UNIT I:

LESSON 1

OPERATIONAL AMPLIFIERS I

Objectives:

- to explain the basic differential amplifier
- Introducing the operational amplifier
- To discuss the properties of operational amplifier.

Structure of the lesson :

1.1 Differential Amplifier

The differential amplifier, also called a difference amplifier, as the name implies, amplifies the difference between two signals. Because of its balanced nature and symmetry, it can amplify very small signals. It usually requires a minimum number of capacitors and can operate without bypass and coupling capacitors. It is the basic building block of operational amplifiers, which are most widely used in integrated circuits.

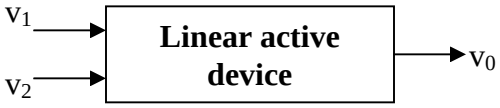


Fig 1.1 Schematic diagram of a differential amplifier.

For a linear active device with two input signals v1 and v2 and the out put signals v0 each measured with respect to ground, we have

$$v_0 = A\,(v_1 - v_2) \text{------} (1.1)$$

where A is the voltage gain of the differential amplifier. In actual practice, the output depends not only upon the difference of the two input signals but also upon the average level. In symmetrical circuits we

talk about the in-phase signals (called common mode (CM) Signals v_c) and the difference or anti-phase signals (called differential mode (DM) signals v_d). They are defined as

$v_c = \frac{1}{2}(v_1+v_2)$, and $v_d = (v_1 - v_2)$---------(1.2)

The output v_0 can be expressed as linear combination of the two input voltages, as

$v_0=A_1 v_1 + A_2 v_2$--------- (1.3)

Where A_1 and A_2 are the voltage amplifications from input 1and 2 respectively. From equations (1.2) and (1.3) we get

$$v_o = A_1(v_c + \frac{1}{2} v_d) + A_2(v_c - \frac{1}{2} v_d) = (A_1 + A_2)v_c + \frac{1}{2}(A_1 - A_2) v_d = A_c v_c + A_d v_d \text{----------(1.4)}$$

where $A_c = A_1 + A_2$ and $A_d = \frac{1}{2}(A_1 - A_2)$, are voltage gains for the signals in common mode and differential modes respectively. They may be defined as

$$A_d = \left(\frac{v_0}{v_d}\right)_{V_C = 0} \text{ and } A_c = \left(\frac{v_0}{v_c}\right)_{V_d = 0} \text{ ----------(1.5)}$$

Thus A_d can be measured directly by setting $v_c = 0$, or $v_2 = - v_1$.

The A_c can be measured by setting $v_d = 0$, or $v_2 = v_1$, generally the desired signals in differential amplifier are DM and undesired signals are CM. The figure of merit for differential amplifier in defined as

$$\rho = \left[\frac{A_d}{A_c}\right] \text{ ----------(1.6)}$$

This is called the common-mode rejection ratio (CMRR) and is also some times referred to as the discrimination factor of a differential amplifier. Ideally A_c=0, and CMRR= $\propto$. In practice, A/c is non-zero but very small, where as A_d is very large. The combination of equations (1.4) and (1.6) gives

$$Vo = A_d v_d \left[1 + \frac{A_c v_c}{A_d v_d}\right] = A_d v_d \left[1 + \frac{v_c}{v_d}.\frac{1}{CMRR}\right]$$

$V_0 = A_d v_d$ (since CMRR = $\propto$) ----------(1.7)

A basic differential amplifier circuit (ckt), consisting of two interlocked common emitter amplifier stages is shown in fig (1.2). The two stages are linked by having both emitters connected to a constant current generator. As current through one emitter increases, current through the other decreases.

The circuit is symmetric about the vertical dashed line and the two transistors and resistors R_c form a bridge circuit that is balanced under zero input signal. The resistors and the transistors are simultaneously fabricated in adjacent areas on a small chip. They will be at the same temperature. A simultaneous change in h_{FE} or v_{BE} will produce equal changes in the voltages at a and b and v_0 will not be affected.

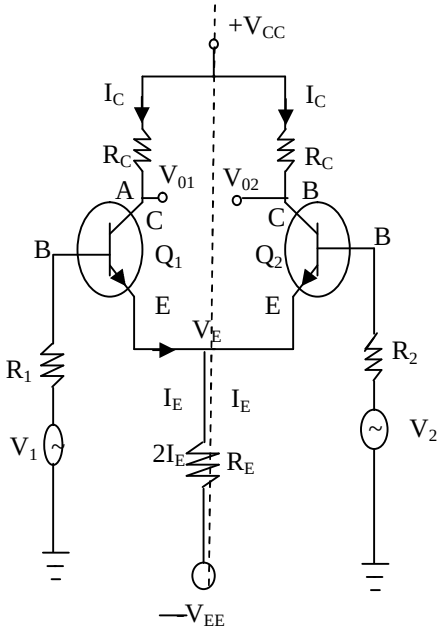


Fig (1.2) Basic differential amplifier circuit diagram

Consider the circuit operation with no input signals. For $v_1=v_2=0$, an emitter current I_E flows in each BJT. Therefore $I_C = I_E$ and

$$V_{01} = V_{02} = V_{CC} - I_C R_C \qquad \text{----------(1.8)}$$

$$\text{thus the base current } I_B = \frac{I_E}{h_{FE}} \qquad \text{----------(1.9)}$$

$$\text{and } V_E = -I_B R_1 - V_{BE} \qquad \text{----------(1.10)}$$

If v_{CE} is chosen large enough to bias each BJT in the center or the linear operating region, then

$$V_{CC} = V_{EE} + 2 I_E R_E + V_{CE} + I_C R_C \qquad \text{----------(1.11)}$$

The ac equivalent circuit for use differential amplifier is shown in fig (1.3). Here it is assumed

$h_{oe} R_C << 1$ or $h_{oe} << 1/R_C$ and thus h_{oe} is omitted in this figure. The collector current $I_c = h_{fe} I_b$. The voltage $h_{re} v_c$ is neglected in comparison with the $h_{ib} I_b$ drop across h_{ie}

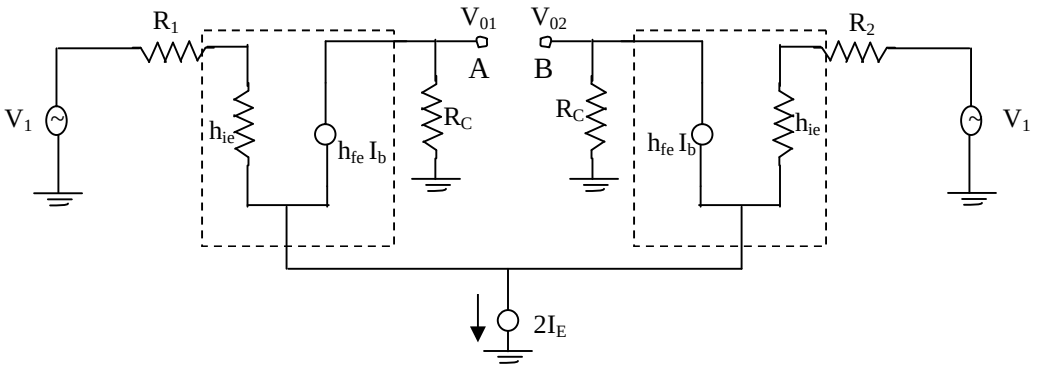


Fig (1.3) AC equivalent circuit of basic differential amplifier

Common mode voltage gain :

Let $v_1 = v_2 = v_s$, user $A_c = \frac{v_o}{v_c} = \frac{v_o}{v_s}$.

Due to symmetry, each input at the base, sees a common emitter circuit, with an un bypassed emitter resister of $2R_E$ (the emitter resistor is effectively doubled, as it carries the emitter current for both transistors). Thus we have

$$Z_i = R_1 + h_{ie} + 2R_E (1+h_{fe}) \text{ and } Z_0 = R_c \quad \text{----------(1.12)}$$

$$\text{Current gain} = A_i = \frac{-I_c}{I_b} = \frac{-h_{fe} I_b}{I_b} = -h_{fe} \quad \text{----------(1.13)}$$

$$\text{Voltage gain } A_v = \frac{A_i Z_0}{Z_i} = \frac{-h_{fe} R_c}{(R_1 + h_{ie} + 2R_E(1+h_{fe}))} \quad \text{----------(1.14)}$$

Since usually $(1+h_{fe}) 2R_E >> h_{ie}$ and $1+h_{fe} = h_{fe}$. The source resistance $R_i << h_{ie}$. Therefore

$$A_v = \frac{-R_C}{2R_E} = \text{common mode voltage gain } A_c.$$

Differential mode voltage gain:-

Let $-v_2 = v_1 = \frac{v_s}{2}$;

Therefore $v_d = v_1$, $-v_2 = v_s$ and $A_d = \frac{v_0}{v_d} = \frac{v_0}{v_s}$. From the symmetry of fig(1.2) for $v_1 = -v_2$,

the emitter of each transistor is grounded for small signal operation i.e $R_E=0$ and

$$Z_i = 2(R_i + h_{ie}),\ Z_o \cong R_c,\ A_i \cong -h_{fe}. \qquad \text{----------(1.15)}$$

$$A_v = \frac{A_i Z_0}{Z_i} = \frac{-h_{fe} R_c}{2(R_i + h_{ie})} \qquad \text{----------(1.16)}$$

= differential mode voltage gain A_d.

Common mode rejection ratio:-

$$\text{CMRR} = \frac{A_d}{A_c} = \frac{-h_{fe} R_c}{2(R_i + h_{ie})} \div \left(-\frac{R_c}{2R_E} \right) = \frac{-h_{fe} R_e}{(R_1 + h_{ie})} \qquad \text{---------(1.17)}$$

Thus we see that CMRR increases with R_E as desirable.

Constant current generators:

Instead of resistor R_E, a constant current generator is used. It may be a JFET with its gate tied to its source. A BJT can be used in a similar way with voltage divider bias. In both cases the output current is approximately constant as long as the voltage across the device is sufficient.

Constant current bias:-

In the differential amplifier discussed so far, the combination of R_E & V_{EE} is used to set up the dc emitter current. We can also use constant bias to set up the dc emitter current if desired. In fact the constant current bias is better because it provides current stabilization and in turn, assures a stable operating point for the differential amplifier. Fig(1.4). shows the dual input balanced output differential amplifier using a resistive constant current bias. Notice that the resistor R_E is replaced by a constant current source transistor (Q3) circuit. The dc collector current in the transistor Q_3 is established by resisters R_1 ,R_2 and R_E and can be determined as follows. Applying the voltage divider rule, the voltage at the base of transistor Q_3 (neglecting base loading effect) is

$$V_{B3} = \frac{-R_2 V_{EE}}{(R_1 + R2)} \qquad \text{----------(1.18)}$$

$$V_{E3} = V_{B3} - V_{BE3} = \left(\frac{-R_2 V_{EE}}{(R_1 + R2)} \right) - V_{BE3} \qquad \text{----------(1.19)}$$

$$I_{E3} = I_{C3} = \frac{(V_{E3} - (-V_{EE}))}{R_E} \qquad \text{----------(1.20)}$$

$$I_{C3} = v_{EE} - \left(\frac{-R_2 V_{EE}}{(R_1 + R2)} \right) . \frac{V_{BE3}}{R_E} \quad \text{----------(1.21)}$$

Because two halves of the differential amplifier are symmetrical, each has half of current I_{ce3}.

That is

$$I_{E1} = I_{E2} = \frac{I_{C3}}{2} = V_{EE} - \left(\frac{R_2 V_{EE}}{(R_1 + R2)} \right) . \frac{V_{BE3}}{2R_E} \quad \text{----------(1.22)}$$

The collector current I_{c3} in transistor Q_3 is fixed and must be invariant because no signal is injected into either the emitter or the base of Q_3. Thus the transistor Q_3 is a source of constant emitter current for transistors Q_1 and Q_2 of the differential amplifier. Besides supplying constant emitter current, the constant current bias also provides a very high source resistance since the ac equivalent of the dc current source is ideally an open circuit.

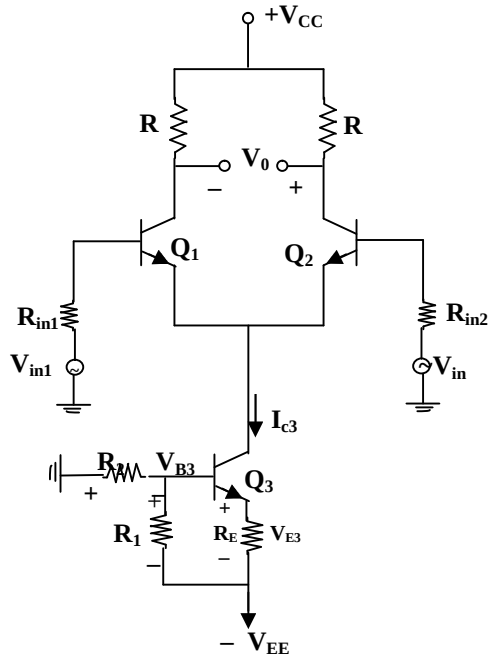


Fig (1.4) Differential amplifier using constant current bias.

1.2 Operational amplifier:

An operational amplifier is a direct-coupled high gain amplifier usually consists of one or more differential amplifiers and usually followed by a level translator and an output stage. The output

stage is generally a push pull or push pull complementary symmetry pair. An operational amplifier is available as a single integrated circuit package. The operational amplifier is a versatile device that can be used to amplify dc as well as ac input signals and was originally designed for computing such mathematical functions as addition, subtraction, multiplication and integration. Thus the name operational amplifier stems from its original use for doing these mathematical operations and so is abbreviated to op-amp. With the addition of suitable external feed back component the modern day operational amplifier can be used for a variety of applications. Such as ac and dc signal amplification, active filters, oscillators, comparators, regulators and others.

Block diagram representation of a typical op-amp:

Since an op-amp is a multistage amplifier. It can be represented by a block diagram shown in figure (1.5)

The input stage is the dual input balanced output differential amplifier. This stage generally provides most of the voltage gain of the amplifier and also establishes the input resistance of the op-amp. The intermediate stage is usually another differential amplifier, which is driven by the output of the first stage. In most amplifiers, the intermediate stage is dual input unbalanced (single ended) output. Because direct coupling is used, the dc voltage at the output of the intermediate stage is well above ground potential. Therefore, the level translator circuit is used after the intermediate stage to shift the dc level at the output of the intermediate stage downward to zero voltage with respect to ground. The final stage is usually a push pull complementary amplifier output stage. The output stage increases the output voltage swing and raises the current supplying capabilities of the op-amp. The well designed output stage also provides low output resistance.

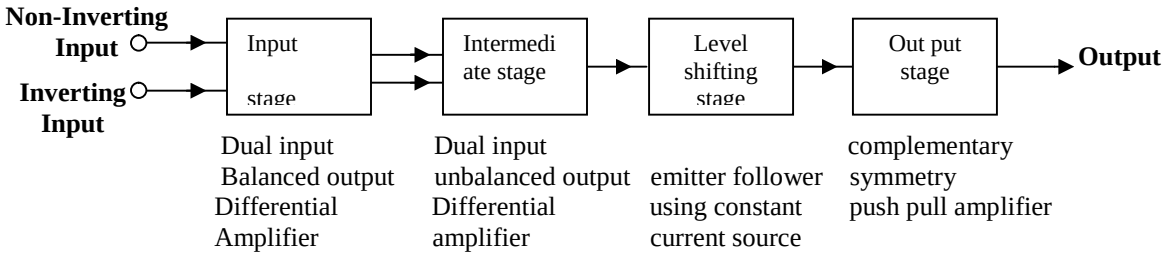


Fig (1.5) Block diagram of a typical op-amp

Schematic symbol:

The most widely used symbol for a circuit with two inputs and one out put is shown in fig (1.6)

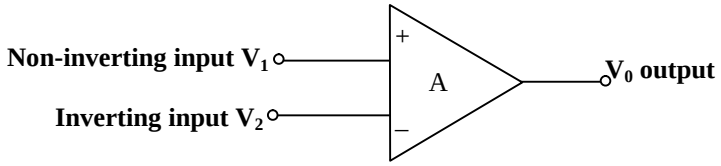


Fig (1.6) Schematic symbol of op-amp

In fig (1.6)

v_1 = voltage at the non-inverting input (volts)

v_2 = voltage at the inverting input (volts)

v_o = output voltage (voltage)

All these are measured w.r.t. ground

A = large signal voltage gain that is specified on the data sheets for an op-amp

For amplifier, power supply and other pin connections are omitted. Since the input differential amplifier stage of the op-amp is designed to be operated in the differential mode, the differential inputs are designated by the (+) and (-) notations, the (+) input is used for non-inverting input. An ac signal (or dc voltage) applied to this input produces an in-phase (or same polarity) signal at the output. On the other hand the (-) input is the inverting input because an ac signal (or dc voltage) applied to this input produces an 180 out of phase (or opposite polarity) signal at the output.

Ideal op-amp:

An ideal op-amp exhibits the following electrical characteristics.

(1) Infinite voltage gain A_V.

(2) Infinite input resistance R_i, so that, almost any signal source can drive it and there is no loading of the preceding stage.

(3) Zero output resistance R_o, so that, output can drive an infinite number of other devices

(4) Zero output voltage when the input voltage is zero.

(5) Infinite bandwidth, so that, any frequency signal from 0 to $\propto$ Hz can be amplified with out attenuation.

(6) Infinite common mode rejection ratio so that output common mode noise voltage is zero.

(7) Infinite slew rate so that voltage changes occur simultaneously with input voltage changes.

There are practical op-amps that can be made to achieve some of these characteristic using a negative feedback arrangement. In particular, the input resistance, the output resistance, and bandwidth can be brought close to ideal values by this method.

Equivalent circuit of an op-amp: Fig (1.7) shows an equivalent circuit of an op-amp. The circuit includes important values from the data sheets: A , R_i and R_o.

Note that A v_{id} is an equivalent Thevenin voltage source, and R_o is the Thevenin equivalent resistance looking back into the output terminal of an op-amp.

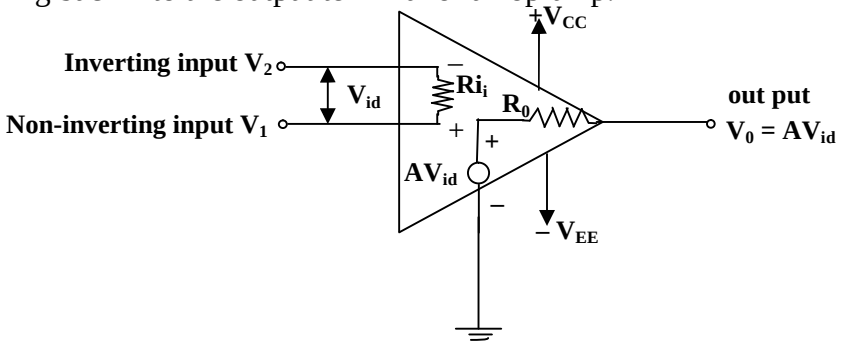


Fig (1.7) Equivalent circuit of op-amp

The equivalent circuit is useful in analyzing the basic operating principles of op-amps and in observing the effectiveness of feed back arrangement. For the circuit shown in fig (1.7), the output voltage is

$$v_0 = Av_{id} = A\,(v_1 - v_2) \qquad \text{----------(1.23)}$$

Where A = large- signal voltage gain

v_{id} = difference input voltage

v_1 = voltage at the non-inverting input terminal w.r.t. ground.

v_2 = voltage at the inverting input terminal w.r.t. ground.

Eqn.(1.23) indicates that the output voltage v_0 is directly proportional to the algebraic difference between the two input voltages. In other words the op-amp amplifies the difference between the two input voltages; it does not amplify the input signal voltages themselves. For this reason the polarity of the output voltage depends upon the polarity of the difference voltage.

1.3 Properties of practical op-amp:

To enhance our understanding of op-amps, we need to define some parameters that appear on data sheet of practical op-amp.

(1) **Voltage gain**:

Fig (1.8) shows an idealized transfer characteristic

In the active region the slope of the curve is the differential mode voltage gain A_d is defined as

$A_d = \frac{\Delta v_0}{\Delta v_d}$, where $v_d = v_1 - v_2$, the differential mode input. A_d is very large ~ 10^6.

