Job Enrichment: Is expanding content of the Job.

Job Empowerment: Is expanding the context of the job.

GENERAL PRINCIPLES OR CHARACTERISTICS FOR' EMPOWERING EMPLOYEES

1. Tell people what their responsibilities are.
2. Given the authority equal to the responsibility assigned to them.
3. Set standards of excellence.
4. Give them knowledge information and feed back.
5. Trust them and treat them with dignity and respect.

CONDITIONS TO CREATE THE EMPOWERED ENVIRONMENT

1. Every one should under stand the need to change
2. The system need to change to new paradigm.
3. The organization must provide information, education, and skill to its employees.

TEAMS AND TEAM WORKS

A team can be defined as a group of people working together to achieve common objectives or goals. Team work is the cumulative actions of the team during which each member of the team subordinates his individual interest and opinions for the fulfilling of objectives of the group.

BENEFITS OF TEAM WORK

Improved solutions to quality problems, ownership of solutions, communication and integration

Objectives – Short Term Planning

Goal – Long Term Planning

TYPES OF TEAMS

Process improvement team:

Involved in improvement of sub processes or processes.

Usually has 6-10 members. Disbanded when the objective is reached. May include the local supplied and customer depending on the location

Cross functional teams:

6-10 members temporary team. Members are Top management level from various functional areas of management. Discuss complex problems and break down into smaller parts to refer it to various departmental teams for further solution.

Natural work teams:

Not voluntary and the total work unit is part of the team. Manager also a part of the team and the management selects the projects to be improved. Managers must also ensure that the entire team is comfortable with each other.

Self directed / self managed work team:

Extension of natural work teams but here the group of individuals is empowered not only to do work but manage it. No manger will present but a coordinator (Which will be normally rotated among members) will be appointed. Additional responsibilities of the team hiring/ dismissal, performance evaluation, customer relations, supplier

relations, recognition/rewards and training.

CHARACTERISTICS OF SUCCESSFUL TEAMS

1. **Sponsor:** In order to have effective liaison with quality council, there should be sponsor. The sponsor is a person from the quality council, he is to provide support to the organization
2. **Team Charter:** A team charter is a document that defines the team's mission boundaries, the background of the problem, the team's authority and duties and resources. It also identifies the members and their assigned roles – leader, recorder, time keeper and facilitator.
3. **Team Composition:** Not exceeding 10 members except natural work team and self managed teams.
4. **Training:** The team members should be trained in the problem solving techniques team dynamics and communication skills
5. **Ground Rules:** The team should have separate rules of operation and conduct. Ground rules should be discussed with the members, whenever needed it should be reviewed and revised
6. **Clear objectives, Accountability :** Periodic status report should be submitted to quality council for review
7. **Well defined decision procedure, Resources:** Adequate information should be provided
8. **Trust by the management, Effective problems solving**: Not by hunches or quick fires
9. **Open communication, Appropriate Leadership, Balanced participation and Cohesiveness**

TEAM MEMBER ROLE

Leader, Facilitator (One who helps the team gets started in the stages), Recorder, Time keeper and Team member.

DECISION MAKING METHODS

Non decision, Unilateral decision, Handclasp decision (Two members with a good idea of the subject decide), Minority-rule decision, Majority rule decision, Consensus (Not everyone need to accept, But every one should be willing to implement)

ELEMENTS OF EFFECTIVE TEAM WORK

Regular scheduling with a fixed time limit, purpose, role and responsibilities, activities, decision, results and recognition.

TEAM MANAGEMENT WHEEL

To make a lean more effective a team management wheel has been evolved. The activities are advising, innovating, promoting, developing, organization, producing, inspecting, maintaining and linking. The roles of wheel are advisor, explore, organizer and controller.

STAGES OF TEAM DEVELOPMENT

Forming stage- Initial stage with only group of individuals and no team work. Team purpose, roles are created.

Storming Stage -Initial agreement roles are challenged. Hostilities, emerge which may be resolved

Norming Stage-Formal informal relations get established.

Performing Stage -Team operates in a successful manner with trust, openness, healthy conflict and decisiveness among the members.

Maintenance stage – Functioning should not deteriorate with time Q

Evaluating Stage – Evaluating team performance

TEN COMMON PEOPLE PROBLEMS

Floundering, overbearing participants, Dominating participants, reluctant participants, unquestioned acceptance of opinions as facts, rush to accomplish, Attribution, Discounts and plops, Wanderlust, Feuding team members.

BARRIERS TO TEAM PROGRESS

Insufficient training, Incompatible rewards and compensation, First-line supervisor resistance, Lack of planning; Lack of management support, Access to information systems, Lack of union support, Project scope too large, Project objectives are not significant, No clear measures of success and No time to do improvement work.

RECOGNITION AND REWARD

Recognition is a process whereby management shows acknowledgement (Verbal or written) of an employee outstanding performance. Recognition is a form of employee +ve motivation.

Reward is a tangible one such as increased salaries, commission, cash bonus, gain sharing etc., to promote desirable behavior. It can be even theatre tickets, dinner for two, a small cash awards, etc.,

The employees are recognized to improve their morale, show the company's appreciation for Better Performance, create satisfied and motivated workplace and stimulate creative efforts.

Various ways for Recognition and Rewards are

1. Recognition can be expressed using verbal and written praise.
2. Rewards may be in the form of certificates and plaques.
3. Reward is normally in the form of cinema tickets, dinner for family etc.
4. The financial compensation (for recognition) can be paid in terms of increased salaries, commissions, gain sharing etc.
5. The efforts of employees can be recognized by promotions, special job assignments etc.
6. A letter of appreciation from the CEO or the Top Management will increase the employee's involvement.
7. Reward may be delayed but recognition should be in a timely basis.
8. Rewards should be appropriate to the improvement level.
9. People like to be recognized than any reward.
10.Special forms of recognition include pictures on the bulletin board, articles in

news letters, letter to families etc.

11.Supervisors can give on-the-spot praise for a job which is done well.

EFFECTS OF RECOGNITION AND REWARD SYSTEM :

1. Recognition and reward go together for letting people know that they are valuable members for the organization.
2. Employee involvement can be achieved by recognition and reward system.
3. Recognition and reward system reveals that the organization considers quality and productivity as important.
4. It provides the organization an opportunity to thank high achievers.
5. It provides employees a specific goal to achieve.
6. It motivates employees to improve the process.
7. It increases the morale of the workers.

PERFORMANCE APPRAISAL

The performance appraisal is used to let employees know how they are performing. The performance appraisal becomes a basis for promotions, increase in salaries, counseling and other purposes related to an employee's future.

IMPORTANCE OF PERFORMANCE APPRAISALS :

1. It is necessary to prevail a good relationship between the employee and the appraiser.
2. Employee should be informed about how they are performing on a continuous basis, not just at appraisal time.
3. The appraisal should highlight strength and weakness and how to improve the performance.
4. Employee should be allowed to comment on the evaluation and protest if necessary.
5. Everyone should understand that the purpose of performance appraisal is to have employee involvement.
6. Errors in performance evaluations should be avoided.

7. Unfair and biased evaluation will render poor rating and hence should be eliminated.

BENEFITS OF EMPLOYEE INVOLVEMENT :

Employee involvement improves quality and increases productivity because

¬Employees make better decisions using their expert knowledge of the process

¬Employees are better able to spot and pin-point areas for improvement.

¬Employees are better able to take immediate corrective action.

¬Employee involvement reduces labour / management friction.

¬Employee involvement increases morale.

¬Employees have an increased commitment to goals because they are involved.

CONTINUOUS PROCESS IMPROVEMENT

Continuous process improvement is designed to utilize the resources of the organization to achieve a quality-driven culture.

Improvement is made by

¬Viewing all work as process.

¬Making all process effective, efficient and adaptable.

¬Anticipating changing customer needs.

¬Controlling in-process performance using measures such as scrap reduction, control charts etc.

¬Eliminating waste and re-work.

¬Eliminating non-value added activities.

¬Eliminating non-conformities.

¬Using Benchmarking.

¬Incorporating learned lessons into future activities.

¬Using technical tools such as SPC, benchmarking, experimental design, QFD etc.

PROCESS :

Process refers to business and production activities of an organization.

There are five basic ways for improvement.

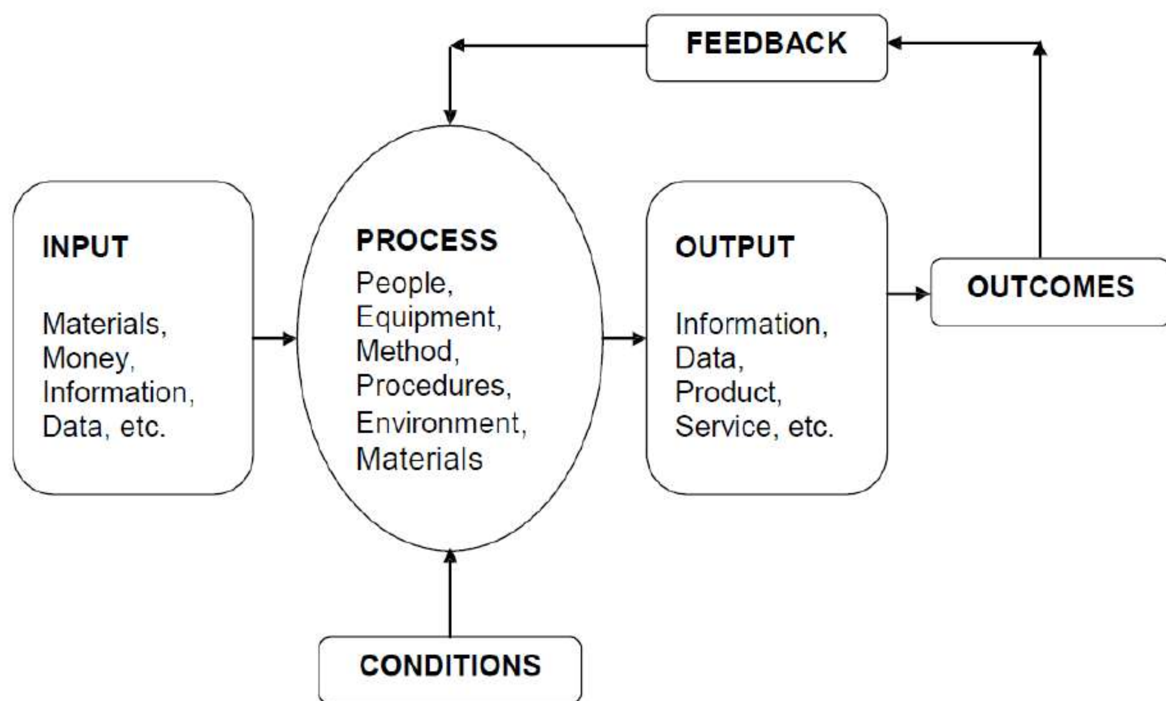
¬Reduce resources.

¬Reduce errors.

¬Meet or exceed expectations of downstream customers.

¬Make the process safer.

¬Make the process more satisfying to the person doing it.



THE JURAN TRILOGY

1. PLANNING

¬Determine internal & external customers.

¬Their needs are discovered.

¬Develop product / service features.

¬Develop the processes able to produce the product / service features.

¬Transfer plans to operations.

2. CONTROL

Control is used by operating forces to help meet the product, process and service requirements.

It consists of the following steps

1. Determine items to be controlled.
2. Set goals for the controls.
3. Measure actual performance.
4. Compare actual performance to goals.
5. Act on the difference.

3. IMPROVEMENT

Aims to attain levels of performance that are higher than current levels.

It consists of the following steps

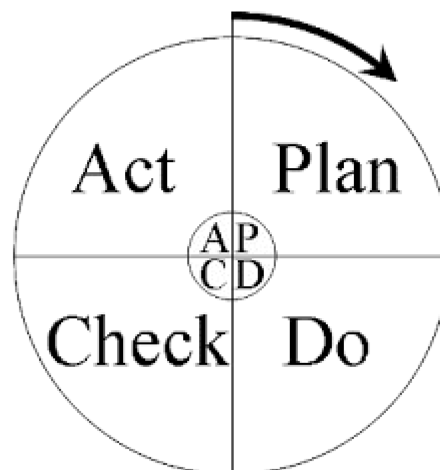
¬Establishment of quality council.

¬Identify the improvement projects.

¬Establish the project teams with a project leader.

¬Provide the team with the resources.

THE PDCA CYCLE :



PROBLEM SOLVING METHOD :

1. IDENTIFY THE OPPORTUNITY

¬Identify the Problem

⟦Pareto analysis of external alarm signals.

⟦Pareto analysis of internal alarm signals.

⟦Proposals from key insiders.

⟦Proposals from suggestion schemes.

⟦Field study of user's needs.

⟦Comments of key people outside the organization.

⟦Customer surveys.

⟦Employee surveys.

⟦Brainstorming by work groups.

¬Form the Team

⟦Team should be selected.

⟦Goals and milestones are established.

¬Define the Scope.

Criteria for a good problem statement is as follows

⟦It clearly describes the problem.

⟦It states the effect.

⟦It focuses on what is known, unknown etc.