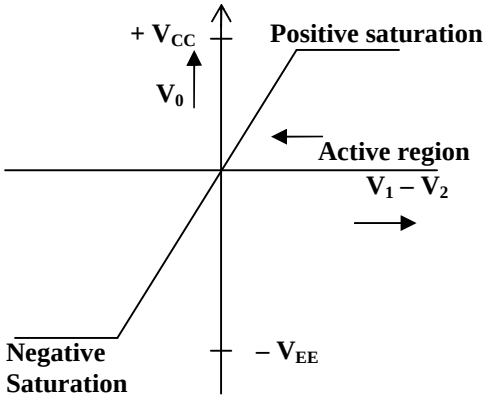


Fig (1.8) Idealized transfer characteristic of op-amp

To stabilize the voltage gain, negative feedback is always used. Fig (1.8) shows that the voltage gain is virtually zero when the output is at saturation level.

(2) Input impedance R_i:

It is the open loop incremental impedance looking into the two input terminals. It is large (~M Ω)

(3) Output impedance R_o:

It is the open loop impedance across the output. It is low (~ 100 Ω)

(4)): **Common mode rejection ratio (CMRR):**

The op-amp should ideally respond to a difference mode only. There will also be an output for the common mode input, because of the nature of the input circuit any (differential amplifier). The common mode gain $A_c = \frac{\Delta v_0}{\Delta v_c}$

The ratio of these two gains is defined as common mode rejection ratio

$$CMRR = \left|\frac{A_d}{A_c}\right| \qquad \text{----------} \quad (1.24)$$

It is usually expressed in decibels, as $20 \log_{10}(CMRR) = 20 \log_{10} |A_d| - 20 \log_{10} |A_c|$

It is having a value 70 dB to 100 dB

Op-amps are further classified into two groups: general- purpose and special purpose. General- purpose op-amps may be used for a variety of applications such as integrator, differentiator, summing amplifier and others. An example of a widely used general-purpose op-amp is the 741/351. On the other hand, special purpose op-amps are used only for the specific applications they are designed for. For example the LM 380 op-amp can be used only for audio power applications;

The pin configuration of most widely used 741 op-amp is given below

It is the most commonly used general-purpose op-amp. It has an integrated 30pF MOS capacitor. It has high input impedance (> MΩ), low output impedance (750 Ω) and large voltage gain (200,000). From here onwards all the discussions are confined to μA741 op-amp.

The electrical parameters of op-amp are defined in the following paragraphs.

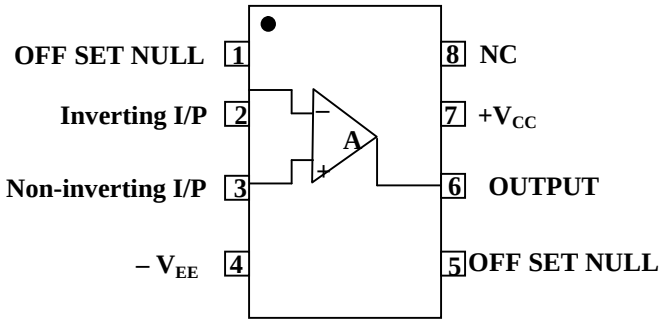


Fig (1.9) pin configuration of μA-741 op-amp

Input offset voltage: Input offset voltage is the voltage that must be applied between the two input terminals of an op-amp to null the output as shown in fig (1.10). In fig (1.10) V_{dc1} and V_{dc2} are dc voltages and R_s represents the source resistance. We denote input offset voltage by V_{i0}. This voltage V_{i0} could be positive or negative;

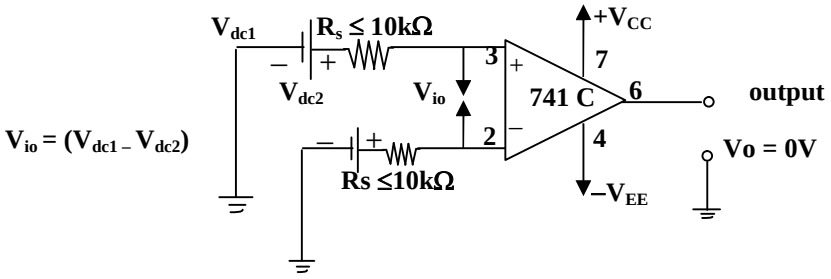


Fig (1.10) Defining input offset voltage V_{i0}

For 741C, the maximum value of V_{i0} is 6 mV DC. The smaller the value of v_{i0}, the better the input terminals are matched.

Input offset Current:

The algebraic difference between the currents into the inverting and non-inverting terminals is referred to as input offset current I_{i0} (see fig.1.11). In the form of an equation.

$$I_{i0} = |I_{B1} - I_{B2}|$$

Where $I_{b1}$is the current into the non-inverting input I_{b2} is the current into the inverting input

The input offset current for the 741C is 200 nA maximum. As the matching between the two input terminals is improved, the difference between I_{B1} and I_{B2} becomes smaller; that is, the I_{i0} value decreases further.

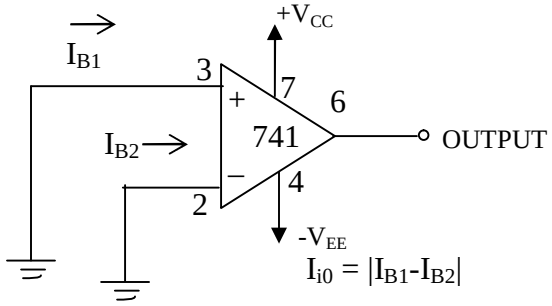


Fig (1.11) defining input offset current I_{i0}

Input Bias current:

Input bias current I_B is the average of the currents that flow into the inverting and non-inverting input terminals of the op-amp. In equation form

$$I_B = \frac{I_{B1} + I_{B2}}{2}$$

I_B = 500 nA maximum for the 741C, where as I_B for the precision 741C is ± 7 nA

Note: that the input currents I_{B1} and I_{B2} are actually the base currents of the first differential amplifier stage.

Differential input resistance:

Differential i/p resistance R_i (often referred to as i/p resistance) is the equivalent resistance that can be measured at either the inverting or non-inverting i/p terminal with the other terminal grounded. For the 741C the i/p resistance is relatively high 2 MΩ.

Input capacitance: Input capacitance C_i is the equivalent capacitance that can be measured at either the inverting or non-inverting terminal with the other terminal grounded. A typical value of c_i is 1.4 pf for the 741C. This parameter is not listed in all op-amp-data sheets.

Offset voltage adjustment ratio: One of the features of the 741 family op-amps is an offset voltage null capability. The 741 op-amps pins 1 and 5 marked as offset null are for this purpose. As shown in figure 1.11 a 10kΩ potentiometer can be connected between offset null pins 1 and 5, and the wiper of the potentiometer can be connected to the negative supply $-V_{EE}$.. By varying the potentiometer the output offset voltage can be reduced to zero volts. Thus the offset voltage adjustment range is the range through which the i/p offset voltage can be adjusted by varying the 10kΩ potentiometer. For the 741C is the offset voltage adjustment range is ±15mV. Very few op-amps have the offset voltage null capability. This means that for most op-amps , we have to design an offset voltage compensating net work in order to reduce the o/p offset voltage to zero.

Common mode rejection ratio: The common-mode rejection ratio (CMRR) is defined in several essentially equivalent ways by the various manufactures. Generally it can be defined as the ratio of the differential voltage gain A_d to the common mode voltage gain A_{cm}. That is

$$\text{CMRR} = \frac{A_d}{A_{cm}}$$

The differential voltage gain A_d is the same as the large signal voltage gain A, which is specified on the data sheets; The common mode voltage gain can be determined from the circuit.

$$A_{cm} = \frac{v_{ocm}}{v_{cm}}.$$

v_{0cm} = o/p common mode voltage.

v_{cm} = i/p common mode voltage.

A_{cm} = common mode voltage gain.

Generally the A_{cm} is very small and $A_d = A$ is very large; therefore the CMRR is very large, being a large value, CMRR is most often expressed is decibels (dB) for the 741C, CMRR is 90 dB typically.

Slew rate (S.R): Slew rate is defined as the maximum rate of change of output voltage per unit of time and is expressed in volts per microseconds. In equation form

$$S.R = \frac{dv_o}{dt} / \text{maximum } V/\mu s$$

Slew rate indicates, how rapidly the output of an op-amp can change in response to changes in the input frequency with input amplitude constant. The slew rate changes with change in voltage gain and is normally specified at unity (+1) gain. The slew rate of an op-amp is fixed, therefore if the slope requirements of the o/p signal are greater than the slew rate, distortion occurs. The slew rate is one of the important factors in selecting the op-amp for an application, particularly at relatively high frequencies. One of the drawback of the 741C is its low slew rate (0.5 V/μs).

Open-loop op-amp configurations:

In the case of amplifiers, the term "open-loop" indicates that no connections, either direct or via another network exists between the o/p and i/p terminals. That is, the o/p signal is not fed-back in any form as part of the input signal and the "loop" that would have been formed with feedback is open.

When connected in open loop configuration the op-amp simply functions as a high gain amplifier. There are three open loop op-amp configurations:

1. The differential amplifier.
2. The inverting amplifier.
3. The non-inverting amplifier.

The differential amplifier:

Fig 1. 12 shows the open loop differential amplifier in which input signals v_{is1} and v_{is2} are applied to the positive and negative input terminals. Since the op-amp amplifies the difference between the two input signals, this configuration is called the differential amplifier.

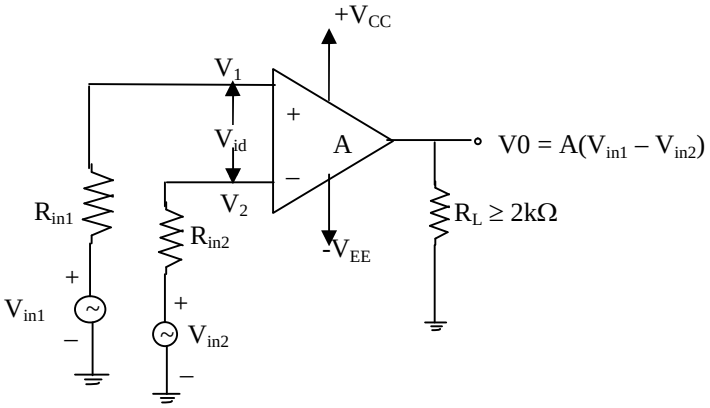


Fig 1.12 Open loop differential amplifier

The op-amp is a versatile device because it amplifies both ac and dc input signals. This means that v_{is1} and v_{is2} could be either ac or dc voltages. The source resistance R_{in1} and R_{in2} are normally negligible compared to the input resistance R_i. Therefore, the voltage drops across the resistors can be assumed to be zero, which then implies that $v_1 = v_{is1}$, and $v_2 = v_{is2}$. Substituting these values of v_1 and v_2 in equation for output voltage, we get

$v_0 = A(v_{in1}-v_{in2})$

Thus, as expected the output voltage is equal to voltage gain A times the difference between the two input voltages. Also notice that the polarity of the output voltage is dependent on the polarity of the input difference voltage ($v_{in1}-v_{in2}$). In open loop configurations gain A is commonly referred to as open loop gain.

The inverting amplifier:

In the inverting amplifier only one input is applied and that, to the inverting input terminal.

The non-inverting input terminal is grounded. Since $v_1 = 0$ and $v_2 = v_{in}$.

$V_0 = - AV_{in}$

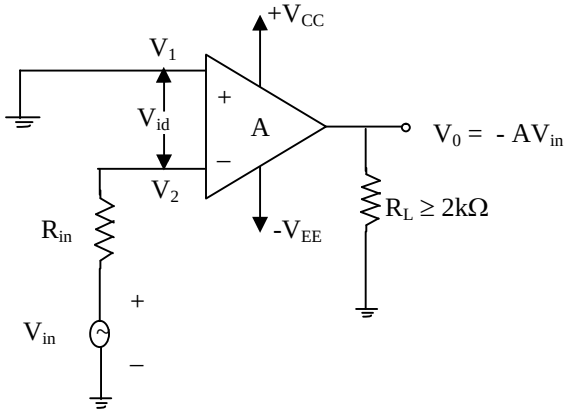


Fig 1.13 Inverting amplifier

The negative sign indicates that the output voltage is out of phase with respect to input by 180^0

or is of opposite polarity. Thus in the inverting amplifier the input signal is amplified by gain A and is also inverted at the output.

The non-inverting amplifier:

Fig 1.14 shows the open loop configurations of non inverting amplifier in this configuration the input is applied to the non inverting input terminal and the inverting input terminal is grounded.

In the circuit shown below $v_1=v_{in}$ and $v_2=0v$. Therefore according to equation $v_0 = Av_{in}$. This means that the output voltage is larger than the input voltage. By gain A and is in phase with input signal.

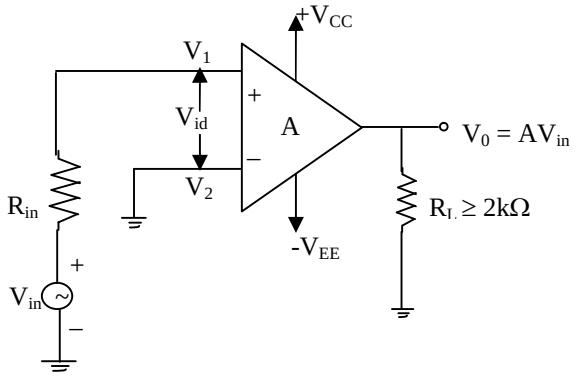


Fig 1.14 Non-inverting amplifier

In all the three open loop configurations any input signal that is only slightly greater than zero drives the output to saturation level. This results from the very high gain A of the op-amp. Thus when operated in open loop, the output of the op-amp is either at negative saturation or positive saturation or switches between positive and negative saturation levels. For this reason the open loop configurations are not used in linear applications.

1.4 Summary of the Lesson:

A differential amplifier consists of two symmetrical common emitter stages and is capable of amplifying the difference between two input signals. Differential amplifier is capable of amplifying ac as well as dc input signals because it employs direct coupling. For proper operation of the differential amplifier, a transistor array and matched components must be used. The differential amplifier can be biased by using emitter bias (a combination of R_E and V_{EE}), a

constant current bias. The constant current bias is more preferable in differential amplifiers because it maintains the infinite input impedance for all the desired values of emitter currents I_E.

An ideal op-amp has infinite voltage gain, input resistance, CMRR and slew rate together with zero output resistance and output offset voltage. The equivalent circuit of op-amp is useful in analyzing the basic operating principles of an op-amp and in observing the effects of feedback arrangements. The voltage transfer characteristic curve of an op-amp is the graph of output voltage versus the differential input voltage. Differential, inverting and non-inverting amplifiers are the three open loop op-amp configurations in which the output signal is not fed back in any form as part of the input signal. When operated in the open loop; generally the op-amp's output is either at positive or negative saturation or switches between positive and negative saturation levels. This action is undesirable in linear applications, hence the op-amp's are rarely used in open loop configuration for the linear applications.

1.5 Key terminology:

Configuration : design of the circuit by employing the various possible networks is called configuration.

Decibel (dB); 1/ tenth of a Bell; Bell is unit of power ratio

Milli volts (mV); 1/1000 of a volts

Nano Ampere (nA); $1nA = 10^{-9}$ ampere

Pico Farad (pF); $1pF = 10^{-12}$ Farad

Saturation; when input voltage in the circuit exceeds a certain value, the output goes to the maximum value is called saturation

Pot (Potentiometer); variable wire wound resistor

1.6 Self-assessment questions:

1. Explain the working of differential amplifier with a diagram.
2. What are common mode and differential mode gain of a differential amplifier?
3. Derive expressions for the common mode voltage gain and differential mode voltage gain of a transistor differential amplifier and show that the common mode rejection ratio increases with emitter resistance.
4. What is meant by constant current bias? Explain its use.
5. What are the important stages of a typical operational amplifier?
6. Mention the important electrical characteristics of an ideal op-amp.

7. How far IC 741 electrical characteristics are nearer to the ideal characteristic of op-amp.
8. Explain the terms input offset voltage, input offset current, input bias current.
9. Explain why op-amp is not used in open loop configuration for linear applications.

1.7 Reference Books:

(1) Operational Amplifiers and Linear Integrated Circuit Technology by Ramakanth A. Gaykwad Prentice Hall Inc.,

(2) Basic Electronics by DC Tayal, Himalaya Publish Co.,

(3) Semi Conductor Electronics by A K. Sharma
New Age International Publishers

(4) Foundations of Electronics
By D. Chattopadhyay, PC Rakshit, B. Saha, M.N. Purkait
Second Edition, Wiley Eastern Ltd.,

(5) Integrated Electronics By Millman & Halkias (MH)

Name of the Author

Dr D.Krishna Rao, M.Sc., M.Phil., Ph.D.,
Associate Professor,
Physics Department,
Nagarjuna University,
Nagarjuna Nagar 522 510.

UNIT I

LESSON 2

OPERATIONAL AMPLIFIERS -II

Objectives of the lesson:

1. To discuss the effect of negative feed back on
   i) closed loop voltage gain
   ii) input resistance
   iii) output resistance and
   iv) band-width of op-amp.
2. To discuss the applications of op-amp

Structure of the lesson:

2.1 Introduction :

As the open loop gain of op-amp is very high, only very small signals (of the order of micro-volts or less) having very low frequency may be amplified accurately with out distortion. However, signals this small are very susceptible to noise. Besides being large the open loop gain of the op-amp is not constant. The voltage gain varies with changes in temperature and power supply as well as with mass production techniques. The variations in voltage gain are relatively large in open-loop op-amp, in particular, which makes the open-loop op-amp unsuitable for many linear applications. In most linear applications the output is proportional to the input and is of the same type.

Further the bandwidth (band of frequencies for which the gain remains constant) of most open-loop op-amps is negligibly small - almost a zero. For this reason, the open-loop op-amp is

impractical in ac applications. For instance the open loop bandwidth of the 741C is approximately 5Hz. However, in almost all ac applications a bandwidth larger than 5 Hz is needed.

Because of the above stated reasons, the open loop op-amp is generally not used in linear applications. Never the less in certain applications the open-loop op-amp is purposely used as a nonlinear device; that is a square ware output is obtained by deliberately applying a relatively large input signal. Open-loop op-amp configurations are most suitable in such applications.

We will be able to select as well as control the gain of the op-amp, if we introduce a modification in the basic circuit. This modification involves the use of feedback, that is, an output signal is fed back to the input either directly or via another network. If the signal fed back is of opposite polarity or out of phase by 180^0 with respect to input signal, the feedback is called of negative feedback. An amplifier with negative feedback has a self - correcting ability against any change in output voltage caused by changes in environmental conditions. Negative feedback is also known as degenerative feedback because when used it degenerates (reduces) the output voltage amplitude and in turn reduces the output gain.

Suppose the signal fed back is in phase with the input signal, the feedback is called positive feedback. In positive feedback, the feed back signal aids the input signal. For this reason it is also known as regenerative feedback. Positive feedback is necessary in oscillator circuits.

Therefore negative fed back stabilizes the gain, increases the bandwidth and changes the input and output resistances, when used in amplifiers. The price paid for these improvements is reduced voltage gain. Other benefits of negative feedback include a decrease in harmonic distortion and reduction in effect of input offset voltage at the output. Negative feed back also reduces the effect of variation in temperature and supply voltages on the output of the op-amp.

2.2 Effect of negative feed back on closed loop op-amp

Amplifier performance can be altered considerably by sampling the output signal and sending a part of it to a mixer circuit where the feedback signal mixes with the input. Depending upon the phase relation ship between the input and feedback signals the resultant input can be less than the source voltage or more than that. If the effective input is less than source voltage it is called negative feed back. In this lesson you will learn the effect of negative feed back on various op amp parameters.

2.2.1 Block diagram representation of feedback configurations:

An op-amp that uses the feedback is called a feedback amplifier. A feedback amplifier is sometimes referred to as a closed loop amplifier because the feedback forms a closed loop between the input and output. A feedback amplifier essentially consists of two parts: an op-amp and a feed back circuit. The feedback circuit can take any form what so ever, depending on the intended application of the amplifier. It means, the feedback circuit may be made up of either passive components, active components, or combinations of both. This chapter, in order to develop the basic feedback concepts presents only purely resistive feedback circuits.

A closed loop amplifier can be represented by using two blocks, one for op-amp and another for a feedback circuit. There are four ways to connect these two blocks. These connections are classified according to whether the voltage or current is fed back to the input in series or parallel as follows:

1. Voltage- series feedback.
2. Voltage- shunt feedback
3. Current - series feedback
4. Current – shunt feedback

The four types of configurations are illustrated in fig (2.1). In fig (2.1)(a) and (b) the voltage across load resistor R_L is the input voltage to the feed back circuit. The feedback quantity (either voltage or current) is the output of the feedback circuit and is proportional to the output voltage. On the other hand in the current series and current shunt feedback circuits. [2.1 (c) and (d)] the load current i_L flows into the feedback circuit. The output of the feedback circuit is the feedback quantity which is proportional to the load current i_L.

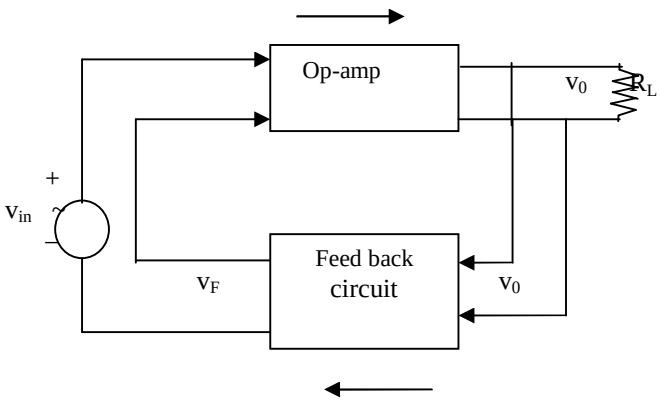


Fig (a): Voltage series feed back

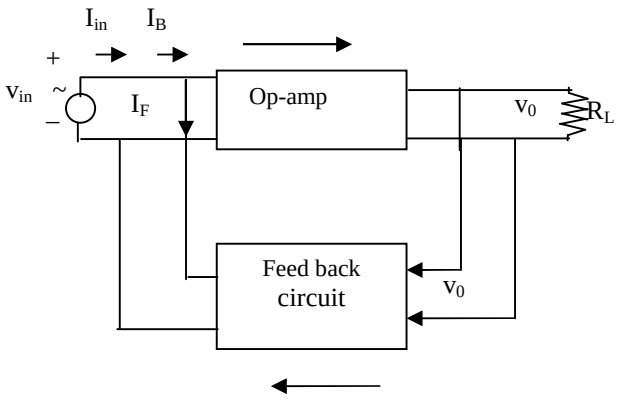


Fig (b): Voltage shunt feedback

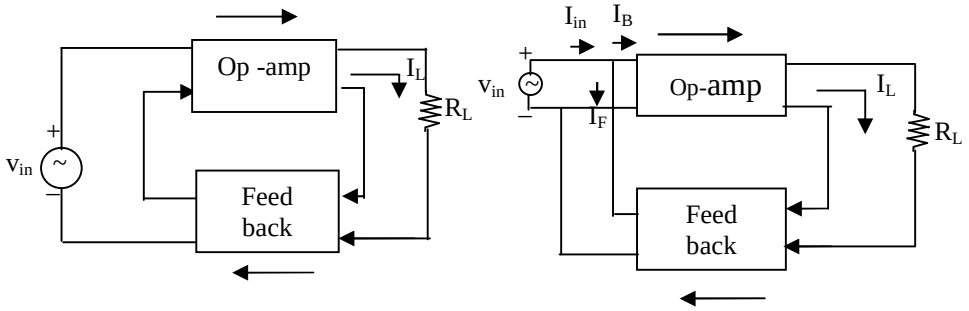


Fig (c): current series feed back **Fig (d): Current shunt feedback**

Fig 2.1. Block diagram represent to that of configurations.

Note that in all four of these configurations, the signal direction through the op-amp is from the input to output. On the other hand, in the ideal case, the signal direction through the feedback circuit is exactly opposite, from output to input.

The voltage-series and voltage - shunt feedback configurations are important because they are most commonly used. In the present chapter the analysis of voltage series feedback is discussed, through computing voltage gain, input resistance, output resistance and bandwidth.

2.2.2 Voltage series feedback amplifier:

The schematic diagram of the voltage series feed back amplifier is shown in fig (2.2). The op-amp is represented by its schematic symbol including the large signal voltage gain A, and the feedback circuit is composed of two resistances R_1 and R_F.

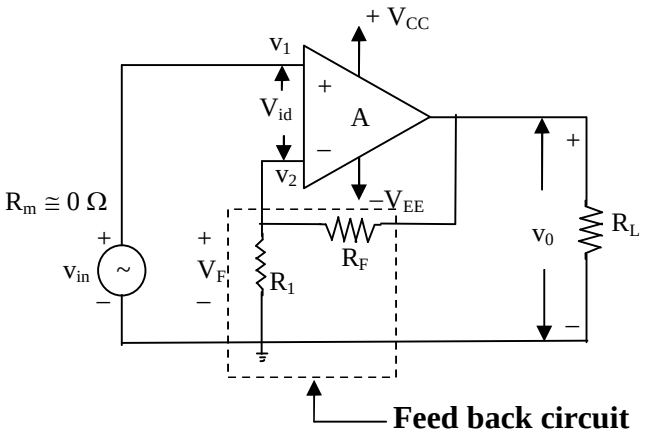


Fig (2.2): voltage series feed back amplifier.

The circuit shown in fig (2.2) is commonly known as a non- inverting amplifier with feedback (or closed loop non-inverting amplifier) because it uses feedback, and the input signal is applied to the non-inverting input terminal of the op-amp.

Before proceeding, it is necessary to define some important terms for the voltage series feed back amplifier of fig (2.2). Specially the voltage gain with and without feedback, and the gain of the feedback circuit are defined as follows.

Open loop voltage gain (or gain without feedback) $A = \frac{v_o}{v_{id}}$.

Closed –loop voltage gain (or gain with feedback) $A_f = \frac{v_o}{v_{in}}$

Gain of the feedback current $B = \frac{v_f}{v_o}$

2.2.3 Negative feedback:

Returning to the circuit of fig (2) Kirchoff's voltage equation for the input loop is

$v_{id} = v_{in} - v_f$ --------(2.1)

Where v_{in} = input voltage

V_f = feedback voltage

V_{id} =difference voltage

However, recall that an op-amp always amplifies the difference input voltage v_{id}. From eqn (2.1) the difference voltage is equal to input voltage v_{in} minus the feedback voltage v_f, in other words the feedback voltage opposes the input voltage (or is out of phase by 180^0 with respect to the input voltage). Hence the feedback is said to be negative. Returning to the analysis of voltage series feedback amplifier we should note that, it is performed by computing closed loop voltage gain, input resistance, output resistance, and the bandwidth.

2.2.4 Closed loop voltage gain:

As defined previously, the closed loop voltage gain

$$A_f = \frac{v_o}{v_{in}},$$

However by equation $v_0 = A(v_1 - v_2)$,

Referring to fig (2.2) we see that

$v_1 = v_{in}$

$v_2 = v_f = \dfrac{R_1 v_0}{R_1 + R_F}$ since $R_i >> R_1$

Therefore $v_0 = A\left(v_{in} - \dfrac{R_1 v_0}{R_1 + R_F}\right)$

Rearranging the above equation, we get

$$v_o = \frac{A(R_1 + R_F)\, v_{in}}{R_1 + R_F + AR_1}$$

$$\text{Thus } A_F = \frac{v_o}{v_{in}} = \frac{A(R_1 + R_F)}{R_1 + R_F + AR_1} \text{ (exact) ---------- (2.2)}$$

Generally, A is very large (typically 10^5), therefore,

$AR_1 >> R_1+R_F$ and $R_1+R_F+AR_1 \cong AR_1$

$$\text{Thus } A_F = \frac{v_o}{v_{in}} = 1 + \frac{R_F}{R_1} \text{ (ideal) --------- (2.3)}$$

Equation (2.3) is important because, it shows that the gain of a voltage series feed back amplifier is determined by the ratio of the two resistors, R_1 and R_F. For instance , if a gain of 11 is desired, we can then choose $R_1 = 1k\Omega$ and $R_F = 10k\Omega$ or $R_1 = 100\Omega$ and $R_F = 1k\Omega$.

Another interesting result can be obtained from equation (2.3). As defined previously the gain or the feed back circuit (B) is the ratio of v_f to v_0. Referring to fig (2.2), the gain is

$$B = \frac{v_f}{v_o}$$

$$B = \frac{R_1 v_0}{(R_1 + R_F) v_0} = \frac{R_1}{R_1 + R_F} \text{ ---------- (2.4)}$$

Comparing equations (2.3) and (2.4) we can conclude that

$$A_F = \frac{1}{B} \text{ (ideal) ---------- (2.5)}$$

This means that gain of the feed back circuit is the reciprocal of the closed loop voltage gain. In other words for given R_1 and R_F the values of A_F and B are fixed. Besides that equation (2.5) is an alternative to equation (2.3)

Finally the closed loop voltage gain A_F can be expressed in terms or open loop gain A and feedback circuit gain B by rearranging equation (2.2) in the following manner we get

$$A_F = \frac{A\left(\frac{R_1 + R_F}{R_1 + R_F}\right)}{\frac{R_1 + R_F}{R_1 + R_F} + \frac{AR_1}{R_1 + R_F}} \quad \text{----------} \quad (2.6)$$

Using equation (2.4) yields

$$A_F = \frac{A}{1 + AB} \quad \text{----------} \quad (2.7)$$

Where A_F = closed loop gain

A = open loop gain

B = gain of the feed back circuit

AB = "loop gain"

2.2.5 Input resistance with feedback:

A voltage series feed back amplifier with the op-amp equivalent circuit is shown in fig (2.3). In this circuit R_i is the input resistance (open-loop) of the op-amp, and R_{iF} is the input resistance of the feed back amplifier. The input resistance with feedback is defined as

$$R_{if} = \frac{v_{in}}{i_{in}} = \frac{v_{in}}{v_{id} / R_i} \quad \text{----------} \quad (2.8)$$

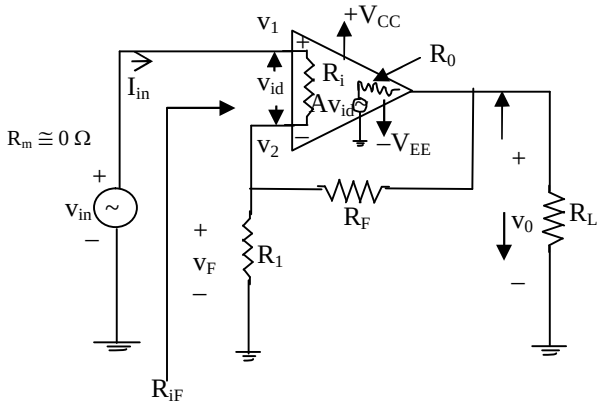


Fig 2.3 Derivation of input resistance with feedback.

However,

$$V_{id} = \frac{v_0}{A} \quad \text{and} \quad v_0 = \frac{A}{1 + AB} v_{in} \quad \text{----------} \quad (2.9)$$

$$R_{if} = R_i \cdot \frac{v_{in}}{v_0 / A} \quad \text{---------- (2.10)}$$

$$R_{if} = AR_i \frac{v_{in}}{\frac{Av_{in}}{(1+AB)}} = R_i (1+AB) \quad \text{---------- (2.11)}$$

It means, the input resistance of the op-amp with feedback is (1+AB) times that without feedback.

2.2.6 Output resistance with feedback:

Output resistance is the resistance determined looking back into the feedback amplifier from the output terminal, as shown in fig 2.4. This resistance can be obtained by using Thevenin's theorem for dependent sources. Specifically, to find output resistance with feed back R_{oF}, reduce the independent source v_{in} to zero, apply an external voltage v_0 and then calculate the resulting current i_o in short , the R_{oF} is defined as follows:

$R_{oF} = \frac{v_o}{i_o}$.

Writing Kirchoff currents equation at the output node N, we get

$$i_o = i_a + i_b \quad \text{----------(2.12)}$$

Since $R_F+R_1 \parallel R_i >> R_0$, $i_a >> i_b$ therefore

$$i_o \cong i_a \quad \text{----------(13)}$$

The current i_o can be found by writing Kirchoff's voltage equation for the output loop

$$v_0 - R_0 i_0 - Av_{id} = 0 \quad \text{---------- (2.14)}$$

$$i_0 = \frac{v_0 - Av_{id}}{R_0}$$

$$v_{id} = v_1 - v_2 = -v_f = -\frac{R_1 v_0}{R_1 + R_F} = -Bv_o$$

therefore $i_0 = \frac{v_0 + A.Bv_0}{R_0}$

substituting the value of i_o in equation, we get

$$R_{oF} = \frac{v_0}{(v_0 + ABv_0)/R_0} = \frac{R_0}{1+AB} \quad \text{---------- (15)}$$

This result shows that the output resistance of the voltage series feedback amplifier is 1/(1+AB) times the output resistance R_o of the op-amp. That is, the output resistance of the op-amp with feedback is much smaller than the output resistance without feedback.

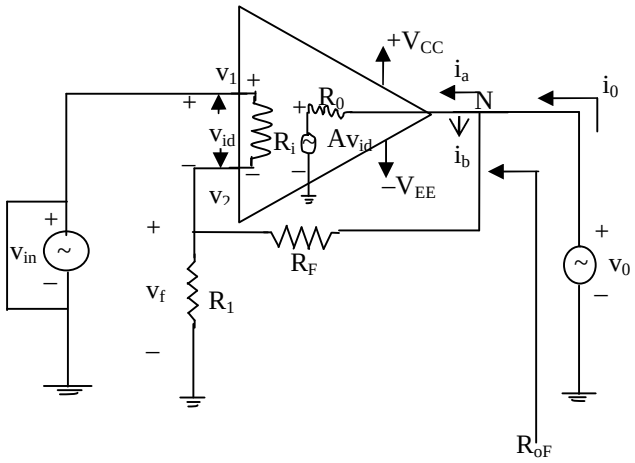


Fig 2.4 Derivation of output resistance with feedback

2.2.7 Bandwidth with feed back:

The bandwidth of an amplifier is defined as the band (range) of frequencies for which the gain remains constant. Manufacturers generally specify either the gain bandwidth product or supply open loop gain versus frequency curve for the op-amp. For the 741 op-amp the latter is typical.

The open loop gain versus frequency curve of the 741C op-amp is shown in Fig 2.5. From this curve for a gain of 200,000 the bandwidth is approximately 5Hz. In other extreme, the bandwidth is approximately 1MHz when the gain is unity. The frequency at which gain equals 1 is known as unity gain-band width. It is the maximum frequency the op-amps can be used for. Further more the gain bandwidth product obtained from the open loop gain versus frequency curve of fig 2.5 is equal to the unity gain bandwidth of an op-amp. However this holds true only for those op-amps like 741,which have just one break frequency below unity gain bandwidth.

Since the gain bandwidth product is constant, obviously the higher the gain the smaller the bandwidth and vice versa. As we have seen if negative feedback is used gain A decreases to A/1+AB. Therefore to obtain the closed loop bandwidth, the open loop bandwidth must be

multiplied by the same factor, by which the gain is divided, that is, by the factor (1+AB). In short.

Band width with feed back = (band width without feedback) (1+AB)

$$f_F = f_0 (1+AB) \text{ ---------- } (16)$$

or alternatively,

f_F = unity gain band width/closed loop gain = U.G.B.W /A_F ----------(17)

The closed loop bandwidth can also be determined from the open loop gain versus frequency plot. To do this we locate the closed loop voltage gain value on the gain axis and draw a parallel line through this value to the frequency axis. Then we project the point of inter section of the line with the curve on the frequency axis and read the value of the closed loop bandwidth. Using this procedure in figure 2.5. The bandwidth is approximately 100 KHz. for a closed loop gain of 10.

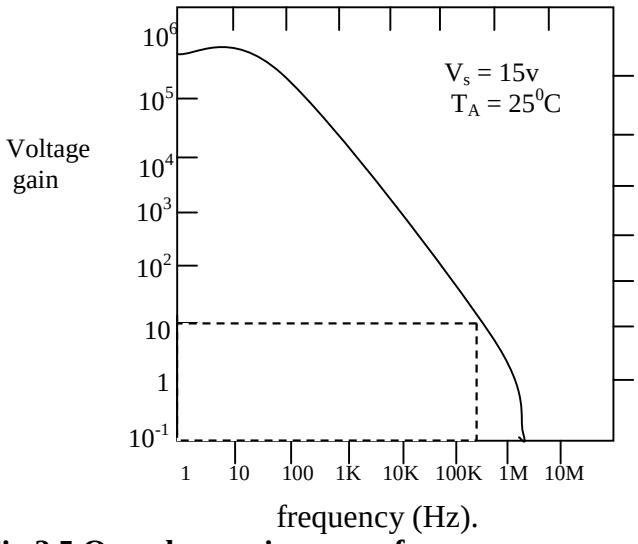


Fig 2.5 Open loop gain versus frequency curve of 741c.

2.2.8 Voltage follower:

The lowest gain that can be obtained from a non-inverting amplifier with feedback is 1. When the non-inverting amplifier is configured for unity gain, it is called a voltage follower

because the output voltage is equal to and in phase with the input. In other words, in the voltage follower, the output follows the input.

Although it is similar to the discrete emitter follower, the voltage follower is preferred because it has much higher input resistance, and the output amplitude is exactly equal to the input.

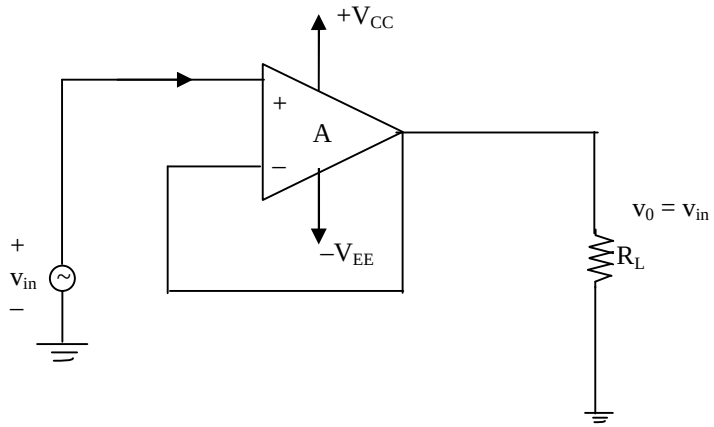


Figure 2.6. Voltage follower.

To obtain the voltage follower from the non-inverting amplifier of figure (2.2), simply open R_1 and short R_F. The resulting circuit is shown in figure 2.6. In this figure all the output voltage is feedback into the inverting terminal of the op-amp; consequently the gain of the feedback circuit is 1. ($B = A_F = 1$).

Since the voltage follower is a special case of the non-inverting amplifier all the formulae developed for the latter are indeed applicable to the former except that the gain of the feedback circuit is 1 (B=1) the applicable formulae are:-

$A_F = 1$

$R_{iF} = A\ R_i$

$R_{oF} = R_o/A$

$f_F = Af_o$

$$\text{vooT} = \frac{\pm v_{sat}}{A}$$

Since $(1+A) \cong A$

The voltage follower is also called a non-inverting buffer because when placed between two networks, it removes the loading on the first network.

2.3 Applications of op-amp:

Op-amps have many applications. Some are described below

2.3.1 Inverting amplifier: The basic inverting amplifier with an input resistor R_1 and feed back resistor R_F is shown in fig 2.7.

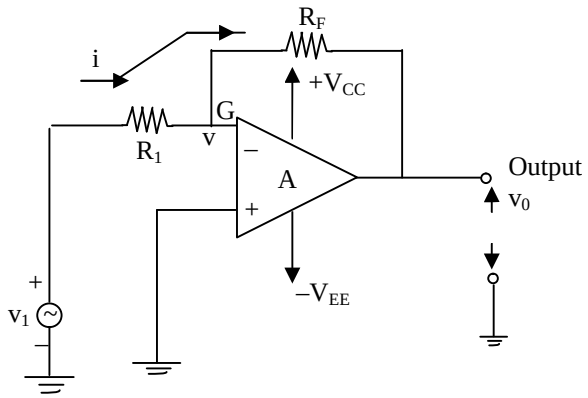


Fig 2.7 An inverting amplifier.