⟦It emphasizes the impact on the customer.

2. ANALYZE THE CURRENT PROCESS

The objective is to understand the process and how it is currently performed.

Step 1 : The team to develop a process flow diagram.

Step 2 : The target performance measures are defined.

Step 3 : Collection of all available data and information.

Common items of data and information are

1. Customer information 2. Design information

3. Process information 4. Statistical information

5. Quality information 6. Supplier information

3. DEVELOP THE OPTIMAL SOLUTION(S)

This phase has the objective of establishing potential and feasible solutions and recommending the best solution to improve the process.

¬Creativity plays the major role, and brainstorming is the principal technique.

¬There are three types of creativity:

⟦Create new processes

⟦Combine different processes

⟦Modify the existing process

4. IMPLEMENT CHANGES

This phase has the objective of preparing the implementation plan, obtaining approval and implementing the process improvements.

¬Approval of the quality council.

¬Obtain the advice and consent of departments, functional areas, teams, individuals etc.

¬Monitor the activity.

5. STUDY THE RESULTS

This phase has the objective of monitoring and evaluating the change by tracking and studying the effectiveness of the improvement efforts.

6. STANDARDIZE THE SOLUTION

¬Institutionalize by positive control of the process.

¬The quality peripherals – the system, environment and supervision must be certified.

¬Operators must be certified.

7. PLAN FOR THE FUTURE

The objective is to achieve improved level of process performance.

¬Regularly conduct reviews of progress by the quality council.

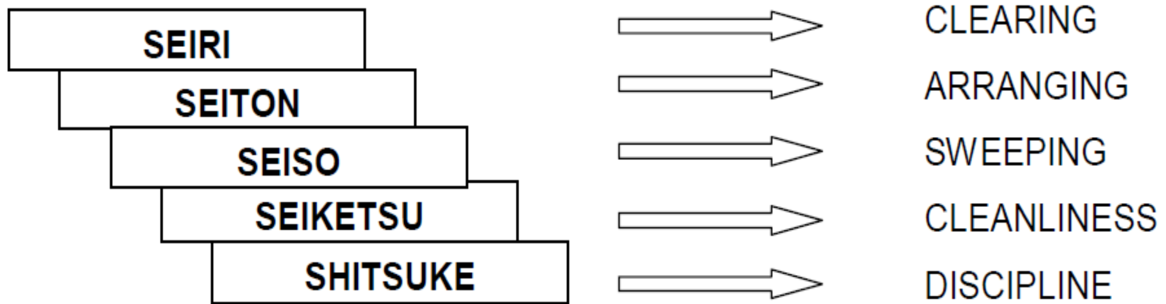
¬Establish the systems to identify area for future improvements.

¬Track performance with respective internal & external customers.

¬TQM tools and techniques are used to improve quality, delivery and cost.

5-S : HOUSEKEEPING

5-S MEANS EVERYTHING IN ITS PLACE



¬There can be no TQM without 5-S.

¬A dirty factory cannot produce quality products.

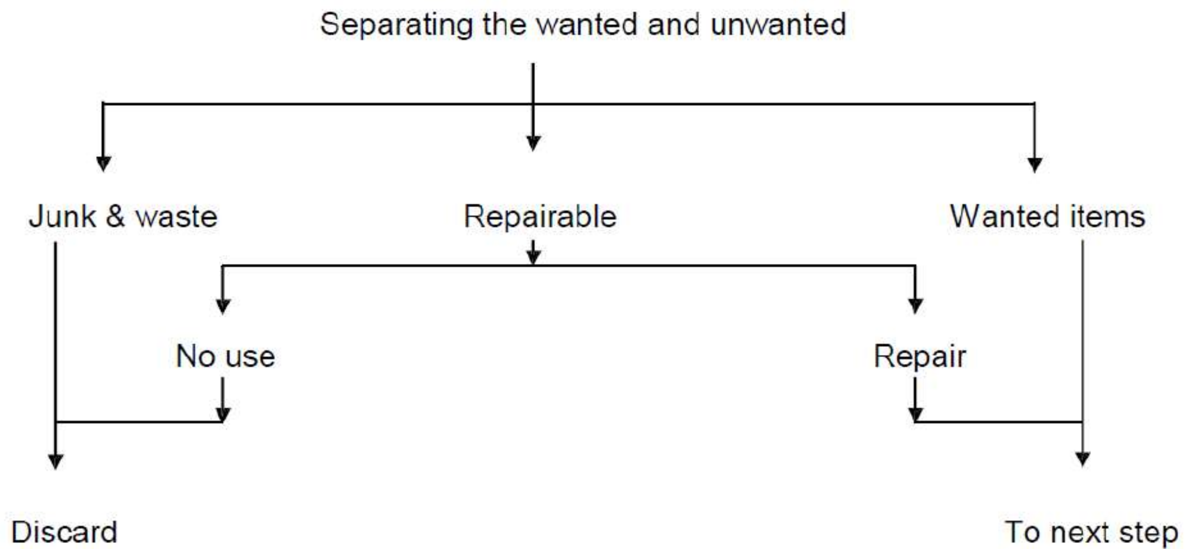
¬Clutter hides problems. A neat workplace promotes easy discovery abnormalities.

The First S : SEIRI : CLEARING

Take out unnecessary items and throw them away

Factory Floor	**Office**	**Home**
] Machines to be scrapped	] Used / Broken pens	] Broken toys
] Rejected material	] Useless paper	] Old clothes
] Expired goods	] Old diaries	] Broken suitcases
] Broken tools, pallets, bins, trolleys.	] Broken furnitures	
] Old notices		

Flow Chart :



Consequences of not practicing SEIRI :

¬The unwanted clutters up the place and the wanted are hard to find.

¬Every place can only hold so much.

¬Clutter sometimes causes misidentification.

The Second S : SEITON : ARRANGING

Arrange everything in proper order so that it can be easily picked up for use.

Factory Floor	Office	Home
] Unlabelled tool crib	] Unlabelled file cabinet	] Clutter
] Cluttered shelves lockers etc.	] Cluttered drawer,shelves, book cases,tables	] No orderly arrangement in the rooms
] Stores – no clear location system.	] Records & documents Not arranged well	
] Things on the floor	] File heaps and papers	

Consequences of not practicing SEITON :

¬Things are seldom available when needed.

¬Items are “lost’ in stores.

¬Items – defectives and good ones get mixed up.

¬Accidents or near-accidents occur due to clutter.

¬Visual control of the shop floor is not possible.

¬Sometimes, production is lost because an item required is available but cannot be found.

¬In some offices, Critical Excise records or tax records may not be traceable. This can lead to finance loss, prosecution or embarrassment.

The Third S : SEISO : SWEEPING

Sweep your workplace thoroughly so that there is no dust anywhere.

Factory Floor	**Office**	**Home**
⟧Dirty machines	⟧Dirty table & furniture	⟧Dirty furniture, floor, window, grills,bookshelves.
⟧Dust on product parts, R.Mtls.		
⟧Dirty jigs, fixtures	⟧Dirty office equipments	
⟧Dirty walls, roofs	⟧Littered floor	
⟧Littered floor	⟧Dirty windows	

Consequences of not practicing SEISO :

¬Most machines are affected by dust & dirt and hence their performance may go down.

¬Dust and dirt on products, materials, packing boxes etc. will affect either their performance quality or their aesthetic look.

¬Unpleasant to work in.

The Fourth S : SEIKETSU : CLEANLINESS

Washing with a strong overtone of keeping things disinfected as well as free of hazardous chemicals.

Factory Floor	**Office**	**Home**
Handling hazardous chemicals	Free of pests	Pest control
Control of fumes, hazardous dust.	Personal hygiene	Personal hygiene
Disinfecting, Personal hygiene		

Consequences of not practicing SEIKETSU :

¬Good health and safety require the practice of Seiketsu.

¬Hazardous chemicals, dusty chemicals, fumes etc. can make it a dangerous place to work in.

¬Washing thoroughly and cleaning a place makes the workplace pleasant.

The Fifth S : SHITSUKI : DISCIPLINE

Discipline especially with regard to safety rules and punctuality.

Consequences of not practicing SEIKETSU :

¬If discipline is not practiced, then the first 4-S would backslide.

¬Lack of Shitsuki means not following the standards. Then, all activities related to safety and quality will be affected.

IMPLEMENTING 5-S

1. Top Management resolve and training.
2. Formation of a top level team.
3. Understanding current circumstances.
4. Establishing priorities and targets.

5. Forming sub-teams and training.
6. Major cleaning.
7. Establishing improvement plans in each priority area.
8. Implementing the plan.
9. Verifying results.
10. Standardizing.
11. Establishing full control.
12. Looking for further improvements.

KAIZEN

Kaizen is a Japanese word for the philosophy that defines management's roles in continuously encouraging and implementing small improvements involving everyone.

It focuses on simplification by breaking down complex progress into their sub – processes and then improving them.

The Kaizen improvement focuses on the use of :

¬Value – added and non – value work activities.

¬Muda, which refers to the seven classes of waste – over-production, delay, transportation, processing, inventory, wasted motion, and defective parts.

¬Principles of motion study and the use of cell technology.

¬Principles of materials handling and use of one – piece flow.

¬Documentation of standard operating procedures.

¬The five S's for workplace organization.

¬Visual management.

¬Just – in – time principles.

¬Poka – Yoke.

¬Team dynamics.

RE-ENGINEERING

It is the fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical measures of performance.

SUPPLIER PARTNERSHIP

The suppliers should be treated as partners to achieve the same quality level as attained within the organization.

The following forces need Supplier Partnership to improve quality, reduce costs and increase market share.

¬Deming Philosophy (Deming's 4th point)

¬Just-in-time

¬Continuous process improvement

¬ISO 9000

CUSTOMER – SUPPLIER RELATIONS :

Dr. Kaoru Ishikawa has given ten principles of customer-supplier relations. They are

1. Both the customer and supplier are fully responsible for the control of quality.
2. Both the customer and supplier should be independent of each other.
3. The customer is responsible for providing the supplier with clear and sufficient requirements so that the customer can know precisely what to produce.
4. Both the customer and supplier should enter into a non-adversarial contract.
5. The supplier is responsible for providing the quality that will satisfy the customer.
6. Both the customer and supplier should decide the method to evaluate the quality of the product or services.
7. Both the customer and supplier should establish in the contract the method by which they can reach an amicable settlement in case of any dispute.
8. Both the customers and supplier should continually exchange information.
9. Both the customer and supplier should perform business activities.
10. Both the customer and supplier should have the best interest of the end user in mind.

PARTNERING

Partnering is a relationship between two or more parties based upon trust, dedication to common goals.

The benefits of partnering are

¬Improved quality

¬Increased efficiency

¬Lower cost

¬Increased opportunity for innovation

¬Continuous improvement

The three key elements to a partnership relationship are

¬Long term commitment

¬Trust

¬Shared Vision

SOURCING

The three types of sourcing are

¬Sole sourcing

¬Multiple sourcing

¬Single sourcing

SUPPLIER SELECTION

The suppliers should be selected with the following ten conditions

1. The supplier should understand clearly the management philosophy of the organization.
2. The supplier should have stable management system.
3. The supplier should maintain high technical standards.
4. The supplier should provide the raw materials and parts which meet quality specifications required by the purchaser.
5. The supplier should have the required capability in terms of production.
6. The supplier should not leak out the corporate secrets.

7. The supplier should quote right price and should meet the delivery schedule. The supplier should be accessible with respect to transportation and communication.
8. The supplier should be sincere in implementing the contract provisions.
9. The supplier should have an effective quality system such as ISO / QS 9000.
10.The supplier should be renowned for customer satisfaction.

SUPPLIER CERTIFICATION :

A certified supplier is one which, after extensive investigation, is found to supply material of such quality that is not necessary to perform routine testing.

The Eight criteria for supplier certification are

1. No product related lot rejections for atleast 1 year.
2. No non-product related rejections for atleast 6 months.
3. No production related negative incidents for atleast 6 months.
4. Should have passed a recent on-site quality system evaluation.
5. Having a fully agreed specifications.
6. Fully documented process and quality system.
7. Timely copies of inspection and test data.
8. Process that is stable and in control.

SUPPLIER RATING :

Supplier Rating is done

¬To obtain an overall rating of supplier performance.
¬To communicate with suppliers regarding their performance.
¬To provide each supplier with a detailed and true record of problems for corrective action.
¬To enhance the relationship between the buyer and the supplier.

RELATIONSHIP DEVELOPMENT :

For establishment of supplier relationship, the following are necessary.

(a) Partnering

(b) Supplier selection

(c) Principles of customer / supplier relations

(d) Certification

(e) Periodic rating

For relationship development, the following are necessary.