The non-inverting input terminal is grounded and the input voltage is v_1 and the output voltage is v_0. Since the gain 'A' of the operational amplifier is very large, the voltage v at the inverting input terminal is very small. In fact, the voltage v is very close to the ground potential. That is, although the point G is not actually connected to the ground, it is held virtually at ground potential irrespective of magnitudes or the potentials v_1 and v_0, the current i. is flowing through resistor R_1 is given by $i = \frac{v_1 - v}{R_1}$. Assuming that the operational amplifier is ideal, having an infinite input impedance the current i will flow through R_F and not into the op-amp. Applying the Kirchoff's current law at the point G. We can have

$$\frac{v_1 - v}{R_1} = \frac{v - v_0}{R_F} \quad \text{---------- (2.18)}$$

Since the point 'G' is virtually ground i.e.$v \approx 0$

From eqn.2.18. we get $\frac{v_1}{R_1} = \frac{-v_0}{R_F}$

Rearranging equation 2.18.

$$\text{We have } \frac{v_0}{v_1} = -\frac{R_F}{R_1} \quad \text{----------} \quad (2.19)$$

The ratio of the output voltage to the input voltage is the gain of the amplifier. Hence the voltage gain is the ratio of feedback resistor R_F to the input resistance R_1. The negative sign indicates that the output voltage is inverted w.r.t. the input voltage. The input resistance (R_{in}) of the whole amplifier is given by the ratio of voltage $v_1$and the input current $\frac{v_1 - v}{R_1}$ that is

$$R_{in} = \frac{v_1}{\frac{v_1 - v}{R_1}} \approx R_1 \quad \text{----------} \quad (2.20)$$

since $v \approx 0$, note that R_{in} refers to the whole of the amplifier not to the op-amp which has an infinite input resistance

2.3.2 Scale changer:

If use ratio $\frac{R_F}{R_1}$ is denoted by k, a real constant, we get $v_0 = -kv_1$. That is, input voltage has been multiplied by the factor '- k' to give the output voltage scale. The circuit shown in fig 2.7 acts as scale changer. For such applications precision resistors are used to get accurate values for the scale factor 'k'

2.3.3 Phase shifter:

If the resistance R_F and R_1 in the circuit are replaced respectively by impedances z_F an z_1 which are equal in magnitude but differ in phase angle, we can write

$$\frac{v_0}{v_1} = -\frac{z_F}{z_1} = -\frac{|z_F|}{|z_1|}.\frac{\exp(j\phi_F)}{\exp(j\phi_1)} = \frac{|z_F|}{|z_1|} e^{j[\Pi + \phi_F - \phi_1]}$$

$$\frac{v_0}{v_1} = \exp j[\Pi + \phi_F - \phi_1] \quad \text{----------} \quad (2.21)$$

Since $|z_F| = |z_1|$ and $e^{j\Pi} = -1$, the angles ϕ_F and ϕ_1 represent respectively, phase angles of the independents z_F and z_1. The circuit is capable of shifting the phase of a sinusoidal input signal

voltage with out changing its magnitude. The amount of phase shift can be any thing between 0 and 360^0.

2.3.4 Non- inverting amplifier:

The circuit diagram of non-inverting amplifier is shown in figure 2.8.

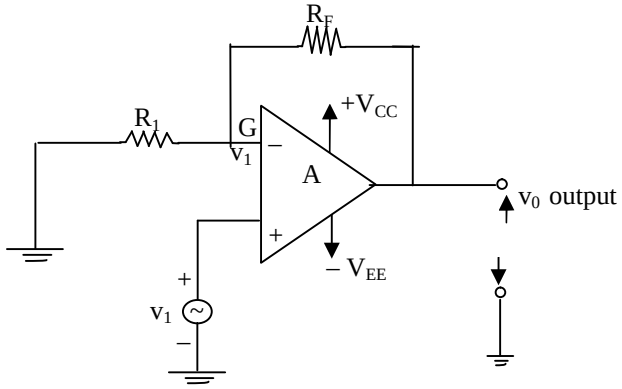


Fig 2.8 Circuit of non-inverting amplifier.

In this case the input voltage v_1 is applied to the non-inverting input terminal. The potential of the point G is also v_1, since the op-amp gain is infinite. Then, as mentioned before, since the input current to the op-amp is negligible, we can write using Kirchoff's current law at the point G.

$$\frac{v_0 - v_1}{R_F} = \frac{v_1}{R_1} \quad \text{----------} \quad (2.22)$$

Therefore, the gain of the amplifier is

$$\frac{v_0}{v_1} = 1 + \frac{R_F}{R_1} \quad \text{----------} \quad (2.23)$$

Since the gain of the op-amp is 1 plus the ratio of the two resistors R_F and R_1. And also note that the output voltage is in phase with the input voltage.

2.3.5 Summing amplifier:

An adder or summing amplifier using op-amp is shown fig2.7. Here the point G is a virtual ground. Since the current flowing into the ground is equal to that flowing out of it we can write

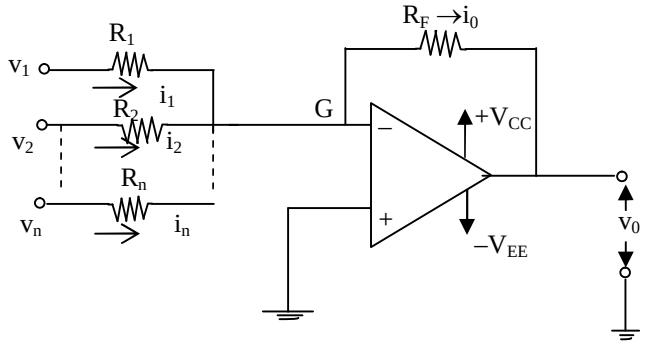


Fig 2.9 an adder or summing amplifier

$$i_1 + i_2 + \ldots\ldots + i_n = i_0 \quad \text{----------} \quad (2.24)$$

$$\frac{v_1}{R_1} + \frac{v_2}{R_2} + \ldots\ldots + \frac{v_n}{R_n} = -\frac{v_0}{R_F} \quad \text{----------} \quad (2.25)$$

Solving equation. (2.25) for v_0, we get

$$v_0 = -\ R_F \left[\frac{v_1}{R_1} + \frac{v_2}{R_2} + \ldots\ldots + \frac{v_n}{R_n}\right] \quad \text{----------} \quad (2.26)$$

If $R_1 = R_2 \ldots\ldots = R_n = R$, we obtain $v_0 = -\frac{R_F}{R}\ [\ v_1+v_2+\ldots+v_n\ ]$ ---------(2.27)

Again if $R_F = R$ we get

$v_0 = -[\ v_1+v_2+\ldots\ldots+v_n]$ --------(2.28)

Equation 2.28 shows that the output voltage v_0 is numerically equal to the negative sum of all input voltages v_1 through v_n. Hence it was named as adder or summing amplifier.

2.3.6 Integrator:

A circuit in which the output voltage waveform is the integral of the input waveform is the integrator or the integration amplifier. Such a circuit obtained by using a basic inverting amplifier configuration, if the feedback resistor R_F is replaced by a capacitor C_F and the relevant circuit is shown fig 2.10.

Fig 2.10 (b, c) input and ideal output waveforms using sine wave and square wave respectively

$R_1 C_F = 1$ second and $v_{oo} = 0$V assumed.

The expression for the output voltage v_0 can be obtained by writing Kirchoff ‘s current equation at node v_2:

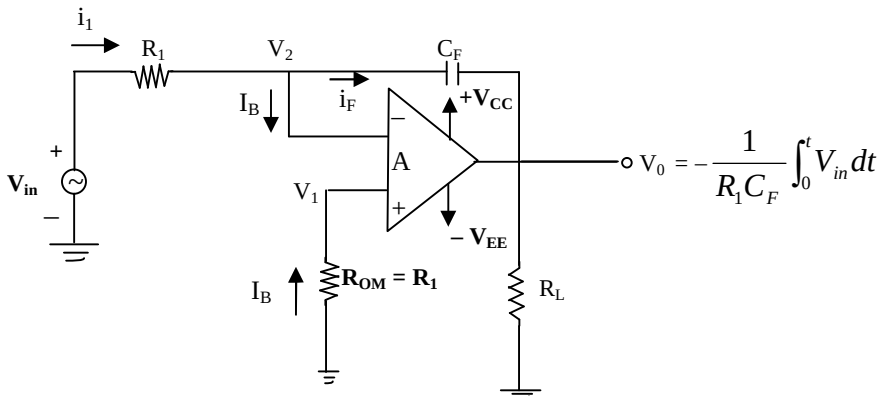


Fig 2.10(a) Integrator circuit.

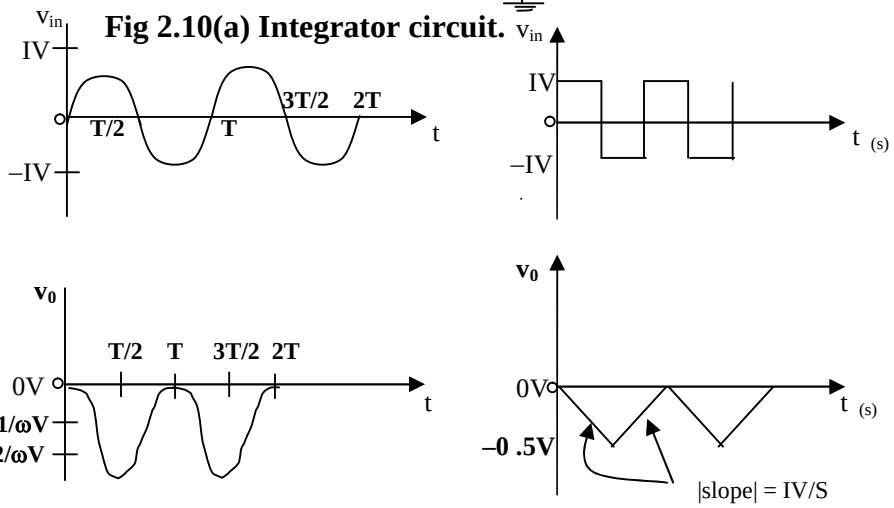


Fig 2.10(b) **Fig 2.10(c)**

$i_1 = I_B + i_F$

Since I_B is negligibly small, $i_1 = i_F$

Recall that the relationship between current through and voltage across the capacitor

$$i_c = C \cdot \frac{dv_c}{dt}$$

Therefore

$$\frac{v_{in} - v_2}{R_1} = C_F \left(\frac{d}{dt}\right)(v_2 - v_0)$$

However, $v_1 = v_2 \cong 0$ because A is very large. Therefore

$$\frac{v_{in}}{R_1} = C_F \frac{d}{dt}(-v_0)$$

The out put voltage can be obtained by integrating both sides w. r. t. time

$$\int_0^t \frac{v_{in}}{R_1} . dt = \int_0^t C_F \frac{d}{dt} (-v_0) . dt = C_F (-v_0) + v_0 \Big|_{t=0}$$

There fore

$$v_0 = - \frac{1}{R_1 C_F} \int_0^t v_{in}\, dt \qquad \text{--------(2.29)}$$

Equation 2.29 says that the output voltage is directly proportional to the negative integral of the input and inversely proportional to the time constant R_1C_F. For example if the input is a sine wave, the output will be a cosine wave of if input is a square wave, the output will be a triangular wave as shown as fig 2.10 (b) and (c) respectively. Note that the wave forms are drawn on the assumption that R_1C_F = 1second and v_{00} = 0v.

For accurate integration of the input waveform, the time period of the input signal T must be longer than or equal to R_1C_F. Again, the R_{om} is used to minimize the effect or input bias current and the output offset voltage. The input offset voltage v_{i0} and the part of the input current charging the capacitor produce the error voltage at the output of the integrator. Therefore in the practical integrator shown in fig 2.10(a), a resistor R_F is connected across the feed back capacitor C_F to produce DC stabilization. In other words R_F limits the low frequency gain and hence minimizes the variations is the output voltage. Generally, $R_F \cong 10R_1$. The integrator is most commonly used in analog computers and analog to digital converters (ADC) and signal wave shaping circuits.

2.3.7 The differentiator:

Fig 2.11 shows a differentiator or differentiating amplifier. As the name implies, the circuit performs the mathematical operation of differentiation: that is the output waveform is the derivative of the input waveform. The differentiator may be constructed form a basic inverting amplifier of an input resistor R_1 is replaced by a capacitor C_1

The expression for the output voltage can be obtained from Kirchoff's current equation written at node v_2 as follows:

$i_c = I_B + i_F$

since $I_B \cong 0$,

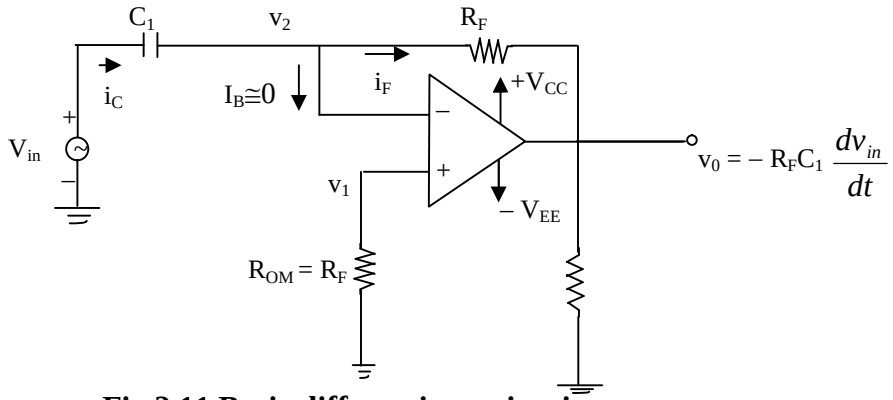


Fig 2.11 Basic differentiator circuit.

$I_c = I_F$

$$C_1 \frac{d(v_{in} - v_2)}{dt} = \frac{v_2 - v_0}{R_F}$$

But $v_1 = v_2 \cong 0v$, because A is very large, therefore

$$C_1 \frac{dv_{in}}{dt} = - \frac{v_0}{R_F}$$

or

$$v_0 = -R_F C_1 \frac{dv_{in}}{dt} \quad \text{----------} \quad (2.30)$$

Thus the output v_0 is equal to the $R_F C_1$ times the negative instantaneous rate of change of the input voltage v_{in} with time. Since the differentiator performs the reverse operation of the integrators function a cosine wave input will produce a sine wave output or triangular input will produce a square wave output. However, the differentiator of fig 2.11 will not do this because it has some practical problems. The gain of the circuit (R_F | Xc_1) increases with increase in frequency at a rate of 20 dB/decade. This makes the circuit unstable. Also, the input impedance Xc_1 decreases with increase in the frequency, which makes the circuit very susceptible to high frequency noise. When amplified, the noise can completely over ride the differentiated output signal.

Both the stability and the high frequency noise problems can be corrected by an addition of two components; R_1 and C_F as shown in fig 2.11 this circuit is a practical differentiator (fig 2.12)

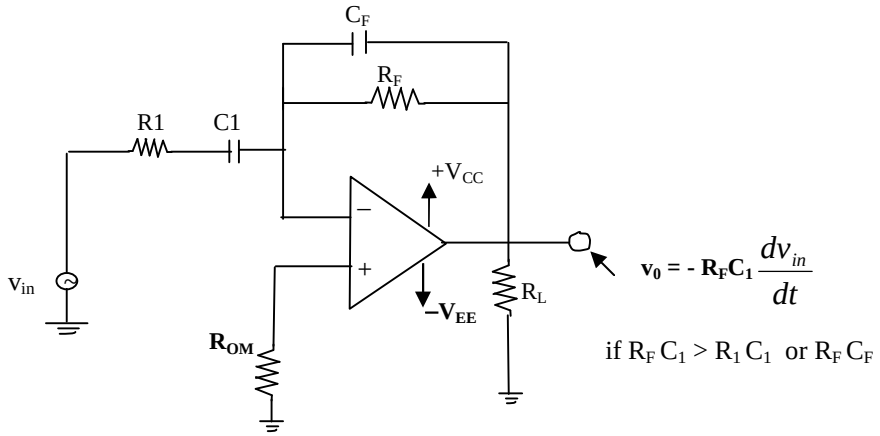


Fig 2.12(A) Practical differentiator.

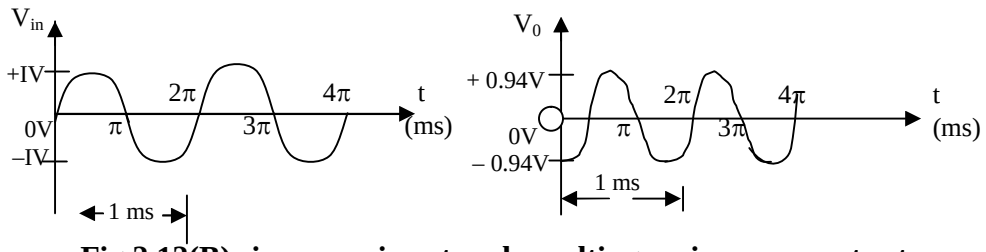


Fig 2.12(B) sine wave input and resulting cosine wave output.

2.4 Summary of the Lesson:

Open loop op amp configuration is unsuitable for linear applications because of its very high voltage gain as well as its variation with temperature, power supply etc.

Therefore, we can select, as well as, control the gain of the op-amp, if we introduce a modification in the basic circuit This modification involves the use of fed-back, that is, the output signal is fed back to the input either directly or via another network. If the signal feed-back is 180^0 out of phase w.r.t. the input signal. Such a type of feedback is called negative feed back. Negative feedback is also known as degenerative feedback, because when used, it reduces the output voltage and in turn reduces the voltage gain and stabilizes the gain.

On the other hand if the signal fed-back is in phase with the input signal such type of feedback is called positive feed back. Positive feedback is also known as regenerative feedback, because the feedback signals aid the input signal. Positive feedback is needed in the oscillator circuits.

There are four types of closed loop configurations using negative feedback: voltage series, voltage shunt, current series and current shunt. Further these configurations are labeled according to whether the voltage or current is fed-back to the input in series or in parallel. The voltage series negative feedback configuration is generally known as non inverting amplifier if the feedback and the ideal closed loop voltage gain of the configuration depends only on the feedback network (components) and is independent of the internal or open loop gain A of the op-amp.

Introduction of the negative feedback in non-inverting amplifier increases the input impedance and bandwidth and decreases the output resistance; total output- offset voltage and the effect of varying environmental conditions on the gain. The voltage follower is the special case of non-inverting amplifier.

Summing, scaling amplifiers can be constructed by using inverting, non-inverting and differential configurations. The integrator and differentiator are most widely used in signal wave shaping applications. Besides the integrator is used in analog computers and the differentiator is used as a rate of change detector in FM modulators.

2.5 Key terminology:

1. Active components: Active components can generate voltage or current or power
   Example: Battery, signal generator.
2. Passive components: The components only dissipate power
   Example: Resistor, capacitor.
3. Degenerative feedback: The output voltage is fed back to the input such that this voltage is 180^0 out of phase w.r.t. the source voltage. As a result, the effective input voltage decreases (degenerates) and this is called degenerative feedback or negative feedback.

2.6 Self-assessment questions:

2.1 Mention two reasons why an open loop op-amp is unsuitable for linear applications.

2.2 What do you mean by feedback? List two types of feedback. Which type is used in linear applications?

2.3 Describe the four negative feedback configurations. Which configuration is most widely used?

2.4 Explain briefly why negative feedback is desirable in amplifier applications.

2.5 How does negative feedback affect the close loop voltage gain, input resistance, output resistance and band width of voltage series feedback amplifier (closed loop non-inverting amplifier)

2.6 In what way is the voltage follower a special case of the non-inverting amplifier?

2.7 Explain the inverting and non-inverting amplifiers?

2.8 Explain the difference between the integrator and differentiator and give one application of each.

2.7 Reference Books:

(1) Operational Amplifiers and Linear Integrated Circuit Technology by Ramakanth A. Gaykwad Prentice Hall Inc.,

(2) Basic Electronics by DC Tayal, Himalaya Publish Co.,

(3) Semi Conductor Electronics by A K. Sharma
New Age International Publishers

(4) Foundations of Electronics
By D. Chattopadhyay, PC Rakshit, B. Saha, M.N. Purkait
Second Edition, Wiley Eastern Ltd.,

Name of the Author

Dr D.Krishna Rao, M.Sc., M.Phil., Ph.D.,
Associate Professor,
Physics Department,
Nagarjuna University,
Nagarjuna Nagar 522 510.

UNIT I

LESSON 3

OP-AMP CIRCUITS

Types of Oscillators-Generators and 555 Timer

Objectives:-

To explain

i) The oscillator principle and discuss the various types of oscillators

ii) The construction of square wave and triangular wave generators

iii) The operation of the 555 Timer as an astable multi-vibrator.