(a) Inspection

¬100% inspection

¬Sampling

¬Audit

¬Identity check

(b) Training

(c) Teams

(d) Recognition and Reward

UNIT III - TQM TOOLS & TECHNIQUES I

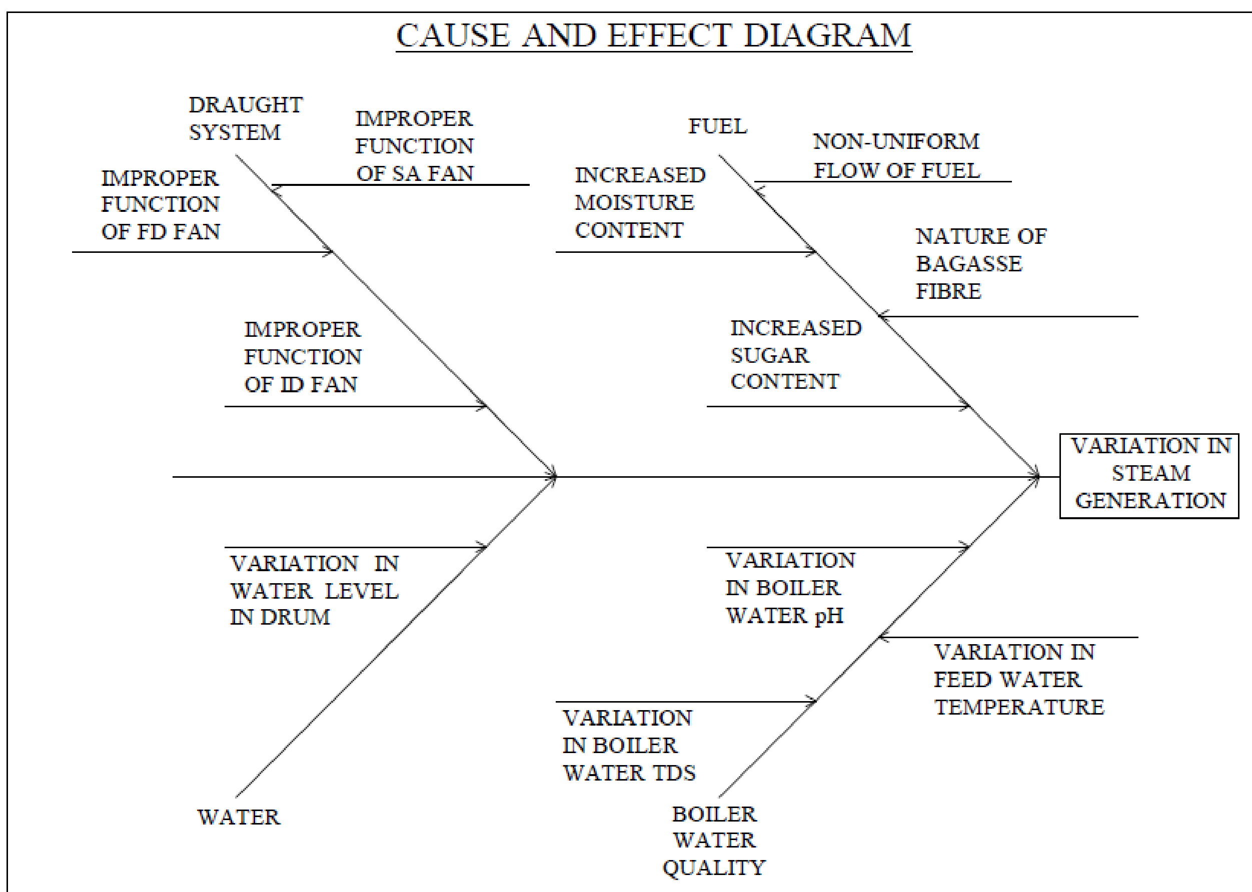
The seven traditional tools of quality – New management tools – Six-sigma: Concepts, methodology, applications to manufacturing, service sector including IT – Bench marking – Reason to bench mark, Bench marking process – FMEA – Stages, Types.

TQM TOOLS (SEVEN TOOLS OF QUALITY)

1. PARETO DIAGRAM

2. FLOW DIAGRAM

3. CAUSE AND EFFECT DIAGRAM



STEPS IN CONSTRUCTING A CAUSE & EFFECT DIAGRAM :

a. Define the problem or effect to be analyzed.

b. Form the team to perform the analysis. Often the team will uncover potential causes through brainstorming.

c. Draw the effect box and the centerline.

d. Specify the major potential cause categories and join them as boxes connected to the centerline.

e. Identify the possible causes and classify them into the categories in step d. Create new categories, if necessary.

f. Rank order the causes to identify those that seem most likely to impact the problem.

g. Take corrective action.

4. CHECK SHEETS

CHECK SHEET

Product : Bicycle

Nonconformity Type	Check	Total
Blister	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ I	21
Light spray	~~IIII~~ ~~IIII~~ ~~IIII~~	15
Drips	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	25
Others	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	25
TOTAL		86

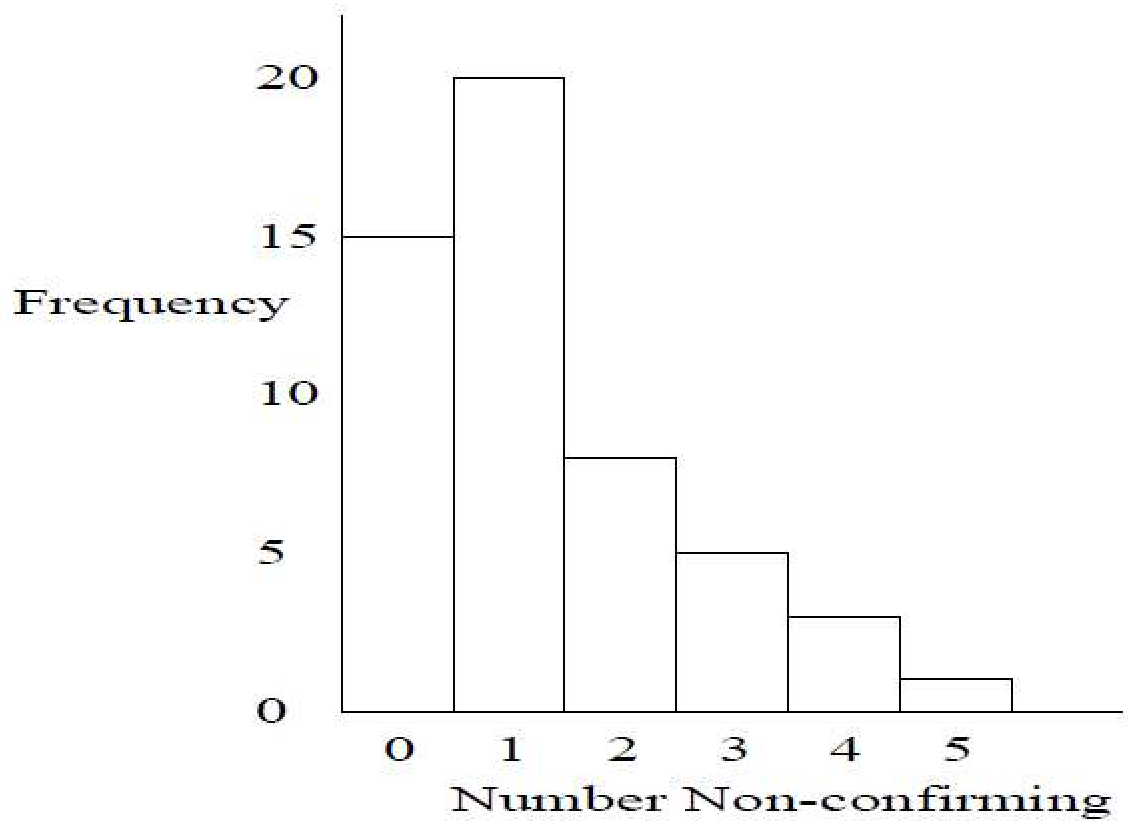
5. HISTOGRAM

Number of Errors

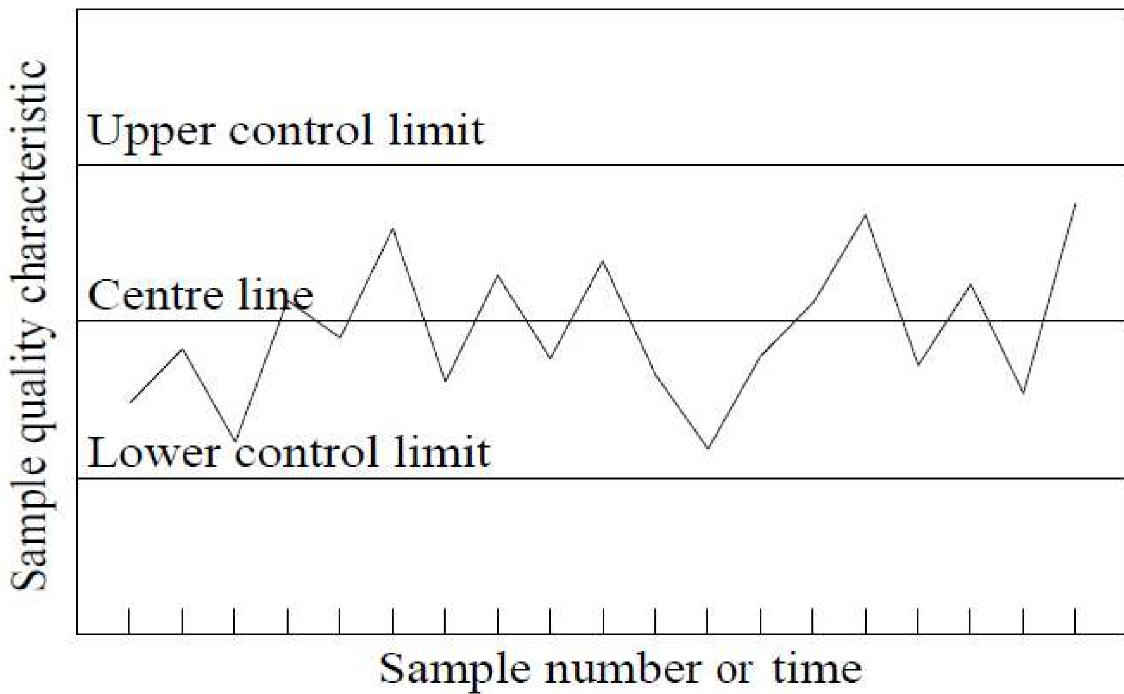
0	1	3	0	1	0	1	0
1	5	4	1	2	1	2	0
1	0	2	0	0	2	0	1
2	1	1	1	2	1	1	
0	4	1	3	1	1	1	
1	3	4	0	0	0	0	
1	3	0	1	2	2	3	

Tally of Number of Errors

Number Non -conforming	Tabulation	Freq.
0	IIII IIII IIII	15
1	IIII IIII IIII IIII	20
2	IIII III	8
3	IIII	5
4	III	3
5	I	1



6. CONTROL CHARTS



7. SCATTER DIAGRAM

In scatter diagram, three types of co-relations exist.

1. Positive correlation.

2. Negative correlation.

3. No correlation.

NEW MANAGEMENT TOOLS

1. WHY,WHY

2. FORCED FIELD ANALYSIS

¬Define the objective.

¬Determine criteria for evaluating the effectiveness of the improvement action.

¬Brainstorm the forces.

¬Prioritize the forces from greatest to least.

¬Take action.

Objective : Stop Smoking

Promoting Forces Inhibiting Forces

Poor Health Habit

Smelly Clothing Addiction

Poor Example Taste

Cost Stress

Impact on Others Advertisement

----

3. NOMINAL GROUP TECHNIQUE
4. AFFINITY DIAGRAM
5. INTER-RELATIONSHIP DIGRAPH
6. TREE DIAGRAM
7. MATRIX DIAGRAM
8. PRIORITIZATION MATRICES
9. PROCESS DECISION PROGRAM CHART
10.ACTIVITY NETWORK DIAGRAM

STATISTICAL FUNDAMENTALS

Statistics is defined as the science that deals with the collection, tabulation, analysis, interpretation and presentation of quantitative data.

Data collected for quality control purposes are obtained by direct observation and are classified as

1. Variables (Measurable quality characteristics like length measured in metres)

2. Attributes (Quality characteristic which are classified as either **conforming** (or) **non-conforming** to specifications, such as “go & no-go” gauge.

MEASURES OF CENTRAL TENDENCY AND DISPERSION

There are two important analytical methods of describing a collection of data as

1. Measures of central tendency.
2. Measures of dispersion.

A measure of central tendency of a distribution is a numerical value that describes how the data tend to build up in the centre. There are three measures in quality as

1. Average
2. Median
3. Mode

Average is the sum of observations divided by the number of observations.

Median is the value which divides a series of ordered observations so that the number of items above it is equal to the number of items below it.

Mode is the value which occurs with the greatest frequency in a set of numbers.
Mode can again classified as

¬No mode
¬Uni mode
¬Bi mode
¬Multimode

Measure of dispersion describes how the data are spread out on each side of the central value.
The two measures of dispersion are

1. Range
2. Standard Deviation

Range is the difference between the largest and smallest values of observations in a series of numbers.

$$Range = R = X_h – X_l$$

Where, R = Range

X_h - highest observation in a series

X_l - lowest observation in a series

Standard Deviation measures the spreading tendency of the data. Larger the standard deviation, greater the variability of data.