Structure of the lesson

Basically the function of an oscillator is to generate alternating current or voltage waveforms. More precisely, an oscillator is a circuit that generates a repetitive waveform of fixed amplitude and frequency with out any external input signal. Oscillators are used in radio, television, computers, and communications. Although there are different types of oscillators all of them work on the same basic principle.

Oscillators are classified into two basic categories: sinusoidal and non-sinusoidal. If the waveform generated looks like a sine or cosine wave, the circuit is called a sinusoidal oscillator and these circuits producing all other waveforms are called non-sinusoidal oscillators. Some times, the oscillators are also classified on the basis of frequency of the generated waveform, viz., audio frequency, radio frequency and ultra high frequency oscillators.

3.1 Oscillator principle:

An Oscillator, as such, is a type of feedback amplifier in which part of the output is feed back to the input via a feedback circuit. If the signal feedback is of proper magnitude and phase, the circuit produces alternating currents or voltages. To visualize the requirement of an oscillator, consider the block diagram of figure 3.1.

However, here the input voltage is zero (v_{in} = 0). Besides that, the feedback is positive because most oscillators use positive feedback. Finally, the closed loop gain of the amplifier is denoted by A_v rather that A_F.

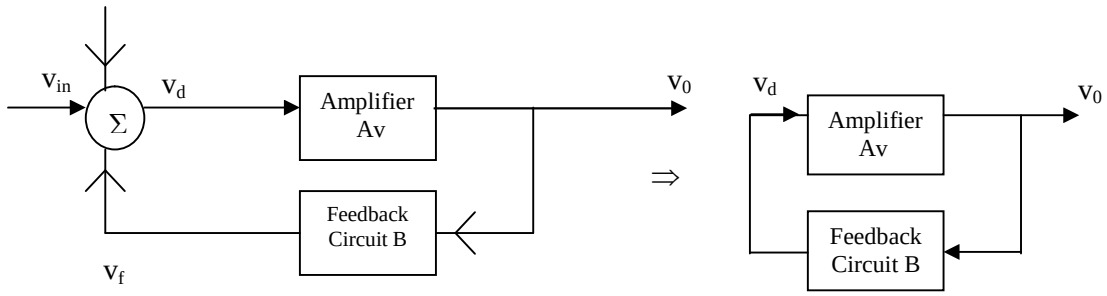


Fig 3.1 Oscillator block diagram

In figure 3.1

$$v_d = v_f + v_{in} \quad \text{--------} \ (3.1)$$

$$v_0 = A_v v_d \quad \text{-------} \ (3.2)$$

$$v_f = B v_0 \quad \text{-------} \ (3.3)$$

Using these relation ships, the following equations are obtained

$$\frac{v_0}{v_{in}} = \frac{A_v v_d}{v_d - v_f} = \frac{A_v}{1 - (\mathbf{v}f / vd)} \quad \text{----------} \ (3.4)$$

From equations (3.2) & (3.3), we have

$$\frac{v_f}{v_d} = A_v B \quad \text{----------} \ (3.5)$$

Substitute equation (3.5) is equation (3.4), we have

$$\frac{v_0}{v_{in}} = \frac{A_v}{1 - A_v B} \quad \text{----------} \ (3.6)$$

However, $v_{in} = 0$, and $v_0 \neq 0$ implies that

$A_V B = 1$

Expressed in polar form

$$A_v B = 1 \text{ and } \phi = 0 \text{ or } 360^0 \quad \text{----------} \ (3.7)$$

Equation (3.7) gives the two requirements for oscillators:

(1) The magnitude of the loop gain $A_v B$ must be at least 1, and

(2) The total phase of the loop gain $A_v B$ must be equal to 0^0 or 360^0. For instance, as indicated in fig (3.1), if the amplifier causes a phase shift of 180^0

The feedback circuit must provide an additional phase shift of 180^0, so that, the total phase shift around the loop is 360^0. The waveforms shown in fig 3.1 are sinusoidal and are used to illustrate the circuit's action. The type of waveform generated by an oscillator depends on the components in the circuit and hence, may be sinusoidal, square or triangular. In addition, the frequency of the oscillation is determined by the components in the feedback circuit.

3.1.1 Oscillator types:

Because of their wide spread use, many different types of oscillators are available. These oscillators are summarized as follows.

Types of Oscillator	Frequency of Oscillation	Types of waveform generated
1. RC Oscillator	1. Audio frequency (AF)	1. Sinusoidal
2. LC Oscillator	2. Radio frequency (RF)	2. Square wave
3. Crystal Oscillator		3. Triangular wave
		4. Saw-tooth wave

3.1.2 Frequency Stability:

The ability of the oscillator circuit to produce waves at one exact frequency is called frequency stability. Although there may be a number of factors that cause changes in oscillator frequency, the primary factors are temperature changes and changes in the dc power supply. Temperature and power supply changes cause variations in the op-amp's gain, in junction capacitances and resistances of the transistors in an op-amp, as well as in an external circuit components. In most cases these variations can be kept small by careful design, using regulated power supplies, and by temperature control.

Another important factor that determine frequency stability, is the figure of merit Q of the circuit. The higher the Q, the greater the stability. For this reason crystals oscillators are far more stable than RC and LC oscillators, especially at higher frequencies. LC circuit and crystal are generally used for the generation of high frequency signals, while RC components are most suitable for audio frequency applications.

3.2 Working of R-C Phase Shift Oscillator:

Fig 3.2 shows a phase shift oscillator, which consists of an op-amp, as amplifying stage and three RC cascaded networks in the feedback circuit. The feedback circuit provides feedback voltage from the output back to the input of the amplifier. The op-amp is used in the inverting mode; therefore any signal that appears at the inverting terminal is shifted by 180^0 at the output. An additional 180^0 phase shift required for oscillation is provided by the cascaded RC net works. Thus the total phase shift around the loop is 360^0 (or 0^0).

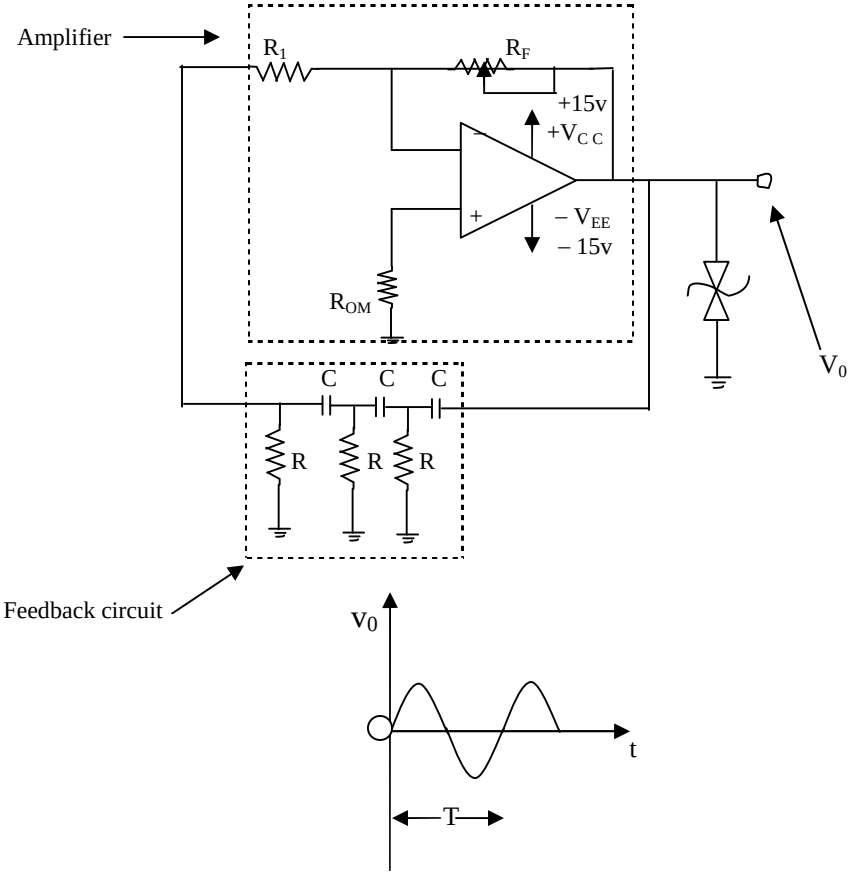


Fig (3.2) Phase shift oscillator and its output waveform

At some specific frequency when the phase shift of the cascaded RC networks is exactly 180^0 and the gain of the amplifier is sufficiently large, the circuit will oscillate at that frequency. This frequency is called the frequency of oscillation f_0 and is given by

$$f_0 = \frac{1}{2\pi\sqrt{6}RC} = \frac{0.065}{RC} \quad \text{---------- (3.8)}$$

At this frequency, the gain A_v must be at least 29. That is

$$\frac{R_F}{R_1} = 29$$

$$R_F = 29\, R_1 \quad \text{---------- (3.9)}$$

Then the circuit will produce a sinusoidal waveform of frequency f_0, if the gain is 29 and the total phase shift around the circuit is exactly 360^0. For a desired frequency of oscillator, choose a capacitor C_1 and then calculate the value of R from equation (3.8). A desired output amplitude, however, can be obtained with back-to-back Zeners connected at the output terminal.

Analysis:-

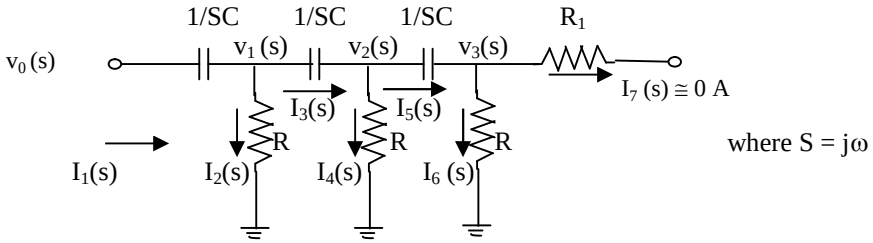


Fig (3.3) RC network of the phase shift oscillator

Nodal equations are transformed into the s domain

Writing Kirchoff's current law (KCL) at node $v_1(s)$, we get $I_1(s) = I_2(s) + I_3(s)$

$$\frac{v_0(s) - v_1(s)}{1/SC} = \frac{v_1(s)}{R} + \frac{v_1(s) - v_2(s)}{1/SC}$$

$$v_1(s)\left[\frac{1}{R} + 2SC\right] + v_2(s)(-SC) = v_0(s)(SC) \quad \text{---------- (3.10)}$$

Writing KCL at node $v_2(s)$

$I_3(s) = I_4(s) + I_5(s)$

$$\frac{v_1(s) - v_2(s)}{1/SC} = \frac{v_2(s)}{R} + \frac{v_2(s) - v_3(s)}{1/SC}$$

$$v_1(s)[-SC] + v_2(s)\left[\frac{1}{R} + 2SC\right] + v_3(s)[-SC] = 0 \quad \text{---------- (3.11)}$$

Writing KCL at node $v_3(s)$

$I_5(s) = I_6(s)$

$$\frac{v_2(s) - v_3(s)}{1/SC} = \frac{v_3(s)}{R}$$

$$v_2(s)\,[SC] + v_3(s)\left[-\frac{1}{R} - SC\right] = 0 \qquad \text{---------- (3.12)}$$

$$v_2(s) = v_3(s)\left[\frac{1}{RSC} + 1\right]$$

Substituting this in equation (3.11)

$$v_1(s)[-SC] + v_3(s)\left[1 + \frac{1}{SRC}\right]\left[\frac{1}{RSC} + 2\right] + v_3(s).(-SC) = 0$$

$$v_1(s) = v_3(s)\left[-1 + \left(1 + \frac{1}{RSC}\right)\left(2 + \frac{1}{RSC}\right)\right]$$

Substituting this in equation (3.10)

$$v_3(s)\left[-1 + \left(1 + \frac{1}{RSC}\right)\left(2 + \frac{1}{RSC}\right)\right]\left[\frac{1}{R} + 2SC\right] + v_3(s)\left[\frac{1}{RSC} + 1\right](-SC) = v_0(s)(SC)$$

$$v_3(s)\left\{-1 + \left(1 + \frac{1}{RSC}\right)\left(2 + \frac{1}{RSC}\right)\left(\frac{1}{R} + 2SC\right)\right\} + v_3(s)\left((-SC) - \frac{1}{R}\right) = v_0(s)(SC)$$

$$v_3(s)\left[-\frac{1}{R} - 2sC + \left[\frac{1}{RSC} + 1\right]\left[2 + \frac{1}{RSC}\right]\left[\frac{1}{R} + 2SC\right] - SC - \frac{1}{R}\right] = v_0(s)(SC)$$

$$v_3(s)\left[(-3sC - \frac{2}{R}) + \left(\frac{RSC+1}{RSC}\right)\left(\frac{2RSC+1}{RSC}\right)\left(\frac{1+2RSC}{R}\right)\right] = v_0(s)(SC)$$

$$v_3(s)\left[\frac{\left(-3R^3S^3C^3 - 2R^2S^2C^2\right) + (1+RSC)\left(1 + 4R^2S^2C^2 + 4RSC\right)}{R^3S^2C^2}\right] = v_0(s)(SC)$$

$$v_3(s)\,[R^3S^3C^3 + 6R^2S^2C^2 + 5RSC + 1] = v_0(s)\,[R^3S^3C^3]$$

$$\frac{v_3(s)}{v_0(s)} = \frac{S^3R^3C^3}{S^3R^3C^3 + 6S^2R^2C^2 + 5SRC + 1} \qquad \text{---------- (3.12 (a))}$$

Next consider the op-amp part of the phase shift oscillator, the voltage gain of the op-amp

$$A_v = \frac{v_0(s)}{v_3(s)} = -\frac{R_F}{R_1} \quad \text{----------} \quad (3.12\ (b))$$

For an oscillator

$$A_v B = +1. \quad \text{----------} \quad (3.13)$$

Therefore, using equations 3.12(a) and 3.12(b), we have

Since it is an inverting amplifier $A_v = -\frac{R_F}{R_1}$

$$\left(-\frac{R_F}{R_1}\right)\left[\frac{S^3R^3C^3}{S^3R^3C^3 + 6R^2S^2C^2 + 5SRC + 1}\right] = 1$$

Substituting S = jω, and equating the real and imaginary parts, respectively, we get

$$\left(-\frac{R_F}{R_1}\right)\left[\frac{-j\omega^3R^3C^3}{-j\omega^3R^3C^3 - 6\omega^2R^2C^2 + j5\omega RC + 1}\right] = 1$$

$$\left(-\frac{R_F}{R_1}\right)(-j\omega^3R^3C^3) = -j\omega^3R^3C^3 - 6\omega^2R^2C^2 + j5\omega RC + 1 \quad \text{----------} \quad 3.14$$

$$-6\omega^2R^2C^2 + 1 = 0 \text{ (Real)}$$

$$\omega^2 = \frac{1}{6R^2C^2} \quad \text{----------} \quad (3.14(a))$$

$$f = \frac{1}{2\pi\sqrt{6}RC} \quad \text{----------} \quad (3.15)$$

$$\left(-\frac{R_F}{R_1}\right)(-\omega^3R^3C^3) = -\omega^3R^3C^3 + 5\omega RC \text{ (imaginary)}$$

$$\frac{R_F}{R_1} = -1 + \frac{5}{\omega^2 R^2 C^2} \quad \text{----------} \quad (3.15\ (a))$$

From equation 3.14(a) and 3.15(a) we have

$$\frac{R_F}{R_1} = 29 \quad \text{----------} \quad (3.16)$$

3.3 Wien Bridge oscillator:

The circuit diagram for typical Wien bridge oscillator is shown in fig3.4. This involves an RC bridge circuit in which a frequency adjusting network is constructed of series combination of R_1 and C_1 connected from the output of the op-amp to the non-inverting input of the op-amp and of a parallel combination of R_2 and C_2 from the non-inverting input of the op-amp to the ground connections. The resistances R_3 and R_4 from the non-inverting input of the op-amp to the ground connections. The resistors R_3 and R_4 make the feedback path and determine the gain of the amplifier. The bridge circuit causes a phase shift of 360^0 between the output and input of the op-amp.

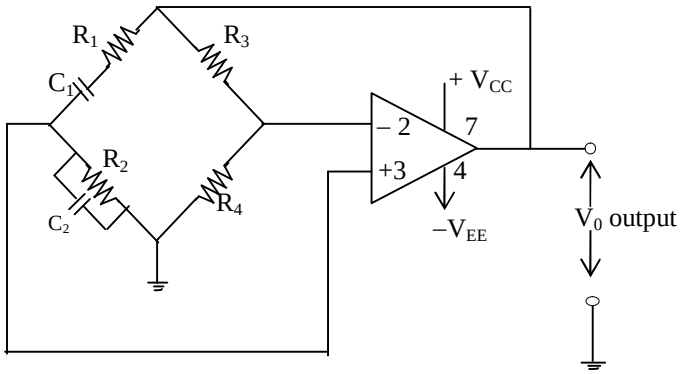


Fig 3.4(a) circuit configuration for Wien bridge oscillator

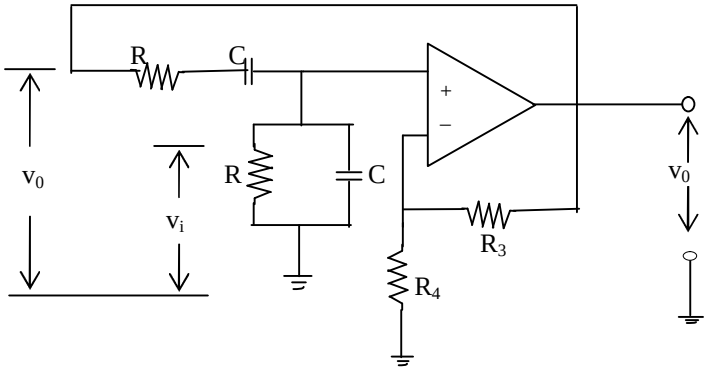


Fig 3.4(b) Redrawn circuit for $R_1 = R_2 = R$, and $C_1 = C_2 = C$

In the circuit of fig.4, the voltage v_i across the parallel combination of R and C is fed back and is expressed as

$$v_i = \frac{v_0 Z_1}{Z_1 + Z_2} \quad \text{----------} \quad (3.17)$$

Where $Z_1 = R\|C = \dfrac{R\dfrac{1}{j\omega C}}{R+\dfrac{1}{j\omega C}}$ and Z_2 (R in series with C) $= R + \dfrac{1}{j\omega C}$ ---------- (3.18)

$$v_i = \frac{v_0\left(\dfrac{R\dfrac{1}{j\omega C}}{R+\dfrac{1}{j\omega C}}\right)}{\left(\dfrac{R\dfrac{1}{j\omega C}}{R+\dfrac{1}{j\omega C}}\right)+\left(R+\dfrac{1}{j\omega C}\right)} \quad \text{----------} \quad (3.19)$$

$$v_i = v_0 \frac{R\dfrac{1}{j\omega C}}{R\dfrac{1}{jC} + R^2 + 2R.\dfrac{1}{j\omega C} - \dfrac{1}{\omega^2 C^2}}$$

$$= v_0 \frac{1}{1 + j\omega RC + 2 - \dfrac{j}{\omega RC}} = v_0 \frac{1}{3 + j\left(\omega RC - \dfrac{1}{\omega RC}\right)} \quad \text{----------} \quad (3.20)$$

In order to have zero phase between v_i and v_0, the j-terms must be zero i.e.,

$$\omega RC - \frac{1}{\omega RC} = 0;$$

$$\omega^2 = \frac{1}{R^2C^2}$$

$$\omega = \frac{1}{RC};\ f = \frac{1}{2\pi RC} \quad \text{----------} \quad (3.21)$$

Substituting this frequency in equation (3.20)

$$v_i = \frac{1}{3} v_0 \quad \text{----------} \quad (3.22)$$

Equation 3.22 shows that there is a loss of voltage gain in the feedback network by a factor of 3. There fore, the amplifier should have minimum voltage gain as 3 for oscillations.