POPULATION AND SAMPLE

In order to construct a frequency distribution of the outer diameter of shafts, a small portion (or) sample is selected to represent all the shafts. The population is the whole collection of shafts.

The population may be an hour's production, a week's production, 10000 pieces and so on.

It is not possible to measure all of the population. Hence, we go for sampling.

Sampling becomes necessary

1. When it is impossible to measure the entire population.
2. When it is more expensive to observe all the data.
3. When the required inspection destroys the product.
4. When a test of the entire population may be too dangerous as in the case of new medical drug.

X is for sample average or sample mean.

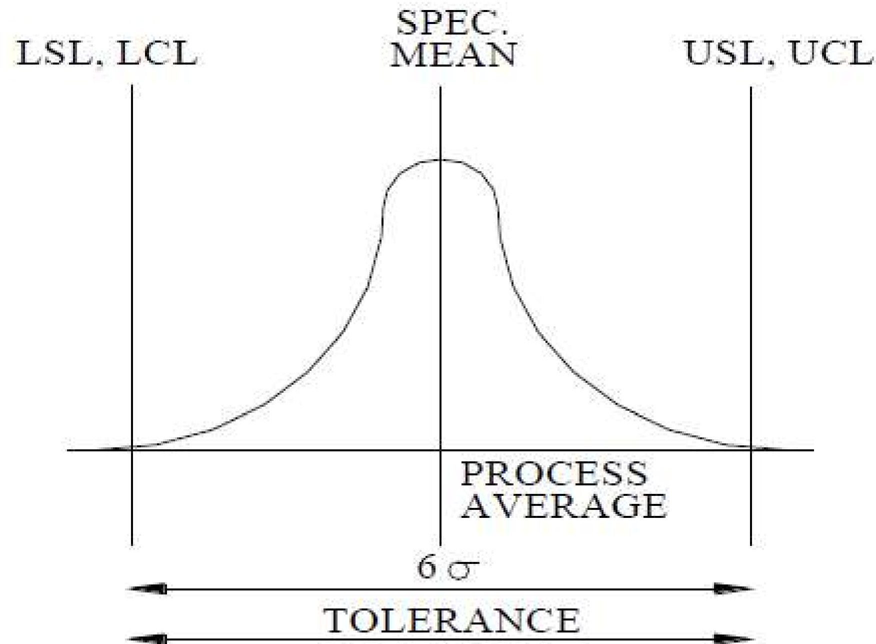
μ is for population mean.

S is for sample standard deviation.

σ is for population standard deviation.

NORMAL CURVE

Normal curve is common type of population. The normal curve is symmetrical, unimodal, bell – shaped distribution with the mean, median and mode all having the same value.



CONTROL CHARTS FOR VARIABLES AND ATTRIBUTES

Variation is a law of nature because no two natural items in any category are the same. Variations are due to the following reasons.

1. Chance causes or Natural causes.

2. Assignable causes.

Chance causes of variation are inevitable. Chance causes affect almost every production process and are inherent in the process. They are purely random, unidentifiable sources of variations.

Hence, when only chance causes are present in a process, the process is said to be in Statistical Control.

Assignable causes result in unnatural variations. The sources of variations may be due to

¬Equipments

¬Materials

¬Environment

¬Operator etc.

The **Control chart** is used to look at variations, seek assignable causes and chance causes. The control chart is a line chart with control limits.

All control charts have three basic components.

1. A centre line, usually the mathematical average of all the samples plotted.
2. Upper and Lower Control Limits that define the constraints of common cause variations.
3. Performance data plotted over time.

A typical control chart is a graphic display of a quality characteristic that has been measured or computed from a **sample** versus **sample number** or **time**. If the process is in control, nearly all of the sample points will fall between **Upper Control Limit (UCL)** and **Lower Control Limit (LCL).**

BENCHMARKING

Benchmarking is a systematic method by which organizations can measure themselves against the best industry practices.

¬Benchmarking is a systematic search for the best practices, innovative ideas, and highly effective operating procedures.

REASONS TO BENCHMARK :

¬It is a tool to achieve business and competitive objectives

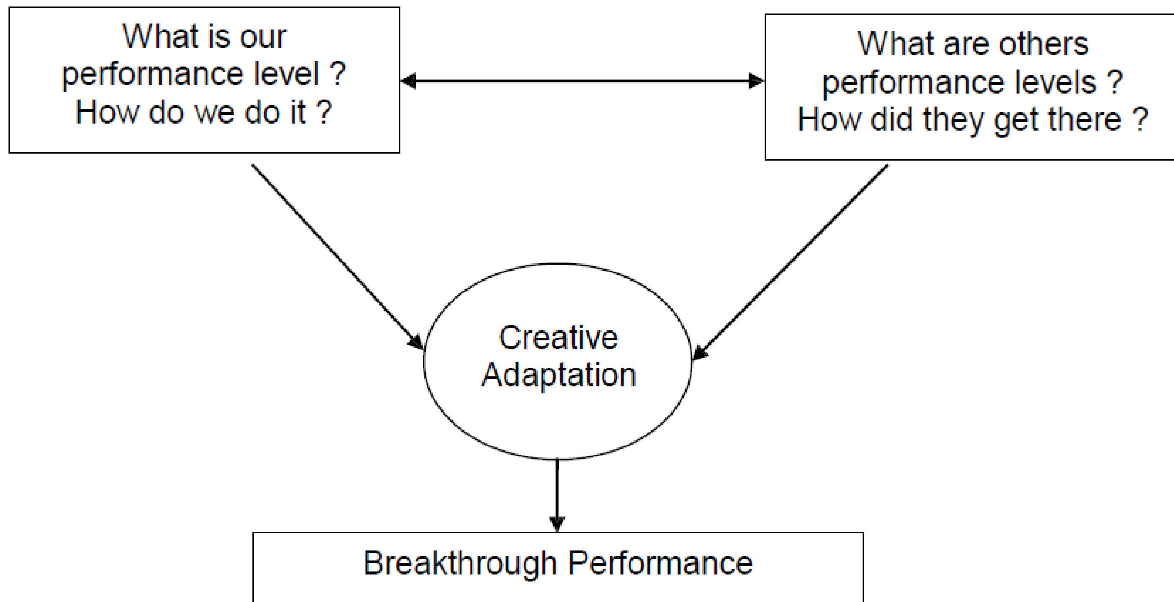
¬It can inspire managers (and Organizations) to compete

¬It is time and cost effective

¬It constantly scans the external environment to improve the process

¬Potential and useful technological breakthroughs can be located and adopted early

BENCHMARKING CONCEPT



PROCESS OF BENCHMARKING

The following six steps contain the core techniques of Benchmarking

1. Decide what to benchmark

¬Benchmarking can be applied to any business or production process

¬The strategy is usually expressed in terms of mission and vision statements

¬Best to begin with the mission and critical factors

¬Choosing the scope of the Benchmarking study

¬Pareto analysis – what process to investigate

¬Cause and Effect diagram – for tracing outputs back

2. Understand current performance

¬Understand and document the current process

¬Those working in the process are the most capable of identifying and correcting problems

¬While documenting, it is important to quantify

¬Care should be taken during accounting information

3. Plan

¬A benchmarking team should be chosen

¬Organizations to serve as the benchmark need to be identified

¬Time frame should be agreed upon for each of the benchmarking tasks

There are three types of benchmarking

a. Internal

b. Competitive

c. Process

4. Study Others

Benchmarking studies look for two types of information

¬How best the processes are practiced

¬Measurable results of these practices

Three techniques for conducting the research are

¬Questionnaires

¬Site visits

¬Focus groups

5. Learn from the data

Answering a series of questions like

¬Is there a gap between the organization's performance and the performance of the best-in-class organizations?

¬What is the gap? How much is it?

¬Why is there a gap? What does the best-in-class do differently that is better?

¬If best-in-class practices were adopted, what would be the resulting improvement?

Benchmarking studies can reveal three different outcomes

¬Negative gap

¬Parity

¬Positive gap

6. Using the findings

The objective is to close the gap. For this

¬Findings must be communicated to the people within the organization

¬Action plans must be developed to implement new processes

Groups that must agree on the change

¬Process owners

¬Upper management

Steps for the development and execution of action plans are

1. Specify tasks
2. Sequence tasks
3. Determine resources needs
4. Establish task schedule
5. Assign responsibility for each task
6. Describe expected results
7. Specify methods for monitoring results

PITFALLS AND CRITICISMS OF BENCHMARKING :

¬Idea of copying others

¬It is not a cure or a business philosophy

¬Some process have to be benchmarked repeatedly

¬It is not a substitute for innovation

SIX SIGMA

The roots of six sigma as a measurement standard can be traced back to Carl Frederick Gauss (1777-1885) who introduced the concept of the normal curve.

Walter Stewarts showed that three sigma from the mean is the point where a process

requires correction.

The objective of six sigma quality is to reduce process output variation so that ± six standard deviations lie between the mean and the nearest specification limit.

This will allow no more than 3.4 Defect Parts per Million (PPM) opportunities, also known as Defects per Million Opportunities (DPMO), to be produced.

Sigma is a Greek alphabet used to represent the distribution on spread (variation) about the mean of a process.

- In manufacturing processes ‘ ’ is used to evaluate the capability of the process to produce defect free output leading to customer’s satisfaction.
- It enables us to make comparisons with similar and dissimilar products and services.
- It is essential to understand the relationship between the sigma value and its implication to meet the required specifications of a process.
- The process capability is estimated from the data collected on a process under statistical control (only chance cause variations are present).
- is called standard deviation in a stable process (process under statistical control). Predictions about process can be made from knowledge of .
- The pattern of variation of measurable characteristics (diameter, lengths, life etc.) will conform to the statistical model known as the normal distribution.
- The total process variation is ± **3** .
- The property of normal distribution is that 99.73% of the entire output will be within 3 **,** only 0.27% output remains beyond ± **3** .

This sort of result can be expected only if the middle value of the specification is set at mean level or at the centre, but in the actual process the mean value may not be at the middle.

When this happens even though process is not affected by any assignable cause variation and is within chance cause variation level only, we call it a positive shift or movement towards higher weight level.

Positive Shift

The process level shifts to the positive side, there will be no change in spread, but there will be increase in defects on the positive side.

Negative Shift

The process level shifts to the negative side without any change in spread and you can see the increase in defects on the negative side.

When we are clear that the process is capable of giving desired output level, then with the help of control charts the process can be monitored to ensure that it is kept under control or within the acceptable level.

Random sampling can be used to monitor the process and also identify' the assignable causes, so that necessary action can be taken to bring back the process under control.

The Japanese quality guru Dr. Hitoshi Kume changed the system of mentioning defects from percentage to parts per million to bring about a psychological change in the people involved in the process.

When 0.27% may look small but 2700 ppm looks big and this makes people react and act which can help in reducing defects.

When we have started taking efforts the chance cause of variations slowly start disappearing and improvements reduce the spread of ± 3 to lesser span as shown in figure.

Six Sigma

Six sigma strategies can be used in an organization to achieve incredible levels of efficiency.

¬The defects can be brought down to a level of 3.4 parts per million.

¬This level is with a shift of 1.5 . If the process can be centered properly the value can be still smaller (two defects per billion).

¬We check the output and if the output is prone to be on the lower or higher side (even if they are within the tolerance).

¬We will be adjusting the process till we get an acceptable level of performance i.e., when the output is closer to the mean value of specification and the variations occur

an positive as well as negative side of the defects level.
¬This essentiates a realistic margin to be provided for dynamic variations (shifts and drifts).
Therefore the long term capability is more significant. In six sigma approach the long term The Motorola has stipulated 1.5 as typical mean shift after a lot of empirical research.
Therefore, 1.5 is the mean shift used in determining long term capability.
T = 12
Process mean is at distance of 6 from LSL and USL.
3.4 defects/million implies 1.5 shifts.
The above process is 6 compliant. The short term capability of their processes is 6a.
The long term performance is 4.5. Process mean is at distance of 6 T = 12 a
Process mean is at distance of 6 from LSL and USL.
3.4 defects/million implies 1.5 shifts.
The above process is 6 a compliant. The short term capability of their processes is 6
The long term performance is 4.5 from LSL and USL.
3.4 defects/million implies 1.5 shift.
The above process is 6 compliant. The short term capability of their processes is 6
. The long term performance is 4.5

Units Used to Measure Defects.
The different units used to measure defects are
Six Sigma and Process Capability
The sigma level corresponding to the defect levels can be known from defects per million or sixSigma tables.
When the **Cp** (capability index) value is higher then the variation will be lesser
Higher the **Cpr** (achieved capability index) value, lesser will be the variation and the process average will be nearer to the target.
A process is said to of 6 when **Cp 2** and **Cpk 1.5**.
THE NEW SEVEN MANAGEMENTAND PLANNING TOOLS

The “advanced” tools that are used to manage cross-functionality include the “seven new QC tools” also known as the “seven management and planning tools” or “7 MP tools”.

These are summarized and illustrated in Hizuno (1979) and Brassard (1989)

- Identify a system owner and team members for each critical system.
- Describe the system under study.
- Identify all Identify subsystems that contribute to the critical system.
- Define the interdependencies of the subsystems.
- Prioritize the subsystems as to their contribution to the critical system.
- Develop a detailed “as is” description of the critical system.
- This includes identifying the interfaces between all system components as well as expanding the level of detail for major contributing subsystems.
- Identify obvious system deficiencies.
- Identify possible cause of system deficiencies.
- Establish “base line” measures for the system and major subsystems.
- Assess the performance of the system and major subsystems.
- Develop a “should be” description of the system and subsystems.
- Recommend changes to improve system and subsystem performance (even if this may mean creating a new system)

These are very important tools which are used in management and planning activities, they are mainly used to organize and bring into a proper structure, the un organized unstructured complex ideas.

The new seven management and planning tools are:

- Affinity diagrams
- Inter relationship diagram
- Tree diagram
- Matrix diagram
- Matrix data analysis

‗Process decision programme chart and
‗Arrow diagram.

Affinity Diagrams

In affinity diagrams large volumes of data is gathered and organized. Ideas, opinions and facts relating to a problem are grouped.

⁀A sequence or pattern formation is the main aim. This is mainly useful in addressing issues such as customer dissatisfaction etc.

⁀Affinity diagrams are tools for verbal data.

⁀Its applications are to organize into groups a large number of ideas, opinions about a particular topic.

Interrelationship Diagram

The relationships between different causative factors and the main issue are established.

‗This tool helps us in identify the relationship between different factors which cause a problem or issue.

‗It also helps in determining the interrelationship between these factors.

‗This tool is used to identify the major causes which help in solving a problem on the basis of logical analysis and linkage of causes associated with the problems.

Tree Diagram

Tree diagram is listed as a tool for non-numerical data. It is used to show the relationships between an issue and its component elements.

⁀This is a tool used for operational planning after initial diagnosis of issues.

⁀The final chart shows steps from the initial problem to the sequential development till the final conclusion.

⁀It looks more like a tree; therefore it is called a tree diagram.

Matrix Diagram

A matrix diagram consists of a set of columns and rows. The intersections of these rows and columns are checked for determining the nature and strength of the problem.

‗This helps us to arrive at the key ideas and determining the relationship and an effective way of perusing the problem.

‗The ideas are conceived on two dimensional relationship basis.

Matrix Data Analysis

It is a multi variate analysis technique also known as the "principal component analysis". This technique quantifies and arranges data presented in a matrix diagram to find more general indicators that would differentiate and give clarity to large amount of complexly intertwined information.

Process Decision Programme Chart (PDPC)

It is a method which maps out conceivable events and contingencies that can occur in any implementation plan along with appropriate counter measures

‗This tool is used to plan each possible chain of events that needs to occur when the problem or goal is unfamiliar one.

‗This is a qualitative tool. Thus PDPC is useful whenever uncertainty exists in a proposed implementation plan. The background situation for a PDPC application is:

⁀The talk on hand is either new or unique

⁀Implementation plan has sufficient complexity

⁀Consequences of failure are serious

⁀Efficiency of implementation is critical

⁀Contingencies must be planable.

Arrow Diagram

Arrow diagram is a tool to plan the most appropriate schedule for the completion of a complex task and its related sub-tasks.

‗It projects likely completion time and monitors all sub-tasks for adherence to

necessary schedule.
‿The total work or task is broken down to sub-tasks or activities.
‿The sub-tasks and the total work is linked by arrows and a diagram is constructed to depict the activities.

Failure Modes and Effects Analysis [FMEA]

Introduction

Customers are placing increased demands on companies for high quality, reliable products. The increasing capabilities and functionality of many products are making it more difficult for manufacturers to maintain the quality and reliability.
‿Traditionally, reliability has been achieved through extensive testing and use of techniques such as probabilistic reliability modeling.
‿Failure Modes and Effects Analysis (FMEA) is methodology for analyzing potential reliability problems early in the development cycle where it is easier to take actions to overcome these issues, thereby enhancing reliability through design.
‿FMEA is used to identify potential failure modes, determine their effect on the operation of the product, and identify actions to mitigate the failures.
‿A crucial step is anticipating what might go wrong with a product. While anticipating every failure mode is not possible, the development team should formulate as extensive a list of potential failure modes as possible.
‿The early and consistent use of FMEAs in the design process allows the engineer to design out failures and produce reliable, safe, and customer pleasing products. FMEAs also capture historical information for use in future product improvement.

Types of FMEA's

There are several types of FMEAs; some are used much more often than others. FMEAs should always be done whenever failures would mean potential harm or injury to the user of the end item being designed. **The types of FMEA are:**
⁀System - focuses on global system functions
⁀Design - focuses on components and subsystems

- Process - focuses on manufacturing and assembly processes
- Service - focuses on service functions
- Software - focuses on software functions

System FMEA

This is used to analyze systems and subsystems in the early concept and design stage. A system FMEA focuses on potential failure modes between the functions of the system caused by system deficiencies. It includes the interactions between systems and elements of the system.

The outputs of the system FMEA are

- A potential list of failure modes ranked by RPN.
- A potential list of system functions that could detect potential failure modes.
- A potential list of design actions to eliminate failure modes, safety issues and reduces the occurrence.

The benefits of the system FEMA are that it: -

- Helps to select the optimum system alternative.
- Helps to determine redundancy.
- Helps to define the basis for system level diagnostic procedures.
- Increases the likelihood that potential problems will be considered.
- Identifies potential system failures and their interaction with other systems.

Design FMEA

This is used to analyze products before they are released to manufacturing. A design focuses on failure modes caused by design deficiencies.

The outputs of the design FMEA are: -

- A potential list of failure modes ranked by RPN.
- A potential list of critical and / or significant characteristics.
- A potential list of design actions to eliminate failure modes, safety issues and

reduces the occurrence.

A potential list of parameters for appropriate testing, inspection, and / or detection methods.

A potential list of recommended action for the critical and significant characteristics.

The benefits of design FMEA are that it: -

Establishes a priority for design improvement actions.

Documents the rationale for changes.

Provides information to help through design verification and testing.

Helps to identify the critical or significant characteristics.

Assists in the evaluation of design requirements and alternatives.

Helps to identify and eliminate potential safety concerns.

Helps to identify product failure early in the product development phase.

Process FMEA

This is used to analyze manufacturing and assembly processes. A process FMEA focuses on
failure modes caused by process or assembly deficiencies. **The outputs of the process
FMEA are**: -

A potential list of failure modes ranked by RPN.

A potential list of critical and/ or significant characteristics.

A potential list of recommended actions to address the critical and significant characteristics.

A potential list to eliminate the causes of failure modes, reduces their occurrences, and improves defect detection if Cpk cannot be improved.

The benefits of the process FMEA are that it: -

- Identifies process deficiencies and offers a corrective action plan.
- Identifies the critical and/ or significant characteristics and helps in developing control plans.
- Establishes a priority of corrective actions.
- Assists in the analysis of the manufacturing or assembly process.
- Documents the rationale for changes.

Service FMEA

This is used to analyze services before they reach the customer. A service FMEA focuses on the failure modes (tasks, errors, mistakes) caused by system or process deficiencies.

The outputs of the service FMEA are: -

- A potential list of errors ranked by RPN.
- A potential list of critical or significant tasks, or processes.
- A potential list of bottleneck processes or tasks.
- A potential list to eliminate the errors.
- A potential list of monitoring system / process functions.

The benefits of the service FMEA are that it: -

- Assists in the analysis of job flow.
- Assists in the analysis of the system and/ or process.
- Identifies task deficiencies.
- Identifies critical or significant tasks and helps in the development of control plan.
- Establishes a priority for improvement actions.
- Documents the rationale for changes.

An FMEA program should start

- When new systems, designs, products, processes, or services are designed.
- When existing systems, designs, products, processes, or services are about to

change regardless of reason.

- When new applications are found the existing conditions of the systems, designs, products, processes, or service.
- When improvements are considered for the existing systems, designs, products, processes, or services.

FMEA Usage

Historically, engineers have done a good job of evaluating the functions and the form of products and processes in the design phase.

- They have not always done so well at designing in reliability and quality. Often the engineer uses safety factors as a way of making sure that the design will work and protected.”
- A large safety factor does not necessarily translate into a reliable product. Instead, it often leads to an over designed product with reliability problems."
- FMEA's provide the engineer with a tool that can assist in providing reliable, safe, and customer pleasing products and processes.
- Since FMEA help the engineer identify potential product or process failures, they can use it to:
- Develop product or process requirements that minimize the likelihood of those failures.
- Evaluate the requirements obtained from the customer or other participants in the design process to ensure that those requirements do not introduce potential failures.
- Identify design characteristics that contribute to failures and design them out of the system or at least minimize the resulting effects.
- Develop methods and procedures to develop and test the product/process to ensure that the failures have been successfully eliminated.
- Track and manage potential risks in the design. Tracking the risks contributes to the development of corporate memory and the success of future products as well.

- Ensure that any failures that could occur will not injure or seriously impact the customer of the product/process.

Benefits of FMEA

FMEA is designed to assist the engineer improve the quality and reliability of design. Properly used the FMEA provides the engineer several benefits. Among others, these benefits include:

- Improve product/process reliability and quality
- Increase customer satisfaction
- Early identification and elimination of potential product/process failure modes
- Prioritize product/process deficiencies
- Capture engineering/organization knowledge
- Emphasizes problem prevention
- Documents risk and actions taken to reduce risk
- Provide focus for improved testing and development
- Minimizes late changes and associated cost
- Catalyst for teamwork and idea exchange between functions

Who Performs the FMEA?

- The FMEA should be initiated by the design engineer for the hardware approach, and the systems engineer for the functional approach.
- Once the initial FMEA has been completed, the entire engineering team should participate in the review process.
- The team will review for consensus and identify the high-risk areas that must be addressed to ensure completeness.
- Changes are then identified and implemented for improved reliability of the product. The following is a suggested team for conducting/reviewing an FMEA.

– Project Manager

– Design Engineer (hardware/software/systems)

– Test Engineer
– Reliability Engineer
– Quality Engineer
– Field Service Engineer
– Manufacturing/Process Engineer
– Safety Engineering
Outside supplier engineering and/or manufacturing could be added to the team.
Customer representation is recommended if a joint development program between user/supplier exists.

Stages of FMEA (Process)

Since the FMEA concentrates on identifying possible failure modes and their effects on the equipment, design deficiencies can be identified and improvements can be made. Priorities on the failure modes can be set according to the FMEA's risk priority number (RPN) system analysis. **The FMEA process consists of the following figure**

- FMEA Prerequisites
- Functional Block Diagram
- Failure mode analysis and preparation of work sheets
- Team Review
- Corrective action

FMEA Timing

The FMEA is a living document. Throughout the product development cycle change and updates are made to the product and process. These changes can and often do introduce new failure modes.

- A new product or process is being initiated (at the beginning of the cycle).
- Changes are made to the operating conditions the product or process is expected to function in.
- A change is made to either the product or process design. The product and

process are inter-related. When the product design is changed the process is impacted and vice-versa.

⁀New regulations are instituted.

⁀Customer feedback indicates problems in the product or process.

FMEA Procedure

The process for conducting an FMEA is straightforward. The basic steps are outlined below.

Describe the product/process and its function.

Create a Block Diagram of the product or process.

Complete the header on the FMEA Form worksheet: Product/System, Subsys./Assy.,

The diagram prepared above to begin listing items or functions

⁀A failure mode in one component can serve as the cause of a failure mode in another component

⁀Describe the effects of those failure modes. For each failure mode identified the engineer should determine what the ultimate effect will be

⁀Establish a numerical ranking for the severity of the effect. A common industry standard scale uses 1 to represent no effect and 10 to indicate very severe with failure affecting system operation and safety without warning Identify the causes for each failure mode.

⁀Enter the Probability factor. A common industry standard scale uses 1 to represent not likely and 10 to indicate inevitable.

⁀Identify Current Controls (design or process detect failures.. Review Risk Priority Numbers (RPN).

⁀The Risk Priority Number is a mathematical product of the numerical Severity, Probability, and Detection ratings:

RPN = (Severity) x (Probability) x (Detection)

The RPN is used to prioritize items than require additional quality planning or action.

- Determine Recommended Action(s) to address potential failures that have a high RPN..
- Assign Responsibility and a Target Completion Date for these actions. This makes responsibility clear-cut and facilitates tracking.
- Indicate Actions Taken. After these actions have been taken, re-assess the severity,
- Update the FMEA as the design or process changes, the assessment changes or new information becomes known.

UNIT IV - TQM TOOLS & TECHNIQUES II

Control Charts – Process Capability – Concepts of Six sigma - Quality Function Deployment (QFD) – Taguchi quality loss function – TPM – Concepts, improvement needs – Cost of Quality – Performance measures.

QUALITY FUNCTION DEPLOYMENT [QFD]

1. INTRODUCTION

Numbers of companies have implemented TQM but few have only reaped its full benefits. There are different approaches as performed by Juran, Crosby, Deming and Feigenbaum.

- Seven QC tools and seven new QC tools are employed for implementing TQM.
- The TQM philosophy emphasizes the role of people, usually in multifunctional teams to bring about improvement from within the organization.
- The heart of the philosophy is the prevention of problem and an emphasis on quality in design and development of product and process.
- The QFD is one technique which is a successful tool in quality planning process.