The input voltage v_1 at the inverting point of the op-amp may be written as

$$v_i = \frac{v_0 R_4}{R_3 + R_4} \quad \text{----------} \quad (3.23)$$

Equaling equation (3.22) and (3.23), we get

$$\frac{1}{3} = \frac{R_4}{R_3 + R_4}$$

$$R_3 = 2R_4 \quad \text{----------} \quad (3.24)$$

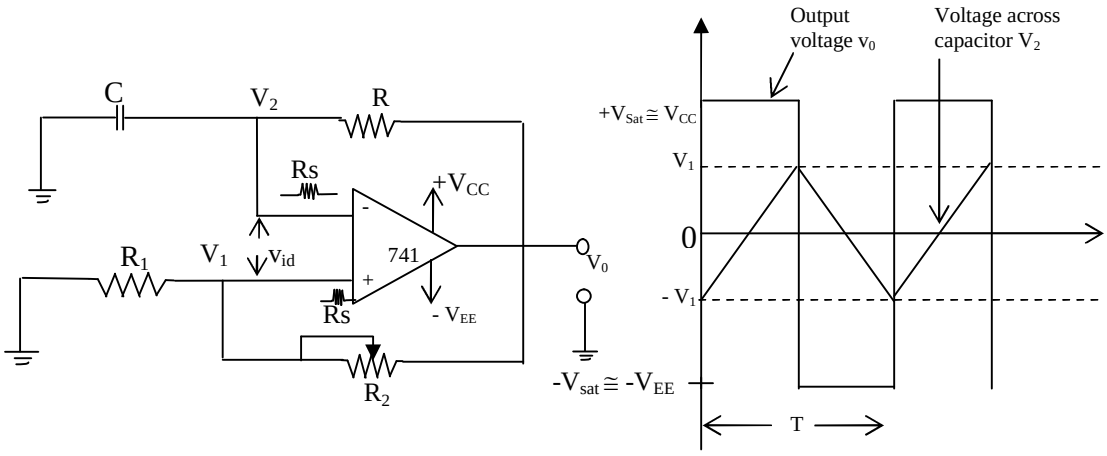
Hence, for a stable operation of the oscillator: R_3 is selected slightly larger than $2R_4$ so that the maximum gain requirement is fulfilled.

3.4 Square wave and Triangular wave generators

3.4.1 Square Wave Generator:

In contrast to sine wave oscillator, square wave outputs are generated when the op-amp is forced to operate in the saturated regions, that is, the output of the op-amp is forced to swing repetitively between positive saturation $+ V_{sat}$ ($\cong + V_{CC}$) and negative saturation $- V_{sat}$ ($\cong - V_{EE}$), resulting in the square wave output. Such a circuit showed in fig(3.5). This square wave generator is also called a free-running multi-vibrator.

Assume that the voltage across the capacitor C is zero volts at the instant the dc supply voltages $+V_{CC}$ and $-V_{EE}$ are applied. This means that the voltage at the inverting terminal is zero initially. At the same instant, however, the voltage v_1 at the non-inverting terminal is very small finite value, that is a function of the output offset voltage V_{ooT} and the values of R_1 and R_2 resistors. Thus the differential input voltage v_{id} is equal to the voltage v_1 at the non-inverting terminal. Although very small, voltage $v_1$will start to drive the op-amp into saturation.



$$|v_1| = \frac{R_1}{R_1 + R_2}\,|V_{sat}|;\ f_0 = \frac{1}{T} = \frac{1}{2RC} \text{ if } R_2 = 1.16R_1$$

Fig 3.5(a) Square wave generator (b) Wave form of output voltage. v_0 and capacitor voltage v_2 of the square wave generator.

For example, suppose that the output offset voltage v_{001} is positive and that, therefore voltage v_1 is also positive. Since initially the capacitor C acts as a short circuit, the gain of the amplifier is very large (A); hence v_1 drives the output of the op-amp to its positive saturation + V_{sat} with the output voltage of the op-amp as $+V_{sat}$, the capacitor C_2 starts charging toward $+V_{sat}$ through the resistor, R. However, as soon as the voltage v_2 across the capacitor is slightly more positive than v_1, the output of op-amp is forced to switch to a negative saturation, $-V_{sat}$. With the op-amps output voltage as negative saturation, $-V_{sat}$ the voltage v_1 across R_1 is also negative, since $v_1 = \frac{R_1}{R_1 + R_2}(-V_{sat})$ ---------- (3.25)

Thus the net differential voltage $v_{id} = v_1 - v_2$ is negative, which holds the output of the op-amp in negative saturation. The output remains in negative saturation until the capacitor C discharges and then recharges to a negative voltage slightly higher than the $-v_1$ [see Fig 3.5(b)]. Now, as soon as the capacitor voltage v_2 becomes more negative than $-v_1$, the net differential voltage v_{id} becomes positive and hence drives the output of the op-amp back to its positive saturation $+V_{sat}$. This completes one cycle. With output at $+V_{sar}$, and voltage v_1, at the non-inverting input is

$$v_1 = \frac{R_1}{R_1 + R} (+v_{sat}) \qquad \text{----------(3.26)}$$

The time period T of the output waveform is given by

$$T = 2RC \ln \left(\frac{2R_1 + R_2}{R_2} \right) \quad \text{---------} \quad (3.27(a))$$

$$\text{or} \quad f_0 = \frac{1}{2RC \ln \left(\frac{2R_1 + R_2}{R_2} \right)} \quad \text{----------} \quad (3.27(b))$$

Equation 3.27(b) indicates that the frequency of the output f_0 is not only a function of the RC time constant but also of the relationship between R_1 and R_2. For example If $R_2 = 1.16\ R_1$, equation 3.27(b) becomes

$$f_0 = \frac{1}{2RC} \quad \text{----------} \quad (3.28)$$

Equation (3.28) shows that the smaller the RC time constant, the higher the output frequency f_0 and vice versa. As in the sine wave oscillator, the highest frequency square wave generated is also set by the slew rate of the op-amp. In practice, each inverting and non-inverting terminal needs a series resistor R_s to prevent excessive differential current flow because the input of the op-amp is subjected to large differential voltage. R_s should have a value of 100 KΩ or higher.

3.4.2 Triangular wave generator: Recall that the output waveform of the integrator is triangular if its input is a square wave. It means that a triangular wave generator can be formed by simply connecting an integrator to the square wave generator of figure 3.5. The resultant circuit is shown in figure 3.6(a). The circuit requires a dual op-amp, two capacitors and at least five resistors. The frequency of the square wave and triangular wave is the same. For fixed R_1, R_2 and C values, the frequency of the square wave, as well as triangular wave, depends on the resistance R. As R is increased or decreased, the frequency of the triangular wave will decrease or increase, respectively. Although the amplitude of the square wave is constant; ($\pm V_{sat}$), the amplitude of the triangular wave decreases with an increase in its frequency, and vice versa.

The input of integrator A_2 is a square wave, while its out put is a triangular wave, however, the output of A_2 to be triangular wave requires that $5R_3C_2 > T / 2$, where T is the time period of the square wave input. As a general rule, R_3C_2 should be equal to T.

To obtain a stable triangular wave, it may also be necessary to shunt the capacitor C_2 with the resistor $R_4 = 10\ R_3$ and connect an offset voltage compensating network at the non-inverting terminal of A_2. As with any other oscillator, the frequency of the triangular wave generator is limited by slew rate of the op-amp. Therefore, high slew rate op-amps should be used for the generation of relatively higher frequencies.

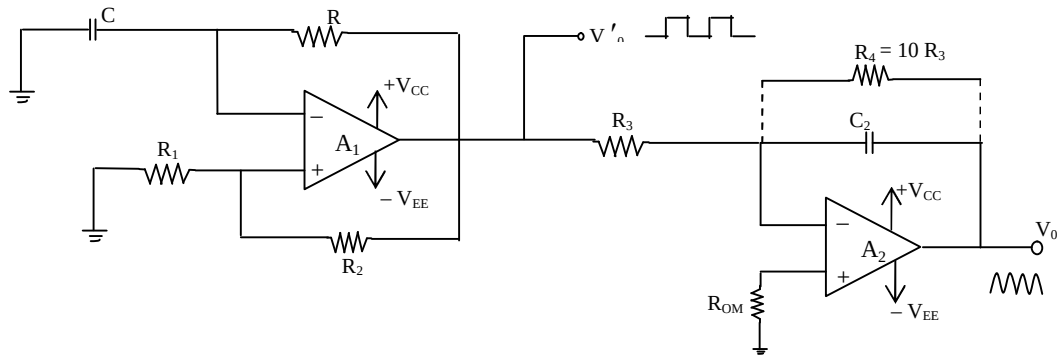


Fig 3.6(a) Triangular Wave Generator circuit

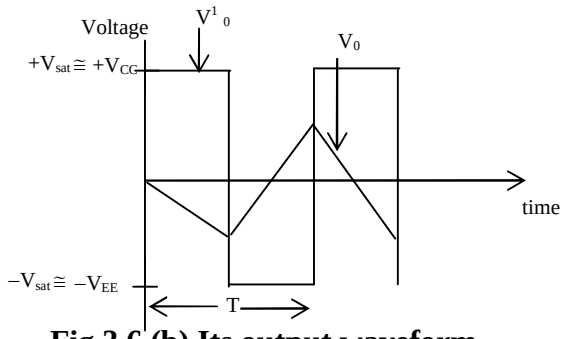


Fig 3.6 (b) Its output waveform.

3.5 The 555 Timer :

3.5.1 The 555 Timer and its pin configurations: Signetic corporation introduced this device as the SE / NE555 in early 1970 for the first time . It is one of the most versatile linear integrated circuit. The applications of 555 timer includes mono-stable and astable multi-vibrators, dc-dc converters, digital logic probes, wave form generators, analog frequency meters and tachometers, temperature measurement and control, infrared transmitters, burglar or toxic gas alarm, voltage regulators, electronic eyes and many others. The 555 is a monolithic timing circuit that can produce accurate and highly stable time delays or oscillation. In other words the timer basically operates in one of the two modes: either as a mono stable multi-vibrator (one shot) or as an astable (free running) multi-vibrator. The device is available as an 8 pin metal can, as an 8-pin mini DIP or a 14-pin DIP. Fig (3.7) shows that the pin connection diagram and the block diagram of the SE/NE 555 timer. The SE 555 is designed for the operating temperature range of –55 to $+125^0$c, while NE 555 operates over a temperature range of 0 to $+70^0$c. The important features of the 555 Timer are

(1) It operates on +5 to +15v supply voltage in both free running (astable) and one shot (mono stable) modes; **(2)** It has an adjustable duty cycle; timing is from micro seconds through hours; **(3)** It has high current output; it can source or sink 200 mA; **(4)** The output can drive TTL and has a temperature stability of 50 parts per million (ppm) per degree Celsius change in temperature.

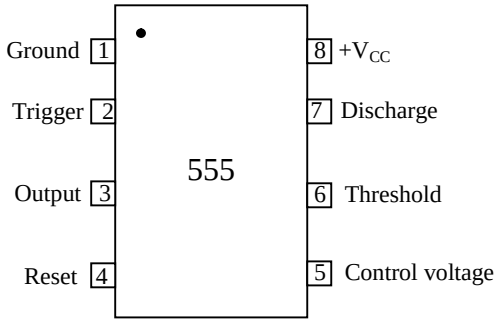


Fig 3.7(a) 555 timer connecting diagram

Before proceeding with the operations of 555 Timer an astable multi-vibrator, it is important to examine its pin functions. The pin number is used in the following discussion refer to the 8 pin mini DIP and 8 pin metal can packages

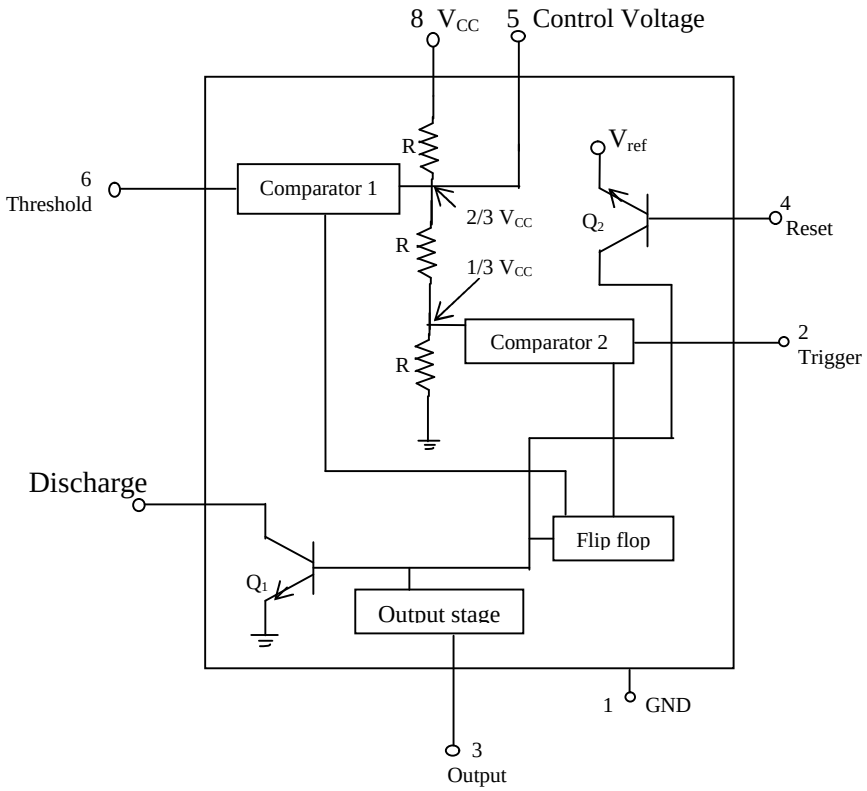


Fig 3.7(b) block diagram

(1) **Pin 1:** Ground. All voltages are measured w.r.t. this terminal.

Pin 2: trigger. The output of the timer depends on the amplitude of the external trigger pulse applied to this pin. The output is low if the voltage at this pin is greater that 2/3 V_{cc}.

However, when a negative going pulse of amplitude larger that 1/3 V_{cc} applied to this pin, the comparator 2 output goes low, which is turn switches the output of the timer high. The output remains high as long as the trigger terminal is held at a low voltage.

Pin 3: there are two ways a load can be connected to the output terminal: either between pin 3 and ground (pin 1) or between pin 3 and supply voltage $+V_{cc}$ (pin 8). When the output is low, the load current flows through the load connected between 3 and $+V_{cc}$ into the output terminal and is called the sink current. However, the current through the grounded load is zero when the output is low. For this reason, the load connected between 3 and $+V_{cc}$ is called normally on load , and that connected between pin 3 and ground normally off load and that connected between pin 3 and ground is called the normally off load.

Pin 4: Reset. The 555 timer can be reset (disabled) by applying a negative pulse to this pin. When the reset function is not in use, the reset terminal should be connected to $+V_{cc}$ to avoid any possibility of false triggering.

Pin 5: control voltage. An external voltage applied to this terminal changes the threshold as well as the trigger voltage. In other words, by imposing a voltage on this pin or by connecting a pot between this pin and ground, the pulse width of the waveform can be varied. When not used, the control pin should be bypassed to ground with a 0.01μF capacitor to prevent noise problems.

Pin 6: Threshold. This is the non-inverting input terminal of comparator 1, which monitors the voltage across the external capacitor. When the voltage at this point is ≥ threshold voltage $\frac{2}{3}V_{cc}$, the output of comparator 1 goes high, which in turn switches the output of the timer low.

Pin 7: Discharge. This pin is connected internally to the collector of transistor Q_1. When the output is high, Q_1 is off and acts as an open circuit to the external capacitor C

connected across it. On the other hand, when the output is low, Q_1 is saturated and acts as a short circuit, shorting out the external capacitor C to ground

Pin 8: $+V_{cc}$. The supply voltage of +5V to +18V is applied to this pin with respect to ground (pin 1)

3.5.2 The 555 as an astable multi-vibrator:

An astable multi-vibrator, often called a free running multi-vibrator, is a rectangular wave generating circuit. The time during which output is either high or low is determined by the two resistors and a capacitor, which are externally connected to the timer 555.

Astable operation:

Fig 3.8(a) shows the 555 timer connected as an astable multi-vibrator. Initially, when the output is high, capacitor C starts charging toward V_{cc} through R_A and R_B. However as soon as voltage across the capacitor equals $\frac{2}{3}V_{cc}$, comparator 1 triggers the flip-flop, and the output switches low [see fig 3.8(b)]. Now capacitor C starts discharging through R_B and transistor. When the voltage across C equals $\frac{1}{3}V_{cc}$, comparator 2's output triggers the flip-flop and the output goes high. Then the cycle repeats. The output voltage and capacitor voltage waveforms are shown in figure 3.8(b).

As shown in this figure, the capacitor C is periodically charged and discharged between $\frac{2}{3}$ V_{cc} and $\frac{1}{3}$ V_{cc} respectively. The time during which the capacitor charges from $\frac{1}{3}$ V_{cc} to $\frac{2}{3}$ V_{cc} is equal to the time the output is high and is given by

$$t_c = 0.69 (R_A + R_B)C \quad \text{----------} \quad (3.29)$$

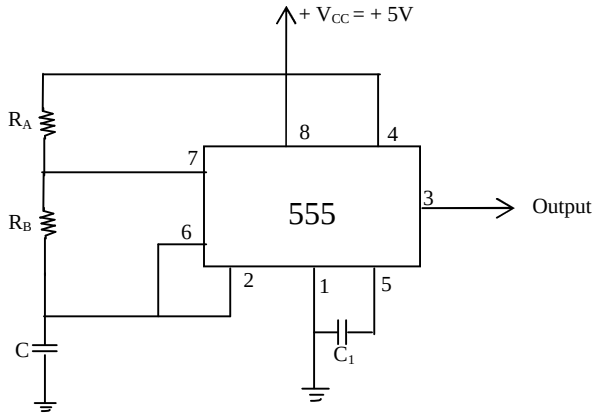


Fig 3.8(a) The 555 astable multi-vibrator circuit

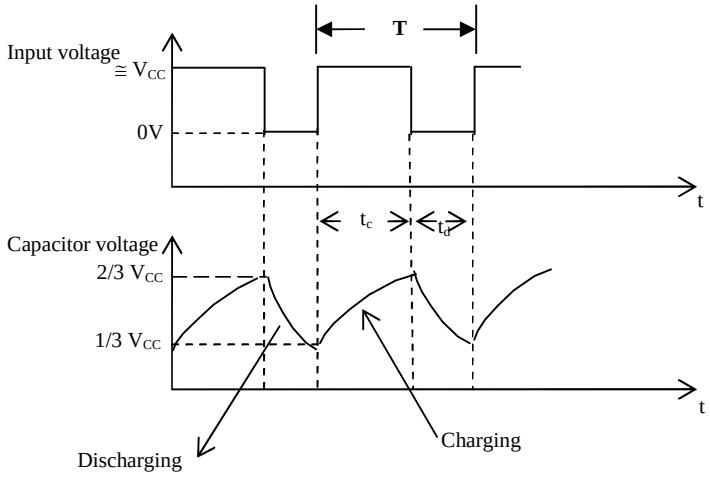


Fig 3.8(b) The 555 astable multi-vibrator output voltage wave form and voltage across the capacitor.

Where R_A and R_B are is ohms and C is in Farads. Similarly the time during which the capacitor discharges from $\frac{2}{3}$ V_{cc} to $\frac{1}{3}$ V_{cc} is equal to the time, the output is low and is given by

$$t_d = 0.69\ R_B C \quad \text{---------- (3.30)}$$

Thus the total time period of the output waveforms is

$$T = t_c + t_d = 0.69\ (R_A + 2R_B)\ C \quad \text{----------(3.31)}$$

This, in turn gives the frequency of oscillation as

$$f_0 = \frac{1}{T} = \frac{1.45}{(R_A + 2R_B)C} \quad \text{----------} \quad (3.32)$$

Equation (3.32) indicates then the frequency f_0 is independent of the supply voltage V_{cc}.

3.6 Summary of the Lesson:

Basically, the function of an oscillator is to generate alternating current or voltage waveforms. The oscillators are classified with two basic categories: sinusoidal and non-sinusoidal. If the waveform generated looks like a sine or cosine wave, the circuit is called a sinusoidal oscillator and the circuits producing all other waveforms are called non-sinusoidal oscillators.

There are two requirements for oscillations:

1. The magnitude of loop gain $A_V B$ must be greater than or equal to 1.
2. The total phase shift of the loop gain must be 0^0 or 360^0.

The Phase shift and Wien bridge are the most commonly used sinusoidal oscillators for frequencies below 100kHz. In all these oscillators the frequency of oscillation is a function of the RC time constant.