2. QUALITY FUNCTION DEPLOYMENT

The products and services are developed quickly and inexpensively without sacrificing quality or losing focus of customer requirements. The QFD is found to be a tool due to the following reasons

- The QFD will save development time, which will reduce the costs and allow the company to respond more easily to changes in the market.
- The QFD will focus the company's resources on providing those capabilities that drive customer satisfaction.
- The QFD ensures that issued critical to the success of the product do not get dropped.

American Supplier Institute defined the QFD as **“a system for translating customer requirements into appropriate company requirements at every stage, from research, through product design and development, to manufacture, distribution, installation and marketing, sales and service.”**

- The QFD methodology was first developed in Japan by Dr. Shigorce Mizuno in 1960’s. But them this was popularly known as quality tables.
- The first application of QFD was at Mitsubishi, Heavy Industries Ltd., in the Kobe shipyard, Japan in 1972.
- Later, this was used by Toyota in the production of mini-vans.
- The QFD was introduced to United States by Dr. Clausing of Xerox in 1984.
- Since then the QFD has become one of the most important tool in Quality Planning Process.
- QFD is a team based management tool which ingrains the views of the customer into the products in a systematic involving all the departments/sections of manufacturing/service. Normally, the views of customers are popularly called as voice of the customers. i.e., the customer requirements have to be referred in their terms of reference.
- The voice of the customers can be collected from market survey and research. In the Phase I of QFD, the voice of the customers will form part of rows in a matrix structure while the part features will be in the columns.
- Phase II of QFD deals with part features Vs part characteristics. Phase III will deal with part characteristics Vs process characteristics.
- Phase IV will deal with process characteristics Vs control parameters. An example of product or process planning is shown in figure

3. STRUCTURE OF QFD

The QFD chart looks like a house as shown in figure

- The left hand side of the chart is used to list what the customer expects in a product and hence this section is referred as “WHATS SECTION”.

⁀The right hand of the chart is used to show importance rating, target value (or) goal, Improvement ratio (or) scale up factor, sales point and customer benchmarking etc.

⁀The left and right of the chart look like a left and right of a house.

Figure Structure of QFD

‗The top side of chart is used to list the Technical requirements or voice of the Engineers to meet the customer requirement. Each row i.e. Customer requirements may be related one or more technical requirements (each column).

‗The body of the chart is to show the relationship between Customer Requirements (CR) and Technical Requirements (TR).

‗The technical requirements are known as “HOWS” to achieve the “WHATS”. And the body looks like a interior of a House.

‗The top triangular part of the chart is used to show the interrelationship of the technical requirements and it looks like the roof the house.

‗The bottom part of QFD chart is to represent technical competitiveness, targets and benchmarking etc. The bottom of the chart looks like a foundation of the house.

‗And hence the QFD chart is very popularly known as House of Quality (HOQ) charts.

‗The others QFD charts can be make deploys columns of the current chart into rows of the next chart and find new “HOWS” to achieve the customer requirements taking them to various phases of manufacturing.

‗The relationship between “WHATS” and “HOWS” are normally classified as weak, moderate and strong and they are rated numerically 1, 3 and 9 respectively. They can also be represented symbolically as follows.

The following case studies are useful in understanding the QFD chart.

House of Quality

CASE STUDY 1

A simple case is that how pencils quality shall be analyzed with QFD. A market research has to be conducted, identifying the voice of customers with importance rating (1 to 5 scales or 1 to 9 scales)

They may be as follows with their importance rating

Importance rating

1. Easy to hold 3
2. Does not smear 5
3. Point lasts long 5
4. Doest not roll 1

The voice of engineers shall be collected by giving them the outcome of market research. They provide the technical requirement to meet the above customer requirements. And they may listed as follows.

1. Length of pencil
2. Time between sharpening

Weak

1 to 9 Scale

1 1

3 3

9 5

Moderate

Strong

1 to 5 Scale

3. Lead dust generated
4. Hexagonally
5. Minimum erasure residue.

⁀A matrix is drawn as shown in fig. Using the voice of the customer and voice of the engineer.

⁀A brainstorming session shall be conducted to know the relationship between

customer requirements and technical requirements.

- In this case study, easy to hold is found to be strong with length of pencil and it is marked with the symbol " ".
- Similarly the relationships are filled for every other cell. The empty cell means that there is no relationship.
- The right hand side of the chart contains goal, improvement ratio, sales point, demanded weight and relative weight.
- Goal with 5 points means that the company wants to fix high the respective customer requirements. In this case "does not smear" and "point lasts" are considered with "5" point goal. While the other two are with 3 point goal.
- Improvement ration = Goal / Average of each row
- Say, the improvement ratio for point lasts = 5/ (3+5+5/3) = 5/4.3
- Likewise, the other rows are filled.The sales point is the customer requirement with which the company wants to sell the product.
- Normally 1.5, 1.2 and 1.0 are the points given for a customer requirement.
- The point lasts is given a 1.5 point in this case while the others are at 1.0.
- The demanded weights are the product of goal, improvement ratio and sales point.
- The Customer requirements with the higher relative weights are normally focused for improving the quality of the pencil.
- The column with average gives the absolute importance. The relative importance of technical requirements is the sum of the product of relative weight and cell relationship values.
- The technical requirements are sincerely attended by the company for which the relative importance is high.
- The right hands side of the chart i.e. graphs shows the present competitive level of the product. A competitor can be compared for each customer requirement and benchmarked for improvement.
- The roof top of the chart represents the co-relationship between the technical requirements.

⁀In this case time between sharpening has a positive co-relation with lead dust generated and minimal erasure residue.

⁀When we involve only one chart to do the quality improvement, shall designate the chart as House of Quality Chart.

⁀When we go in for move than one deployment chart, then the charts shall be called as Quality Function Deployment Charts.

Figure. Fishbone diagram of castings

Figure QFD External Customer requirements-Castings Manufacturing industry

‗The customer requirements timely delivery, graphite size, free from dimensional deviation and good surface finish are the areas in which the improvements are expected.

‗They are indicated by higher demanded weights as it is shown in figure 7. (see Excel file).

‗The functional weights indicates that the design of pattern and construction, schedule as per order and sand preparation are the control parameters need to be attended for improvement of quality while they continue to maintain the other parameters.

CASE STUDY 4: A QFD APPLICATION IN A TECHNICAL INSTITUTION

An upcoming technical institution is focused in this case study for QFD application.

‗The customers of a technical institution may be classified as students, industries and parents & society.

‗The stages of the QFD had been employed to translate the Voice Of the Customer (VOC) into operation requirements.

‗In the first stage, the VOC are translated into service elements.

Customer Competitive Evaluation (CCE) = Average of each row

Improvement Ratio = Goal/CCE

Absolute Weight = Importance * Improvement ratio * sales point

Demanded Weight = Percentage weight of absolute weight
Functional Weight (Absolute) = Sum of product of relationship of voice
Customer & technical requirements and
respective demanded weight.
Weight in % = Percentage weight of Functional weight
The SE is translated into Key Process Operations (KPO) in the second stage.
Finally the KPO are translated into Operations Requirements.
The authors have focused the specific area "Course Content" for this study and identified the VOC through brainstorming.
The Service Elements (SE), Key Process Operations (KPO) and Operational Requirements (OR) have also been identified as shown in table 1,2 an 3 by brainstorming with a group of faculty members.

Voice of Customers Requirement (VOC) and Service Elements (SE)

Service Elements (SE) and Key Process Operations KPO)

Sl.No Voice of Customers (VOC) Service Elements (SE)
Recent Advancements Material Collection
2 Detailed Course Contents Lecture Material
3 Course Material Technical Knowledge
4 Technical Skills Lecture Plan
5 Clarity in Explanation Qualification
6 Conceptual Skills Competency in Teaching
7 IT Skills Competency in Software and IT Skills
8 Presentation Skills Laboratory Component
9 Communication Skills Computer Based tutorials
10 Analytical Skills Class discussions
11 Assignments/ Term paper
12 Mini Projects
13 Case studies

14 Consultation Hours

15 Project works

Sl.No Service Elements Key Process Operations (KPO)

1 Material Collection Link with Universities

2 Lecture Material Link with Industries

3 Technical Knowledge Prepare Lecture Notes

4 Lecture Plan Attend Seminars/Conferences

5 Qualification Consulting experts

6 Competency in Teaching Specify the Min. qualification

7 Competency in Software and IT skills Get feed back

8 Laboratory Component Prepare Laboratory Experiments

9 Computer based tutorials Prepare CBTs

10 Class Discussions Encourage Group Discussions

Table 3: Key Process Operations (KPO) Operation Requirement (OR)

The stage I, stage II and stage III of QFD are shown in FiguresThe importance ratings, co-relationship ratings, goal points and sales points are assigned based on the discussions with the faculty and students.

_The above formulas are used and the required weight ages are calculated. The technical competitive evaluation and the functional weight in percentages are also calculated.

_The Customer Competitive Evaluation (CCE) and Technical Competitive Evaluation (TCE) have been plotted as shown in Figures.

11 Assignments/ Term paper Design Criteria /Structure

12 Mini Projects Attach to Industries

13 Case studies Specify Consultation hours

14 Consultation Hours

15 Project works

Sl.No Key Process Operations (KPO) Operation Requirements (OR)

1 Link with Universities Establishing Good Relationship

2 Link with Industries Holding Meetings

3 Prepare Lecture Notes Inviting Experts

4 Attend Seminars/Conferences Conducting Course Evaluation

5 Consulting experts Organizing Seminars

6 Specify the Min. qualification Holding Publisher fair

7 Get Feedback Attending Seminars/Conferences

8 Prepare Laboratory Experiments Organizing Faculty Exchange Programme

9 Prepare CBTs Organizing Faculty Development Programme

10 Encourage Group Discussions Arranging Transparent Feedback

11 Design Criteria /Structure Arranging Software / IT Training

12 Attach to Industries Providing Intranet and internet

13 Specify Consultation hours Providing Teaching Assistant

14 Providing Power Point Presentations

15 Providing Resources for CBTs

The important VOC, SE, KPO and OR are identified based on required weight ages and functional weight ages respectively. Each stage the concept of Value Engineering should be used to know the value addition by the respective elements.

ANALYSIS

The VOC of the customers with respect to "Course Content" have been identified in their terms.

The respective Service elements are found through through brainstorming. The existing relationship VOC and SE are studied for improving the quality.

And Its Customer Competitive Evaluation (CCE) have been plotted and compared with that of competitors. Likewise, it has been carried out for stage II and stage III.

Figure QFD Chart of VOC and SE (Course Content) in a Technical Institution

QFD Chart of SE and KPO (Course Content) in a Technical Institution

QFD Chart of KPO and OR (Course Content) in a Technical Institution

The QFD charts indicate the areas that are to be considered immediately. For example, the VOC, recent advancements, IT kills and technical skills are to be focused as per the higher required weight age as shown in Figure. 8. Similarly, the service elements, the key process operations, etc. may be decided based on their higher functional weight age. The importance of the VOC are changing with time and environment. Accordingly the service elements, key process operations and operations requirements may also change. The consistent and persistent implementation of strategic QFD would lead to identify important areas for continuous improvement of quality of engineering education.

TAGUCHI'S QUALITY LOSS FUNCTION

INTRODUCTION

There are various measures used to measure the quality level. Percentage defectives, process capability index and warranty cost are some of the measures of quality level of shipped products.

⁀The percent defectives (or) warranty costs are understandable quality measures by anybody because they are related or converted cost.

⁀However, the customers are put to hardships due to time lag in taking action against warranty claims.

⁀On the other hand, the process capability indexes are used to measure quality but not as a cost problem.

⁀The actual significance of process capability index changes from a particular value to another value cannot be equally understood.

Say for example, standard deviation of a particular product of a company is 1/6 while the rival company's standard deviation is 1/ 12. The tolerance level is specified by the customers and it is common for both the companies.

Tolerance

Company ‘A’:

Cp = 6 x Std. Deviation

Tolerance = 1

time of tolerance = 6 x 1/6 Tolerance

Company ‘B’:

Cp = 6 x Std. Deviation

Tolerance = 0.048

time of tolerance = 6 x 1/ 12

The difference in Cp values does not reflect an understandable measure of quality for customer and it does not consider the problem as a cost related. Hence a quality level measure is introduced to consider the loss of customers if the product characteristic (y) deviates from the target values (m).

Uses of the loss function

The total loss due to the deviation of the product from its target value can be calculated.

It can be used to compare the current process and improved process.

Loss function and types of tolerances

Normally three different types of tolerances are employed in production/service.

They are normally

The Nominal –The Best (N – type)

The Smaller- The Better (S-type)

The Larger- The Better (L-type)

The Nominal –The Best (N – type)

Bilateral tolerances are used in this type of tolerance system. Under Ntype, the tolerance is related to a basic dimension and is given in two directions

in plus and minus.

The S- Type

The specification of certain quality characteristics such as wear, shrinkage, noise level and impurity etc will be zero or positive but not negative.
In this type, the quality characteristics will be greater than or equal to zero.
i.e. Y   0 and the target value ‘m’ is zero.

The L Type

The quality characteristics such as the strength of materials and fuel efficiency fall in this type of tolerance. The larger the values of these characteristics better the situation. The L-type can be written as ‘S’ type by assuming a characteristics as follow; Y – L – type quality characteristic

TOTAL PRODUCTIVE MAINTENANCE

Introduction

Total Productive Maintenance (TPM) is keeping the current plant and equipment at its highest productive level through cooperation of all areas of the organization
_This approach does not mean that such basic techniques as predictive and preventative maintenance are not used; they are essential to building a foundation for a successful TPM environment.
_Predictive maintenance is the process of using data and statistical tools to determine when a piece of equipment will fail, and preventative maintenance is the process of periodically performing activities such as lubrication on the equipment to keep it running.

Analyzing TPM into its three words, we have:

Total = All encompassing by maintenance and production individuals working together.
Productive = Production of goods and services that meet or exceed customers

expectations.

Maintenance = Keeping equipment and plant in as good as or better than the original condition at all times.

The overall goals of TPM are

- Maintaining and improving equipment capacity.
- Maintaining equipment for life.
- Using support from all areas of the operation.
- Encouraging input from all employees.
- Using teams for continuous improvement.

THE PLAN

Total Productive Maintenance (TPM) is an extension of the Total Quality Management (TQM) philosophy to the maintenance function. Seven basic steps get an organization started toward TPM:

- Management learns the new philosophy.
- Management promotes the new philosophy.
- Training is funded and developed for everyone in the organization.
- Areas of needed improvement are identified.
- Performance goals are formulated.
- An implementation plan is developed.
- Autonomous work groups are established.

PROMOTING THE PHILOSOPHY

Senior management must spend significant time in promoting the system.

- They must sell the idea and let the employees know that they are totally committed to its success.
- Like TQM or any other major change in an organization, there must be total commitment from the top. If the belief in the new philosophy and commitment are not there, then positive results will not happen.

‗Too often lip service is given to a "new idea."
‗This action is usually brought on by a belief that the new system will solve some immediate problems and lead to an immediate return on investment.
‗A long-term commitment to the new philosophy is required. It has been proven by other organizations to be a better way of doing business.
‗Introducing TPM with a huge fanfare leads employees to shrug it off as the latest method for getting them to work harder.
‗Management must first build credibility, and the best way to accomplish that task is to change first and lead the way.

TRAINING

Teach the philosophy to managers at all levels. Begin with senior management, and work down to first-line supervisors.
‗Senior management must spend time learning about and understanding the ramifications of applying this philosophy to their organization.
‗Middle management must learn how to deal with the team approach and how small autonomous work groups function.
‗This organizational level seems to have the greatest difficulty with this type of change. In recent years, downsizing has come at the expense of middle managers.
‗First-line supervisors need to learn their role in what most likely will be a new environment. Supervisors who have been used to guiding their groups will find this an easy transition.
‗There needs to be some instruction in the area of jobs that maintenance people do and jobs that production people do.
‗A great benefit of TPM is the cross-pollination of ideas between maintenance technicians and production operators.

IMPROVEMENT NEEDS

Employees who work with the equipment on a daily basis are better able to identify these conditions than anyone else in the Organization. An implementation team of

operators and technicians to coordinate this process is essential. This action will build credibility and start the organization towards TPM.

Six major loss areas need to be measured and tracked:

Downtime Losses

- Planned
- Unplanned Downtime
- Idling and minor stoppages
- *Slow-downs*

Poor quality losses

- *Pro*cess nonconformities
- Scrap

Goal

Goals should be set after the improvement needs are identified. A good first goal is to establish the timeframe for fixing the first prioritized problem.

- Technicians and operators will probably want it done faster than management because it causes them more problems on a daily basis.
- Identifying needs and setting goals begins the process of getting the organization to work together as a team.

Developing Plans

First develop and implement an overall plan of action for training all employees. Plans for developing the autonomous work groups should take place during the training phase.

- Using the team approach will set the stage for the development of autonomous work groups, which are teams established for daily operations
- Part of the planning process should take into consideration that autonomous work groups will change over time.

As processes and procedures are improved, the structure of the whole organization will change. It would be unreasonable not to expect autonomous work groups to change also.

Autonomous Work Groups

Autonomous work groups are established based on the natural flow of activity.

o First, make the operator responsible for the equipment and the level of maintenance that he is capable of performing.

Operators and maintenance personnel are brought together, resulting in an autonomous **work group.**

QUALITY COST

Quality costs are defined as those costs associated with the non-achievement of product or service quality as defined by the requirements established by the organization and its contracts with customer and society.

Quality costs:

Costs of control

Prevention costs

Appraisal costs

Costs of failure of control

Internal failure costs

External failure costs

Quality costs are broadly classified into:

1. Failure costs
2. Appraisal costs
3. Prevention costs

Failure costs:

Failure costs are the direct and indirect costs incurred on those products or services, which fail to comply with their prescribed specifications.

¬Internal failure costs – before delivered to customers

¬External failure costs – after delivered

Appraisal costs:

Appraisal costs are those costs involved in actual checking of the quality, viz., the cost of carrying out actual inspection.

Prevention costs:

Prevention costs are those, which are involved in ensuring that faulty or defective **work, or rejections are not produced in very first instance.**

PERFORMANCE MEASURES

Performance measures are required for the managers for managing an organization perfectly.

Performance measures are used to achieve the following objectives.

¬To establish performance measures and reveal trend.

¬To identify the processes to be improved.

¬To determine the process gains and losses.

¬To compare the actual performance with standard performance.

¬To provide information for individual and team evaluation.

¬To determine overall performance of the organization.

¬To provide information for making proper decisions.

WHAT SHOULD BE MEASURED?

Human resources

1. Lost time due to accidents, absenteeism.
2. Employee turnover.

3. Employee satisfaction index.

4. Training cost per employee.

5. Number of grievances.

Customers

1. Number of complaints from customers.

2. Number of on-time deliveries.

3. Warranty data.

4. Dealer satisfaction.

Production

1. Inventory.

2. SPC Charts.

3. Amount of scrap / rework.

4. Machine down time.

Research and Development

1. New product time to market.

2. Design change orders.

3. Cost estimating errors.

Suppliers

1. On-time delivery.

2. Service rating.

3. Quality performance.

4. Average lead time.

Marketing / Sales

1. Sales expense to revenue.

2. New product sales to total sales.

3. New customers.

Administration

1. Revenue per employee.
2. Purchase order error.
3. Billing accuracy.
4. Cost of poor quality.

STRATEGY :

The quality council has the overall responsibility for the performance measures.It ensures that all the measures are integrated into a total system of measures.

A typical system contains the following function

¬Quality

¬Cost

¬Flexibility

¬Reliability

¬Innovation

PERFORMANCE MEASURE PRESENTATION :

There are six basic techniques for presenting performance measures. They are

1. Time series graph.
2. Control charts.
3. Capability Index.
4. Taguchi's loss function.
5. Cost of poor quality.
6. Malcolm Baldrige National Quality Award.

In MBNQA, five categories are analyzed. They are

a) Manufacturing

b) Service

c) Small business

d) Health care

e) Education

UNIT V
QUALITY SYSTEMS

ISO (International organisation for standardisation)

Definition:

"The quality system is the organisation structure, responsibilities, procedures, process and resources for implementing quality management."

- ✓The quality system should function in such a manner as to provide proper confidence that :
  1. The system is well understood and effective
  2. The products or services actually do satisfy customers expectations.
- ✓In short a quality system involves:

  How- Methods and process description

  Who- Responsibilities and authorities

  When- Records and evidence

  Where- identification and traceability

ISO 9000

The ISO 9000 family is primary concerned " quality management". This means what the organisation does to fulfill.

- The customers quality requirement
- Applicable regulatory requirements
- Enhance customers satisfaction
- Continuous improvement

Need For ISO 9000

Every organisation is concerned with the quality of its product or service.

Reasons for implementing a quality system:

✓Improved employee involvement

✓Improved housekeeping

✓Improved decision making

✓Improved customers satisfaction

✓Improved safe working

✓Reduced customer complaints

✓Reduced quality cost

ISO 9000 Quality system

The ISO 9000 standards are a collection of formal international standards, technical reports, had books and web based documents on quality management and quality assurance

ISO technical committee and it sub committees are responsible for the development of the standards

ELEMENTS OR CAUSES OF ISO 9000

- ✓ Management responsibility
- ✓ Contract review
- ✓ Design control
- ✓ Document control
- ✓ Purchasing
- ✓ Purchaser supplied product
- ✓ Product identification and traceability
- ✓ Process control

- ✓ Inspection and testing
- ✓ Measuring and test equipment
- ✓ Inspection and test status
- ✓ Control of non- conforming products
- ✓ Corrective and preventive action
- ✓ Handling, storage, packaging and delivery
- ✓ Control of quality records
- ✓ Internal quality audits
- ✓ Servicing
- ✓ Stastical techniques

ISO 9001:2008 Quality Management System

- ✓ ISO 9001:2008 was developed in order to: – Introduce clarifications to the existing requirements of ISO 9001:2000 and Introduce changes that are intended to improve compatibility with ISO 14001:2004, Environmental Management Systems.
- ✓ ISO 9001:2008 Quality Management System certification enables you to demonstrate your commitment to quality and customer satisfaction, as well as continuously improving your company's operations.

- ✓ It includes all requirements for quality management of the product or service such as a quality planning requirement along with policies, objectives and quantifiable targets.
- ✓ It is an international standard that defines good management practices.
- ✓ Provides a global standard that spells out quality and trust.
- ✓ The purpose of ISO 9001:2008 is to facilitate international trade by providing a single standard that is recognized and respected globally.
- ✓ ISO 9001:2008 is important because of its international acceptance. It is recognized in more than 120 countries and is the logical choice for any organization involved in international business or committed to quality.

The 8 Quality Management Principles

1. Customer Focus
2. Leadership
3. Involvement of People
4. Process approach
5. Systems approach to management
6. Continual Improvement
7. Factual approach to decision making
8. Mutually beneficial supplier relationships

Steps To Registration Of Quality System

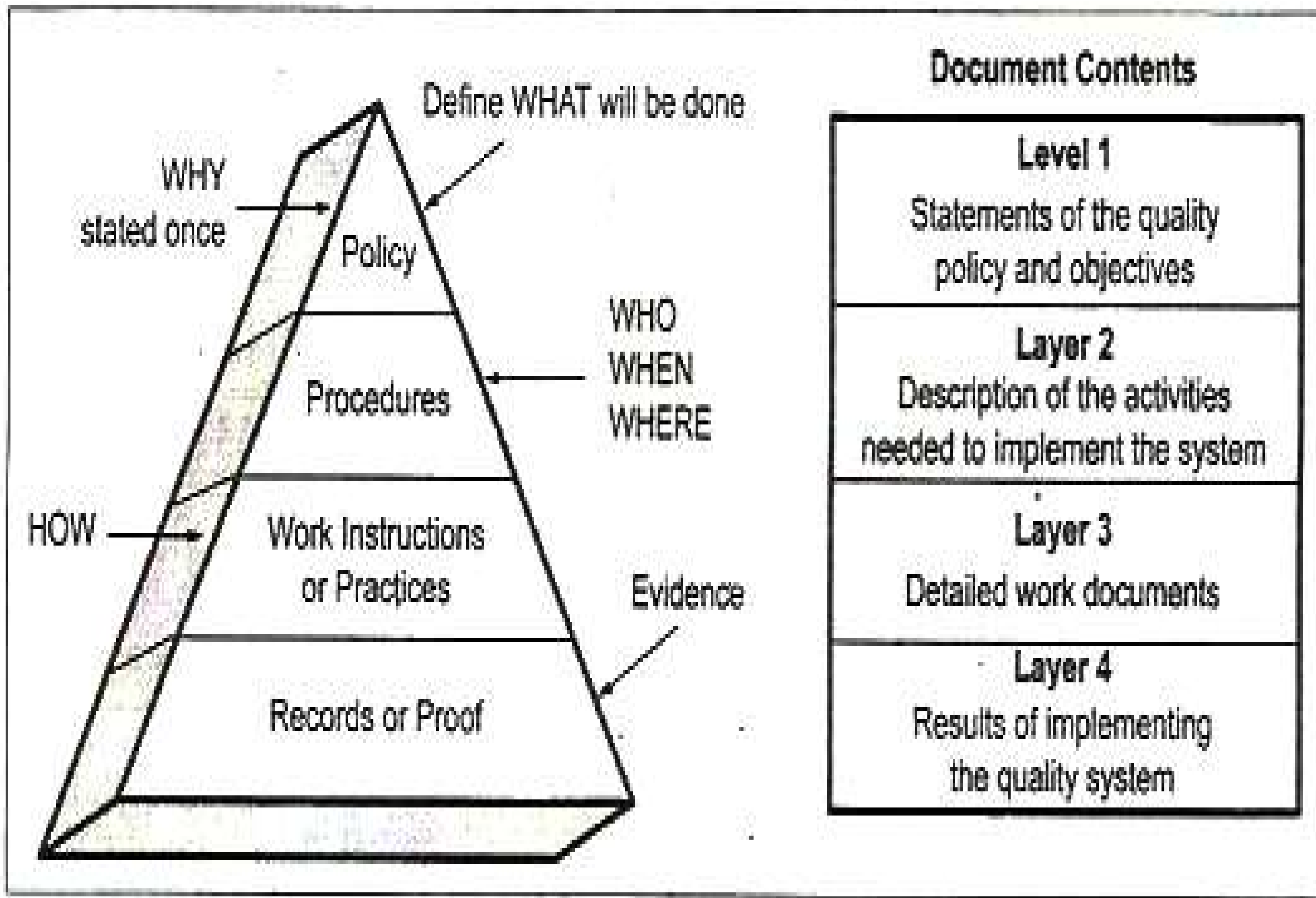
- ✓ Top management commitment
- ✓ Appoint the management representative
- ✓ Awareness
- ✓ Appoint an implementation team
- ✓ Training
- ✓ Time schedule
- ✓ Select element owners