For frequencies above 100kHz, tunable oscillators like Hartley, Colpit's oscillators and crystal oscillators are preferable.. The theory regarding these oscillators is available in standard text books and is left as an exercise to the students. For information on this topic students may refer to books mentioned under reference.

Square wave outputs are generated when the op-amp is forced to operate in the saturated regions, that is, the output of the op-amp is forced to swing repetitively between positive saturation $+V_{sat}$ ($\cong V_{CC}$) and $-V_{sat}$ ($\cong -V_{EE}$) resulting in the square wave output. One way to obtain the triangular wave is to integrate the square wave. Therefore, the triangular wave generator can be formed by using a comparator and integrator.

Signetics NE / SE 555 is a monolithic timing circuit, that can produce accurate and highly stable time delays or oscillations. Thus the 555 may be used as either a mono-stable or an astable multi-vibrator. This device can be used in such applications as waveform generators, digital logic probes, integrated transmitters, burglar alarms, toxic gas alarms and electronic eyes.

3.7 Key terminology:

1. Crystal oscillator: A crystal oscillator is also a resonance frequency oscillator in which a piezo electric crystal is used as tuned / resonant circuit because properly cut crystal shows the characteristic of a resonant circuit. The crystal oscillator possesses excellent frequency stability.

2. Radio frequencies: The frequency of electromagnetic radiation with in the range 1000kHz – 1000MHz.

3. Saw-tooth wave: A wave form in which the shape resembles the teeth of a saw. The voltage builds slowly and linearly up to a peak value (maximum) and when falls almost instantaneously to zero or valley (minimum) value in each cycle.

4. Potentiometer: A variable wire wound resistor with three leads, the value of resistor varies between extreme lead and wiper.

5. Wien bridge: This involves an RC bridge circuit in which a frequency adjusting network is constructed of a series combination of R and C connected from the output of the op-amp to the non-inverting input of the op-amp and of a parallel combination of R and C from the non inverting input of the op-amp to the ground connection.

6. Dual in line package: In the duel – in – line package (DIP) the chip is mounted inside a plastic or ceramic case.

7. Metal can type: The chip is encapsulated in a metal or plastic case. The metal can is best suited for power amplifiers because metal is a good heat conductor and consequently has better dissipation capability than the flat-pack or dual – in – line package.

8. Flip-Flop: The most important memory element is the Flip – Flop which is made up of an assembly of logic gates. It can store one bit of binary data (logic 1 or 0)

3.8 Self-assessment questions:

3.1) Define an oscillator.

3.2) Why is the positive feedback required for oscillators?

3.3) What are the two conditions for oscillators?

3.4) How are oscillators classified?

3.5 Explain the frequency stability and its significance.

3.6 Describe the phase shift oscillator and obtain an expression for frequency of oscillations.

3.7 Derive an expression for the frequency of oscillation for Wien bridge oscillator. How is the minimum gain requirement is full-filled in this oscillator?

3.8 What are the important features or of the 555 Timer ?.

3.9 What are the two basic modes in which the 555 Timer operates?

3.10 Explain the operation of 555 timer as an astable multi-vibrator.

3.9 Reference Books:

(1) Operational Amplifiers and Linear Integrated Circuit Technology by Ramakanth A. Gaykwad Prentice Hall Inc.,

(2) Basic Electronics by DC Tayal, Himalaya Publish Co.,

(3) Semi Conductor Electronics by A K. Sharma
New Age International Publishers

(4) Foundations of Electronics
By D. Chattopadhyay, PC Rakshit, B. Saha, M.N. Purkait
Second Edition, Wiley Eastern Ltd,.

(5) Integrated Electronics by Millman & Halkias (MH)

Name of the Author

Dr D.Krishna Rao, M.Sc; M.Phil; Ph.D.,
Associate Professor,
Physics Department,
Acharya Nagarjuna University,
Nagarjuna Nagar 522 510.

UNIT – I

LESSON - IV

POWER CIRCUITS

Objective: The objective of this chapter is

1) To explain the need of power amplifiers and their working.
2) To explain the working of IC voltage regulated power supplies.

Structure of Lesson:

4.1 Power amplifers

In small signal amplifiers, the signal voltage and current are smaller, therefore, the amount of power handling capacity and power efficiency have little importance. The only requirements of small signal amplifiers are linearity and gain. When large signal are to be amplified for the operation of devices such as speakers and motors, the amplifier must be capable of handling large amount of power and its efficiency of converting input dc power to output ac power must be high. Such an amplifier is known as power amplifier. Hence the important requirements of power amplifiers are the power efficiency of the circuit, the maximum amount of power which the circuit can handle, and impedance matching between amplifier output and load, so that the maximum power is transferred to the output device.

4.1.1 Classification of Power amplifiers:

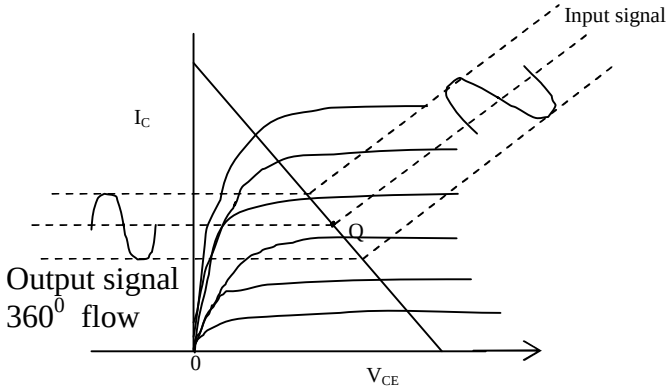


Fig 4.1(a) Operating conditions for power amplifiers class - A

The power amplifier are classified on the basis of their operating power as class A, B, C, D, E, F, and G. In other words, this classification depends upon the duration of the collector current flow w.r.t. the input signal, which is assumed to be a full 360^0. In class A amplifiers, there is always collector current regardless of the time in the cycle of the applied signal. The operating point Q is selected at the middle of the linear region of the transistor characteristic as shown in fig 4.1(a). This class of amplifiers is considered to be linear because output signal is an exact replica of the input signal.

In class – B amplifiers, the collector current flows only for one half of input signal. The operating point Q is located near the cut – off region as shown in fig 4.1(b). In order to get the output current for full cycle, two Class–B amplifiers are used in a combination known as push – pull.

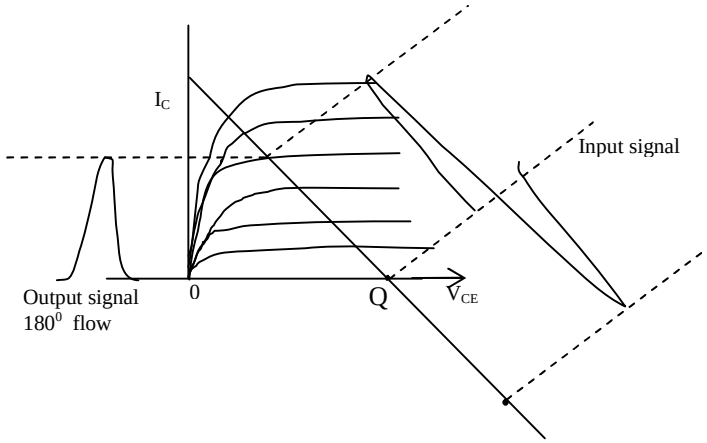


Fig 4.1(b) Operating conditions for power amplifiers class - B

In class – C amplifiers, the collector current flows for less than the half cycle of the input signal. The operating point Q is located beyond the cut-off point as shown in fig4.1(c). There are some amplifiers, which have operating parameters in between class-A and Class-B amplifiers. These amplifiers are designated as class-AB.

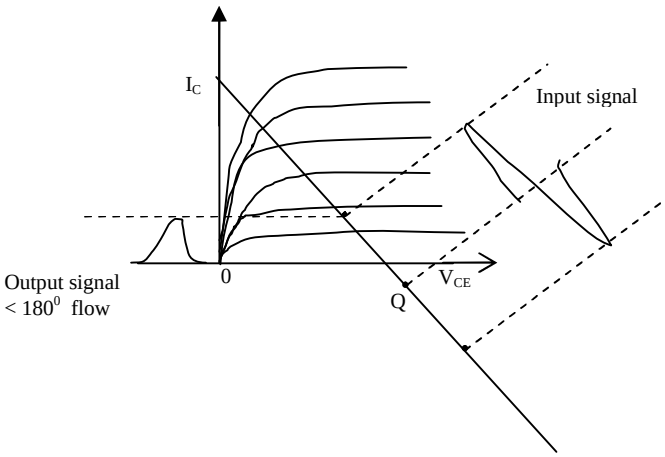


Fig 4.1(c) Operating conditions for power amplifiers class - C

4.1.2 Efficiency of a power amplifier:

Fig 4.2 shows a typical power amplifier circuit. The power P_s supplied by the battery V_{cc} is distributed in both load resistor R_L and the transistor as P_R and P_D respectively, i.e.,

$$P_S = P_R + P_D \quad \text{----------(4.1)}$$

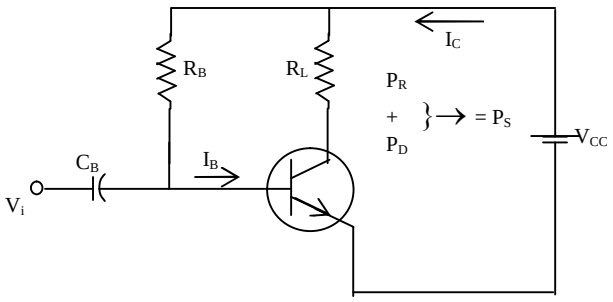


Fig4.2 Power distribution in Class – A power amplifier circuit

In an amplifier circuit with the input signal, the current drawn from the battery V_{CC} has a wave shape having an ac signal component riding on a dc bias component. That is, the dc voltage source in an amplifier circuit furnishes a dc plus an ac current, which is not the case with input circuits having only passive elements such as resistors, capacitors and inductors. Hence the power across R_L is the sum of the powers owing to ac and dc currents i.e.,

$$P_R = P_{RD} + P_{RA}$$

The power occurring due to ac current P_{RA} is the only useful component in an amplifier because it represents the ac output power P_0 i.e.,

$$P_{RA} = P_O$$

$$\text{Hence } P_S = P_O + P_{RD} + P_D \quad \text{---------- (4.2)}$$

The terms P_{RD} and P_O represent the losses in the circuit. In fact, the operation of the power amplifier is to convert as much of the dc power P_S drawn from the battery into the ac power P_O across $R_{L.}$ Hence, they efficiency of the circuit is defined as

$$\% \text{ Efficiency} = \eta = \frac{P_{O(ac)}}{P_{S(dc)}} \times 100$$

The power P_S and P_O can be expressed as

$$P_S = V_{CC} \,.\, I_C,\ P_O = I_C{}^2{}_{(rms)}.R_L \quad \text{---------- (4.3)}$$

Where I_C is the average current drawn from the battery and $I_{C(rms)}$ is the r.m.s. value of the ac current through R_L.

4.1.3 Class A power amplifier:

The circuit of a typical class A transistor power amplifier operating in the common emitter mode is shown in fig.4.3. This mode of operation gives the largest power gain of all the three possible configurations. The amplifier is directly coupled to the load resistance R_L. The transistor is biased in class A condition, so that the collector current flows during the whole of the input signal cycle. The capacitor C_E is the emitter bypass capacitor and prevents the ac voltage from appearing across R_E. C_B is the blocking capacitor. This prevents the ac input signal voltage from interacting with the dc voltage in the base circuit.

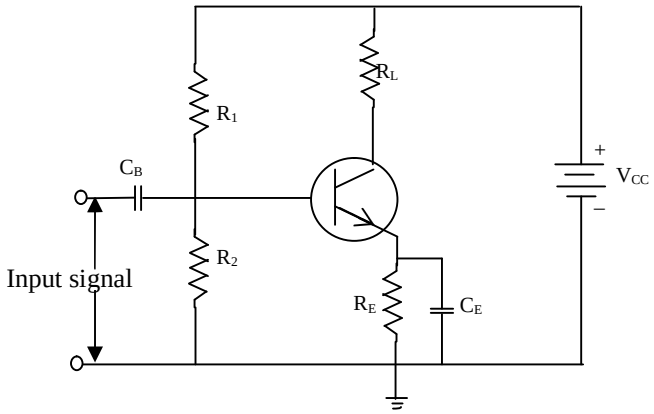


Fig4.3. The circuit of a typical class – A power amplifier directly coupled to the load resistance

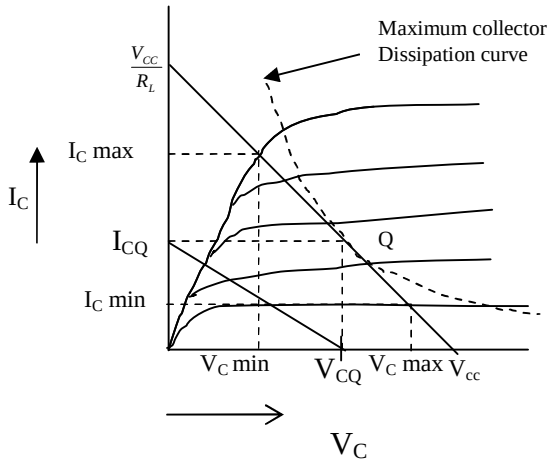


Fig 4.4 Collector characteristics with the load line and the quiescent operating point 'Q' for a direct coupled class – A power amplifier.

Fig 4.4 shows the collector characteristics of the transistor. The load line corresponding to R_L is drawn on the characteristics. Q is the quiescent operating point. This is located mid way upon the load line to ensure maximum output power. When the maximum input signal is applied, the quiescent operating point swings between cut off region and saturation. At cut off, the collector voltage is a maximum and the collector current is a minimum. At saturation, reverse is the case, i.e., the collector voltage is a minimum and the collector current is a maximum. In fig 4.4 V_{Cmax} and I_{Cmin}, represent the collector voltage and current corresponding to cut off condition; and V_{Cmin} and I_{Cmax} represent the same quantities corresponding to the saturation.

The input dc power supplied by the collector power supply is

$$P_{CS} = V_{CC} \quad I_{CQ} \quad \text{----------} \quad (4.1)$$

Where I_{CQ} is the quiescent collector current.

The ac output power is given by

$$P_{ac} = V_{Crms} \quad I_{Crms} \quad \text{----------} \quad (4.2)$$

Where V_{Crms} and I_{Crms} are the r.m.s. values of ac collector voltage and current respectively.

Assuming that the operating point swings equal distances on each side of the quiescent point Q, the peak value of the ac voltage V_{cm} and current I_{cm} are given by (see fig 4.4)

$$V_{Cm} = \frac{V_{Cmax} - V_{Cmin}}{2} \quad \text{----------} \quad (4.3)$$

$$I_{Cm} = \frac{I_{Cmax} - I_{Cmin}}{2} \quad \text{----------} \quad (4.4)$$

Therefore, the ac output power is

$$P_{ac} = \left(\frac{V_{Cmax} - V_{Cmin}}{2\sqrt{2}}\right)\left(\frac{I_{Cmax} - I_{Cmin}}{2\sqrt{2}}\right)$$

$$P_{ac} = \frac{(V_{Cmax} - V_{Cmin})(I_{Cmax} - I_{Cmin})}{8} \quad \text{----------} \quad (4.5)$$

Therefore, the collector efficiency is

$$\eta = \frac{P_{ac}}{P_{CS}} = \frac{(V_{Cmax} - V_{Cmin})(I_{Cmax} - I_{Cmin})}{8V_{CC}I_{CQ}} \quad \text{----------} \quad (4.6)$$

If the collector characteristics are ideal, then $V_{Cmin} = 0$, and $I_{Cmax} = 2I_{CQ,}$ $V_{Cmax} = V_{CC}$ and $I_{Cmin} = 0$, substituting these values in eq.(4.6), we get,

$$\eta = \frac{2V_{CC}I_{CQ}}{8V_{CC}I_{CQ}} = 0.25 \text{ ---------- (4.7)}$$

Therefore, the maximum efficiency that can be obtained from a class-A power amplifier when coupled directly to the load resistance, is 0.25 or 25% .

4.1.4 Transformer coupled class-A power amplifier:-

When the transistor works into a load resistance that is different from its output resistance, to obtain maximum power output, the load resistance is coupled to the power amplifier by a transformer. This is shown in fig 4.5.

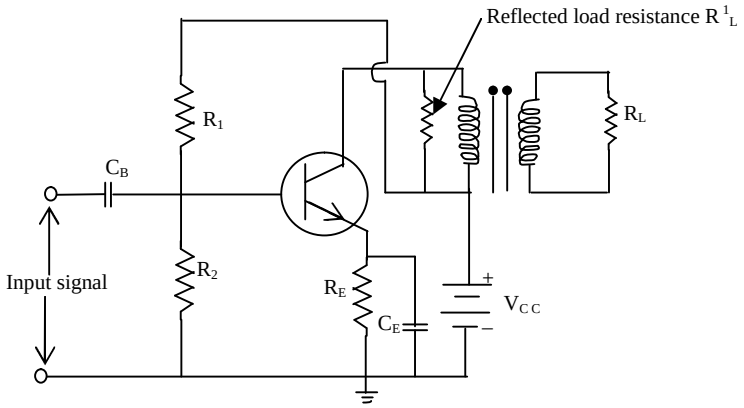


Fig.4.5. A Transformer coupled class – A, transistor power amplifier

In this circuit, the value of the base bias resistors R_1 and R_2 are adjusted to give class A operation with the expected input signal voltage. C_E is the emitter bypass capacitor and C_B is the blocking capacitor that prevents the ac input signal voltage from being interacted by dc base bias voltage.

The dc resistance of the transformer primary is small. This resistance serves as the collector load of the amplifier for dc currents. Hence the load line corresponding to this resistance will be almost vertical. This load line is the dc load line, and is shown in fig 4.6. When the ac input signal source is connected to the input of the amplifier, resistance in the collector circuit is formed by the reflected resistance of the load R_L. The load resistance reflected to the transformer primary is given by

$$R_L^1 = a^2 R_L \text{ ---------- (4.8)}$$

Where 'a' is the ratio of the number of turns N_p, in the primary to the number of turns N_s in the secondary and is called the turns ratio of the transformer.

If the resistance of the secondary windings of the transformer is R_s, then eq.(4.8) becomes

$$R_L^1 = a^2(R_S + R_L) \text{ ----------(4.9)}$$

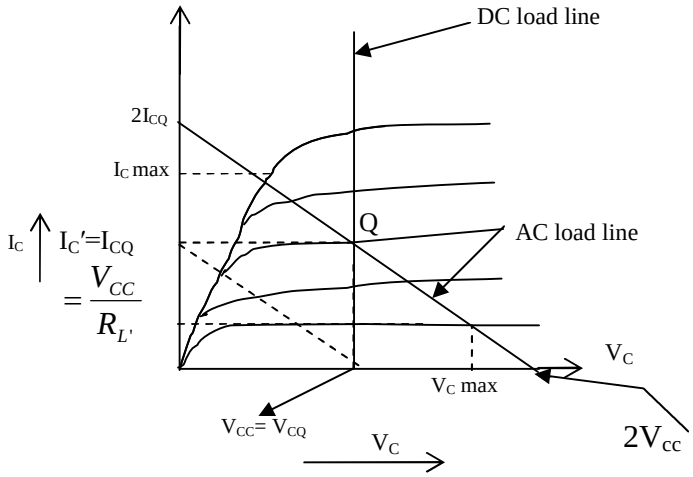


Fig 4.6 Collector characteristics with the dc load line, quiescent operating point and ac load line for the transformer coupled class A power amplifier

$R_L^{'}$ is the load resistance actually seen by the collector for ac currents. This resistance is made equal to V_{Cmax} / I_{Cmax} by properly choosing a. The load line corresponding to this resistance is called dc load line and is shown in fig (4.6). This line has a slope. $-1 / R_L^{'}$. To locate this ac load line, the following method is adopted.

a) First assume some convenient value of the power supply voltage, say V_{CC}

b) Determine the current $I_C^{'} = \frac{V_{CC}}{R_L^{'}}$ and locate the point on the I_c axis.

c) Draw the line joining $(V_{CC}, 0)$ and $(0, I'_C)$ on the collector characteristics (the dotted line in fig 4.6)

d) Now, draw a line through the quiescent operating point Q and parallel to the line V_{CC} I'_C. This line through the Q point is the desired ac load line.