- ✓ Review the present system
- ✓ Write the documents
- ✓ Install the new system
- ✓ Internal audit
- ✓ Management review
- ✓ Registration
- ✓ Award of ISO 9000 certificate

Documentation Of Quality System

The documents required by most organisations are:

- ✓Quality policy manual
- ✓Quality system procedures
- ✓Work instructions
- ✓Record/Formats/Forms


Define WHAT will be done
WHY
stated once
Policy
Procedures
WHO
WHEN
WHERE
HOW
Work Instructions
or Practices
Evidence
Records or Proof
Document Contents
Level 1
Statements of the quality
policy and objectives
Layer 2
Description of the activities
needed to implement the system
Layer 3
Detailed work documents
Layer 4
Results of implementing
the quality system

Quality Auditing

- ✓ Quality auditing should be carried out in order to verify whether a quality system is effective and suitable

Definition:

A quality system audit is define as " a systematic and in dependent examination to determine whether quality activities and related results comply with planned arrangements, whether these arrangement are implemented effectively and whether these are suitable to achieve objectives."

PURPOSE OF QUALITY AUDIT :

- To establish the adequacy of the system.
- To determine the effectiveness of the system.
- To afford opportunities for system analysis.
- To help in problem solving.
- To make decision making easier etc.

Features Of Quality Audits

- ✓The quality audit typically applies to quality system or elements such as process, products or service.
- ✓Quality audits are carried by staff who are not directly responsible in the areas being audited

Types Of Audits

- ✓ First party audit or internal audit: Where the auditee is its own client
- ✓ Second party audit: This refers to audit by one organisation on another organisation
- ✓ Third party audit: This refers to audit by an independent organisation on a supplier, for accreditation assessment purpose

Objectives Of Quality Audits

- ✓To determine the conformity or non-conformity of the quality system elements with regard to specified requirements.
- ✓To determine the effectiveness of the implemented quality system in meeting specified quality objective
- ✓To meet regulatory, if applicable
- ✓To evaluate an organisation's own quality system against a quality system standard

Stages Of Audit

1. Audit planning
2. Audit performance
3. Audit reporting
4. Audit follow-up

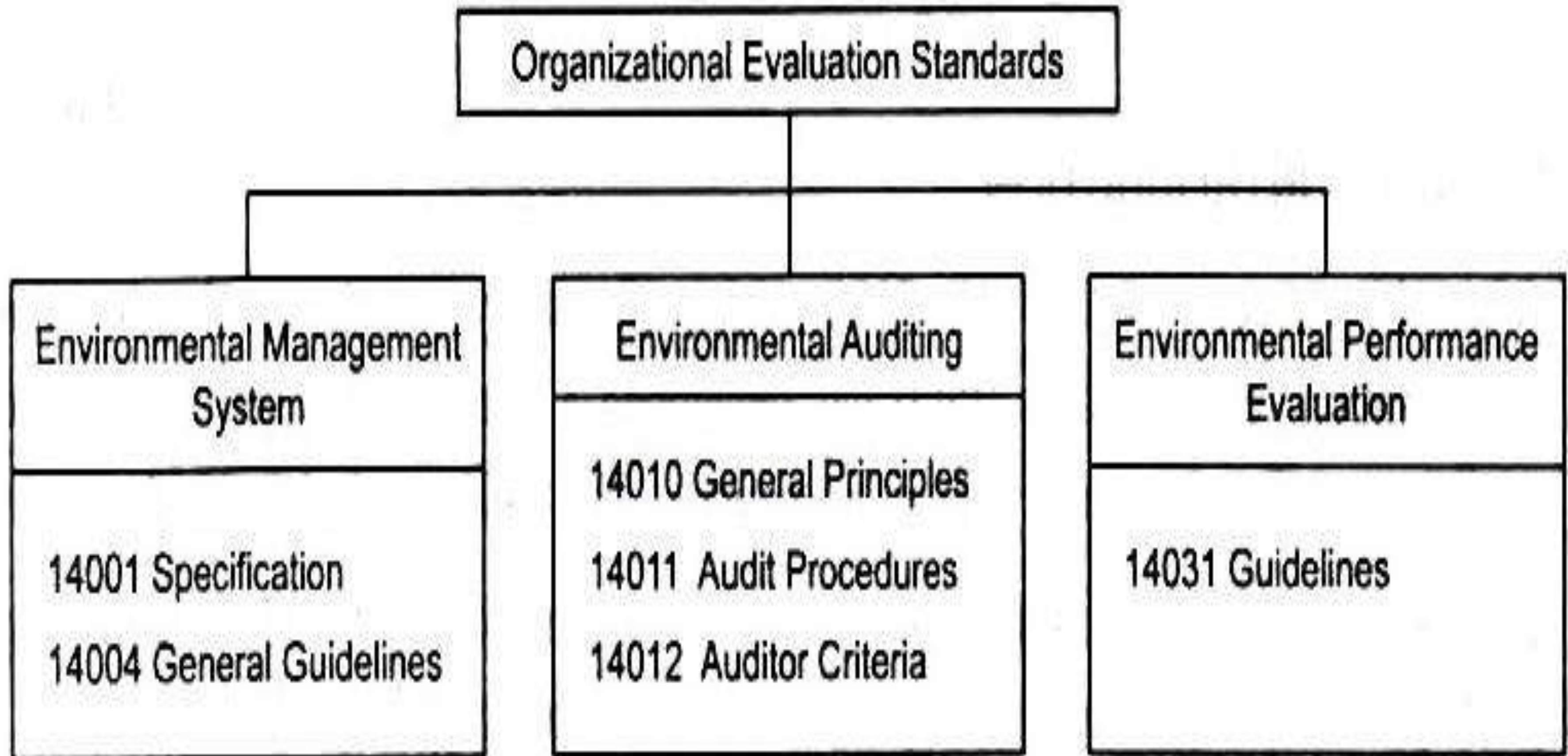
ISO 14000

The ISO 14000 standards are a set of norms for environmental management system either at organisation and process level or product level

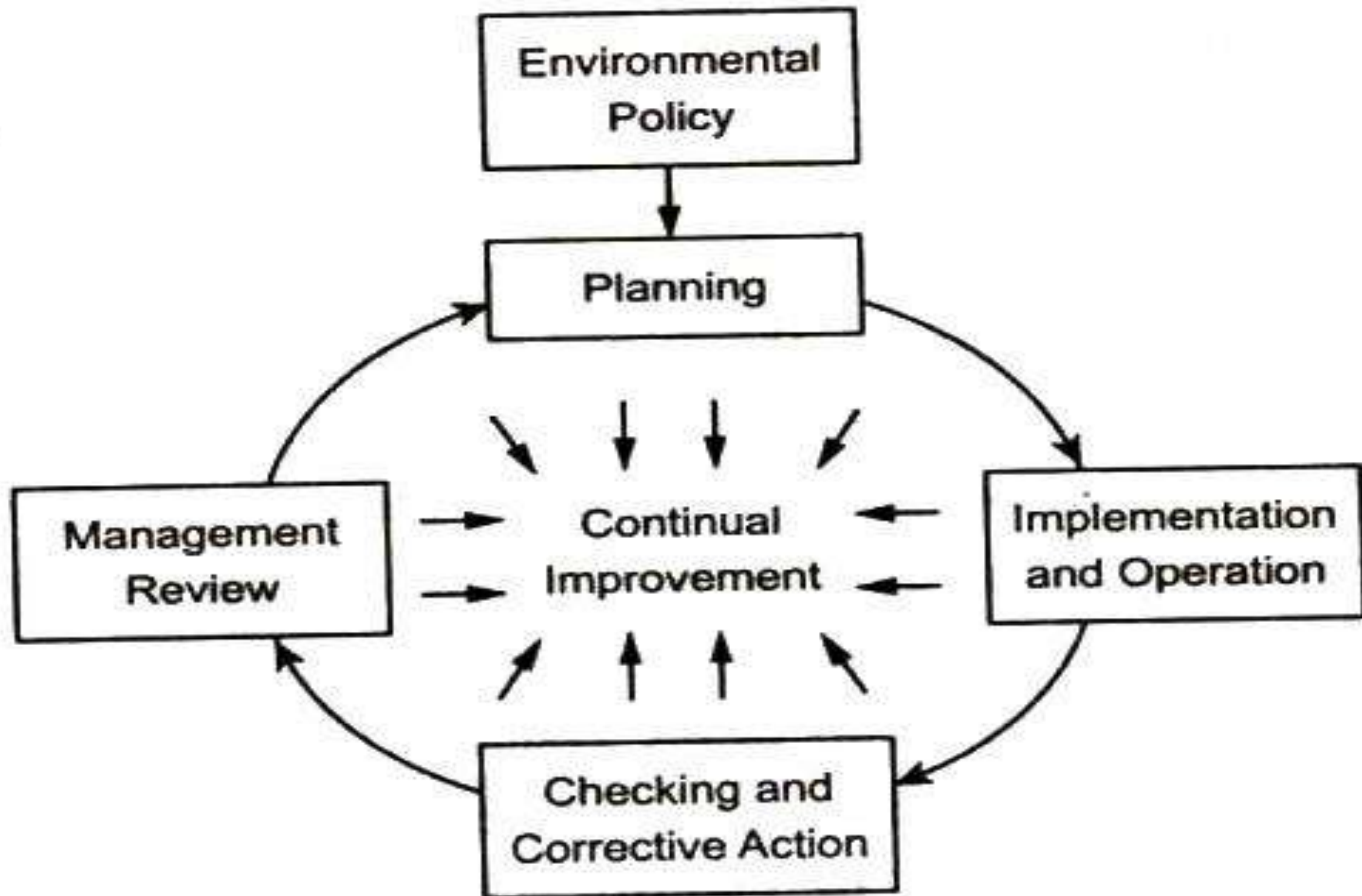
The series of ISO14000 standards are designed to cover:

i. Environmental management system
ii. Environmental auditing
iii. Environmental performance evaluation
iv. Life- cycle assessment
v. Environmental aspects in product standards

Organisation Evaluation Standards



Concepts Of ISO14000



Requirements Of ISO 14000

- ✓General requirements
- ✓Environmental policy
- ✓Environmental aspects
- ✓Legal aspects
- ✓Objective and targets
- ✓Environmental management program
- ✓Structure and responsibility

- ✓Training and awareness program
- ✓Communications
- ✓Document control
- ✓Operational control
- ✓Monitoring and measuring
- ✓Recording
- ✓Environmental management system audit
- ✓Management review

Benefits of ISO14000

The benefits of ISo14000 quality system can be viewed in two categories

1. Global benefits
   - ✓ To facilitate trade and remove trade barriers
   - ✓ To improve environmental performance of plant earth
2. Organizational benefits
   - ✓ Satisfying customers requirement
   - ✓ Improving team work
   - ✓ Improving industry/ government relations
   - ✓ Maintaining a good relationship with public /community / society

Importance of TQM for services sector

- ✓ Total Quality Management is the key mantra for the manufacturing industry, but its benefits have been better realized by intense customer-oriented service industries — be it fast moving consumer goods (FMCG), retail, hospitality, telecom or banking.
- ✓ In service organisations, the TQM challenge lies in establishing smooth connectivity between business processes so as to retain the customer.
- ✓ A quality control approach to cover all processes would be beneficiaspect oal to every Organisation.. Since in a service industry every fdepartment has a key and quality is associated with every employee, quality Control a very important role to play.

- ✓ Putting in place an effective TQM mechanism in a service industry requires patience and commitment on the part of the management and the workforce to satisfy the customer.
- ✓ Over 60 % of the organization's future revenue will come from the existing customers.
- ✓ A 2 percent increase in customer retention has an equivalent impact upon profitability as a 10 percent reduction in operating costs.
- ✓ Upto 96% of unhappy customers do not in fact complain, but they are three times more likely to communicate a bad experience to other customers than a good one .
- ✓ If a customer complains and the organization responds effectively to the product or service failure, then the loyalty of the customer can actually increase
- ✓ It costs 5 times as much to attract a new customer as it costs to keep an old one.

CASE STUDY