When the ac signal is applied to the input of the amplifier, the collector current fluctuates and the operating point moves along the ac load line both sides of the Q-point. Thus the instantaneous value of the collector voltage V_C will exceed the collector supply voltage V_{CC} during a part of the cycle. This is because of the fact that the induced voltage across the transformer primary windings adds to the supply voltage during part of the cycle.

For maximum output, the Q point is located on the dc load line to give a quiescent collector current $I_{CQ} = \frac{I_{C\,max}}{2}$. Hence, when the transistor operates at maximum capacity and its collector

characteristics are ideal, the instantaneous collector voltage V_C swings from V_{Cmax} to zero and the collector current from zero to I_{Cmax}. This gives

$$V_{Cmax} = 2\ V_{CC} \quad \text{---------- (4.10)}$$

and $$I_{Cmax} = 2\ I_{CQ} \quad \text{---------- (4.11)}$$

Also, for ideal collector characteristics, $V_{Cmin} = 0$, and $I_{Cmin} = 0$,

The expression for the actual collector efficiency is the same as that for a direct- coupled class – A amplifier and is given by eqn (4.6).

$$\eta = \frac{2V_{CC}.2I_{CQ}}{8V_{CC}I_{CQ}} = 0.5 \quad \text{---------- (4.12)}$$

Thus the theoretical maximum efficiency of transformer- coupled class A power amplifier is 50%. This is twice as great as that for direct- coupled class A amplifier.

4.1.5 Push-pull connections:-

The amplifier circuits of figs 4.3and 4.5 employ a single transistor as the amplifying active device. Hence, they are called single ended power amplifiers. There is a definite limit to the power output obtainable from single ended amplifier. The power output can, however, be increased by using two transistors connected in parallel. Another method of obtaining greater output power is to connect two transistors in such a way that the collector current in one transistor decreases when that in the other transistor increases. This type of connection is commonly called the push-pull connection. The push-pull connection can produce a power output of the same order of magnitude that a parallel connection can produce, but other additional advantages over parallel operation, the most important one of which is the elimination of the even order harmonic distortion. Hence, a push-pull operation is widely used when a greater maximum power output with a prescribed amount of harmonic distortion is required. The push-pull connection can be used for class A, B, AB, and C operations.

4.1.6 Push-pull Class – B power amplifier:

The circuit diagram of a push-pull Class – B transistor power amplifier is shown in fig 4.7. The bases and the collectors of the two transistors are connected to the opposite ends of the center-tapped input and output transformers T_1 and T_2 respectively. The transistors are biased to cut-off so that no collector currents, except the leakage or cut-off currents, flow when the input signal is zero.

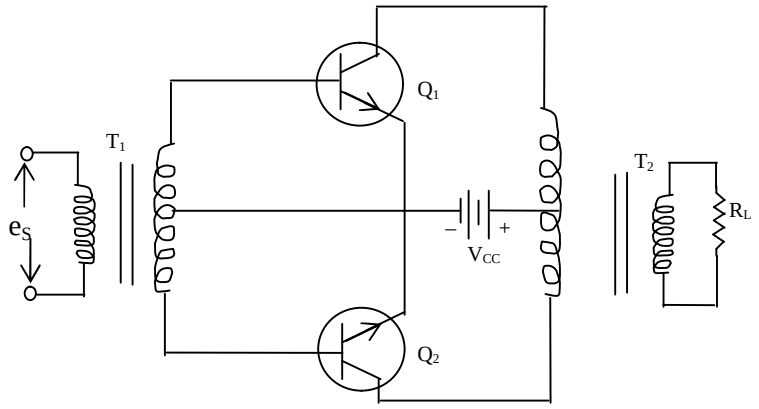


Fig 4.7 Push-pull class – B Transistor power amplifier

During one half-cycle of the input signal (fig.4.7), the upper end of the secondary of the input transformer T_1 becomes positive with respect to its center, and the lower end becomes negative during the next half cycle, the lower end is positive and the upper end is negative with respect to the center.

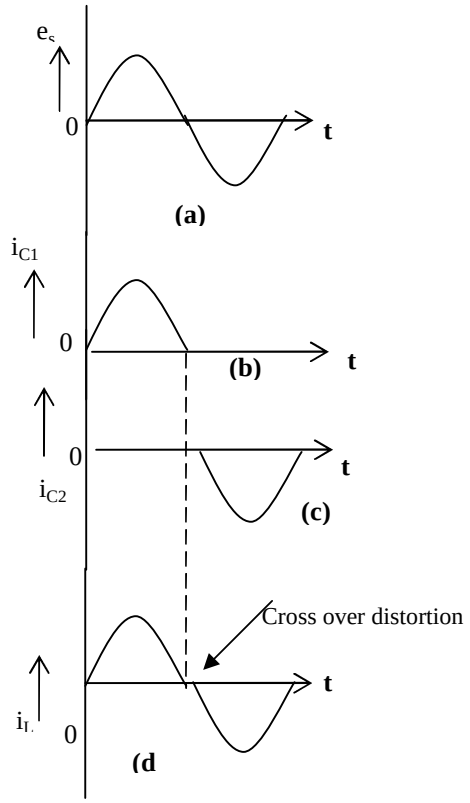


Fig 4.8 Waveform of (a) input voltage (b) collector current of Q_1 (c) collector current of transistor Q_2 and (d) the total load current of the class – B push-pull amplifier

Hence, during one half cycle of the input signal, the transistor Q_1 is forward biased and conducts and the transistor Q_2 remains at cut-off. During the next half cycle of the input signal, the reverse is the case; i.e., Q_2 conducts and Q_1 is cut off. Therefore the collector current for each transistor has half sinusoidal waveform as shown in fig 4.7 (b) and (c).

The total load current i_L is, however, a complete sine wave [fig 4.8(d)] because the collector current for each transistor flows through each half of the primary of the output transformer T_2.

Assuming the dc resistance of the primary of the transformer T_2 to be very small. The dc load line is drawn almost vertical. A collector characteristic of a transistor operating under class – B conditions is shown in (fig 4.9)

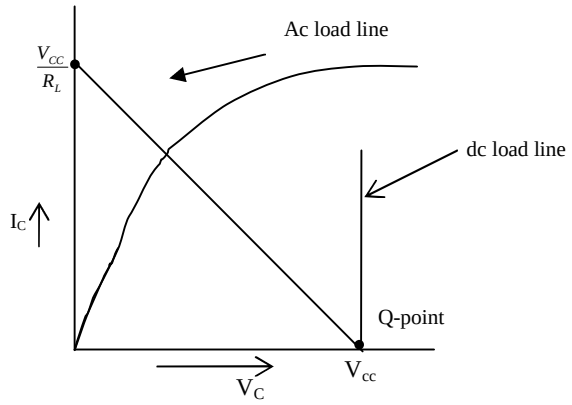


Fig 4.9 The ac and dc load lines on a collector characteristics of a transistor used in a push pull class – B power amplifier.

The reflected resistance across the primary of T_2, if the secondary winding resistance R_S is neglected, is given by

$$R_L^{11} = a^2 R_L \quad \text{---------- (4.13)}$$

The resistance appears from the collector to collector of the two transistors. The load impedance R_L^1 seen by each transistor is that appearing between one end of the transformer and the center tap. Therefore,

$$R_L^1 = \left(\frac{a}{2}\right)^2 R_L = \frac{a^2}{4} R_L \quad \text{----------(4.14)}$$

The ac load line corresponding to R_L^1 is drawn on the collector characteristic of a transistor in the fig 4.9. Since the amplifier operates under class-B condition, the voltage across a transistor is $V_C = V_{CC}$ when the collector $I_C = 0$ for that transistor. Also for $V_C = 0$, I_C is given by $\frac{V_{CC}}{R_L^1}$. Therefore the ac load line passes

through (V_{CC}, 0) and (0, $\frac{V_{CC}}{R_L^1}$) co-ordinates. Since the collector current and voltage are half sinusoidal for class B operation, the relationship between the root mean square (rms) and the peak values are given by

$$V_{C\,rms} = \frac{V_{C\,rmax}}{\sqrt{2}} \quad \text{---------- (4.15)}$$

$$\text{And} \quad I_{C\,rms} = \frac{I_{C\,rmax}}{\sqrt{2}} \quad \text{---------- (4.16)}$$

Hence, the output power due to one transistor is given by

$$P_{ac} = V_{C\,rms}\, I_{C\,rms} = \frac{V_{C\,max}\, I_{C\,max}}{2} \quad \text{---------- (4.17)}$$

As only one transistor is conducting at a time, the total output power is

$$(P_{ac})_{total} = \frac{V_{C\,max}\, I_{C\,max}}{2} \quad \text{---------- (4.18)}$$

Fig 4.9 shows that the maximum collector voltage swing that can occur is given by

$V_{Cmax} = V_{CC}$, hence Eqn, 4.18 becomes

$$(P_{ac})_{total} = \frac{V_{C\,C}\, I_{C\,max}}{2} = \frac{V_{CC}^2}{2R_L^1} \quad \text{---------- (4.19)}$$

Eqn.4.19 gives the maximum output power. Further the dc power supplied by the transistor is given by

$$P_{Cs} = V_{CC}\, I_{Cav} \quad \text{---------- (4.20)}$$

Where I_{Car} the average value of half sinusoidal and is given by

$$I_{Cav} = \frac{I_{Cmax}}{\pi} \quad \text{---------- (4.21)}$$

There fore, the total power supplied by the two transistors is given by

$$(P_{cs})_{total} = 2\,P_{cs} = \frac{2V_{C\,C}\, I_{C\,max}}{\pi} = \frac{2V_{CC}^2}{\pi\, R_L^1} \quad \text{---------- (4.22)}$$

Finally, the maximum efficiency of the class B push-pull power amplifier is

$$\eta_{max} = \frac{(P_{ac})_{total}}{(P_{cs})_{total}} = \frac{V_{CC}^2}{2R_L^1} \times \frac{\pi R_L^1}{2V_{CC}^2} = \frac{\pi}{4} \quad \text{---------- (4.23)}$$

Hence, the efficiency that can be obtained with a push-pull class B operative is $\frac{\pi}{4}$ or 78.5%

4.2 POWER CIRCUITS

4.2.1 VOLTAGE REGULATORS:

A voltage regulator is a circuit that supplies a constant voltage regardless of changes in load currents. Although voltage regulators can be designed using op-amps, it is quicker and easier to use IC (Integrated circuit) voltage regulators. Further more, IC voltage regulators are versatile and relatively inexpensive and are available with features such as a programmable output, current/voltage boosting, internal short circuit current limiting, thermal shutdown, and floating outputs for high voltage applications. IC voltage regulators are of the following types.

i) Fixed output voltage regulators: positive and / or negative output voltage

ii) Adjustable output voltage regulators: positive or negative output voltage.

iii) Switching regulators

iv) Special regulators:

Except for the switching regulators all other types of regulators are called linear regulators. The impedance of a linear regulator's active element may be continuously varied to supply a desired current to the load. On the other hand, in the switching regulator, a switch is turned on and off at a rate such that, the regulator delivers the desired average current in periodic pulses to the load. Because the switching element dissipates negligible power in either on or off state, the switching regulator is more efficient than the linear regulator. In addition, most loads cannot accept the average current in periodic pulses. Therefore, most practical regulators are of the linear type.

Voltage regulators are commonly used for on-card regulator and laboratory type power supplies. Voltage regulators especially the switching type, are used as control circuits in pulse width modulator (PWM), push-pull bridges, and series type switch mode supplies. Almost all power supplies use some type of voltage regulator IC because, voltage regulators are simple to use, reliable, low cost and above all, available in variety of voltage and current ratings. A vast number of voltage regulators are available, data sheets and application notes provided by the manufacturers contain information on the design and use of these devices.

4.2.2 Fixed Voltage Regulators:

4.2.2(a) Series of positive voltage regulators with seven voltage options: The 7800 series consists of three terminal positive voltage regulators with seven voltage options are shown in Fig 4.10(a).

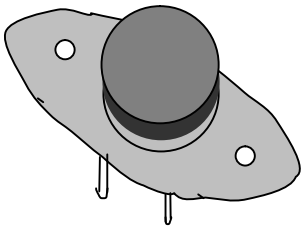
These ICs are designed as fixed voltage regulators with adequate heat sinking and can deliver output currents in excess of one ampere (1A). Although these devices do not require external components, such components can be used to obtain adjustable voltage and currents. These ICs also have internal thermal

over load protection and internal short circuit current limiting. As shown in fig 4.10(c), proper operation requires a common ground between input and output voltages. In addition, the difference between input and output voltages ($V_{in} - V_0$), called drop-out voltage, must be typically 2.0 volts even during the low point on the input ripple voltage. Further more, the capacitor C_i is required if the regulator is located an appreciable distance from the power supply filter. Even through C_0 is not needed it may be used to improve the transient response of the regulator.

Device Type	Output Voltage(V)	Maximum input voltage(V)
7805	5.0	35
7806	6.0	35
7808	8.0	35
7812	12.0	35
7815	15.0	35
7818	18.0	35
7824	24.0	40

Fig 4.10 The 7800 series regulators (a) voltage options

Metal package
(TO-3 type)
Pin1 – Input
Pin2 – Output
Case: Ground



Plastic package
(TO-220 type)
Pin1 – Input
Pin2 – Ground
Pin3- output

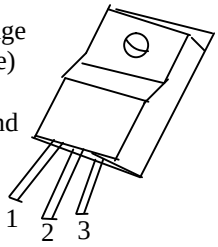
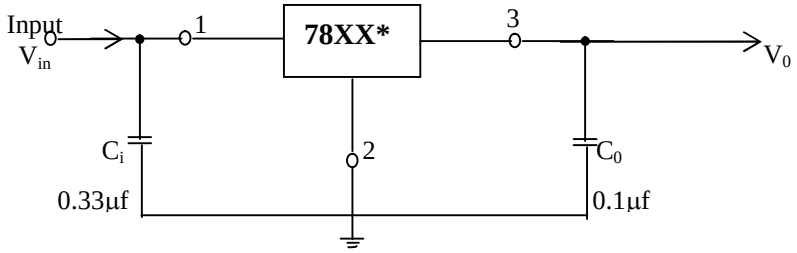


Fig 4.10 (a)The 7800 series regulators (b) Package types



4.10 The 7800 series regulators (c) Standard application
*** XX - These two numbers in type number indicate output voltage**

Typical performance parameter for voltage regulators is line regulation, load regulation, temperature stability and ripple rejection. Line or input regulation is defined as the change in the output voltage for a change in the input voltage and is usually expressed in milli volts or as a percentage of V_0. Temperature stability or average temperature coefficient of output voltage ($T_C V_0$) is the change in output voltage per unit change in temperature and is expressed in either milli volts/^{0}c or parts per million (ppm)/^{0}c. Ripple rejection is the measure of a regulator's ability to reject ripple voltages. It is usually expressed in decibels. The smaller the values of line regulation, load regulation and temperature stability, the better the regulator.

The 7800 regulators, can also be used as current sources. A typical connection diagram of the 7805c as a 0.5A current source is shown in fig 4.11. The current supplied to the load is given by the equation

$$I_L = \frac{V_R}{R} + I_Q \quad \text{----------} \quad (4.24)$$

Where I_Q = quiescent current (amperes)

= 4.3 mA typically for the 7805C

Referring to fig 4.11, $V_R = V_{23} = 5V$ and $R = 10\,\Omega$;

Therefore,

$$I_C \cong 0.5 \text{ A}$$

The output voltage V_0 with respect to ground is

$$V_0 = V_R + V_L \quad \text{----------} \quad (4.25)$$

Where $V_L = I_L R_L$, the load resistance $R_L = 10\,\Omega$, hence, $V_L = 5V$, therefore $V_o = 10V$, since the dropout voltage for the 7805c is 2V, the minimum input voltage required is given by the equation

$$V_{in} = V_0 + \text{Dropout voltage}$$

$$V_{in} = 10 + 2 = 12V$$

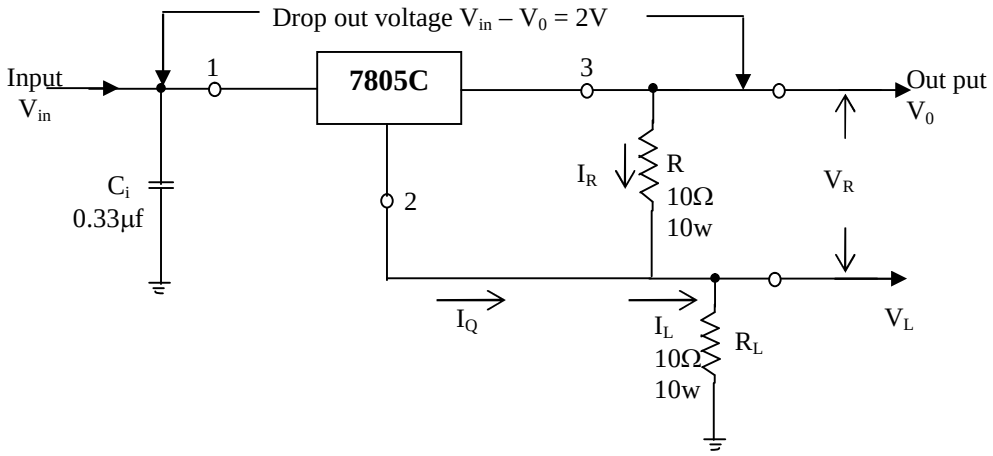


Fig 4.11. The 7805C as a 0.5A current source.

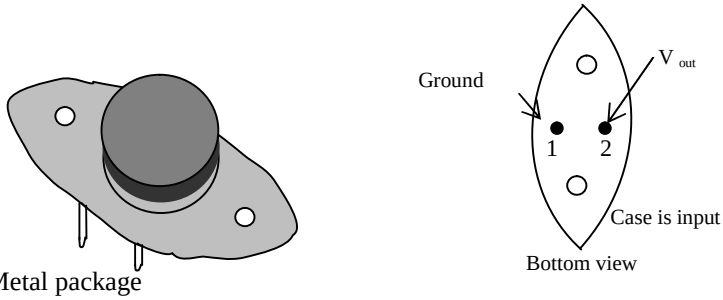
In short, a current source circuit using a voltage regulator can be designed for a desired value of load current (I_L) simply by selecting an appropriate value for R.

4.2.2 (b) Series of negative voltage regulator with mine voltage options:

Device Type	Output Voltage (v)	Maximum input voltage (v)
7902	– 2.0	– 35
7905	– 5.0	– 35
7905.2	– 5.2	– 35
7906	– 6.0	– 35
7908	– 8.0	– 35
7912	– 12.0	– 35
7915	– 15.0	– 35
7918	– 18.0	– 35
7924	– 24.0	– 40

Fig 4.12 The 7900 series voltage regulators (a) Voltage options

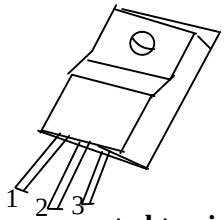
The 7900 series of fixed output negative voltage regulators are complements to the 7800 series devices. Further, these negative voltage regulators are available with the same seven voltage options as the 7800 devices and besides, two extra voltage options, - 2V and –5.2V, are also available in the negative 7900 series as shown in fig 4.12(a), the package types in which the 7900 series voltage regulators available are shown in fig 4.12(b).



Metal package
(TO-3 type)

TO-3 steel metal can

Plastic package
(TO-3 type)
Pin: 1 Ground
2 Input
3 out put



(Heat sink surface connected to pin 2)

Fig 4.12 The 7900 series regulators (b) Package types

4.2.3 Adjustable Voltage Regulators:

The many manufactures that used fixed voltage regulators like the 7800 and 7900 series, in their line of products had to stock and hold an inventory quantities of each voltage in order to always have on hand a specific device for a particular system. Adjustable voltage regulators provided the answers to the excessive inventory and production costs because a single device satisfies many voltage requirements from 1.2 up to 57V. In addition, they have the following performance and relative advantages over the fixed types.

i) Improved system performance by having line and load regulators of a factor of 10 or better

ii) Improved over load protection allows greater output current over operating temperature range.

iii) Improved system reliability with each device being subjected to 100% thermal limit burn-in.

4.2.4 Adjustable Positive Voltage Regulators:

The LM317 are a popular series adjustable three-terminal positive voltage regulators. The different grades of regulators in the series are available with output voltage of 1.2 to 57V and the output current from 0.1 to 1.5 A as shown in figure 4.13(a). The LM 317 series regulators are available in standard transistor packages that are easily mounted and handled [4.13(b)]. The three terminals are V_{IN} , V_{OUT} and

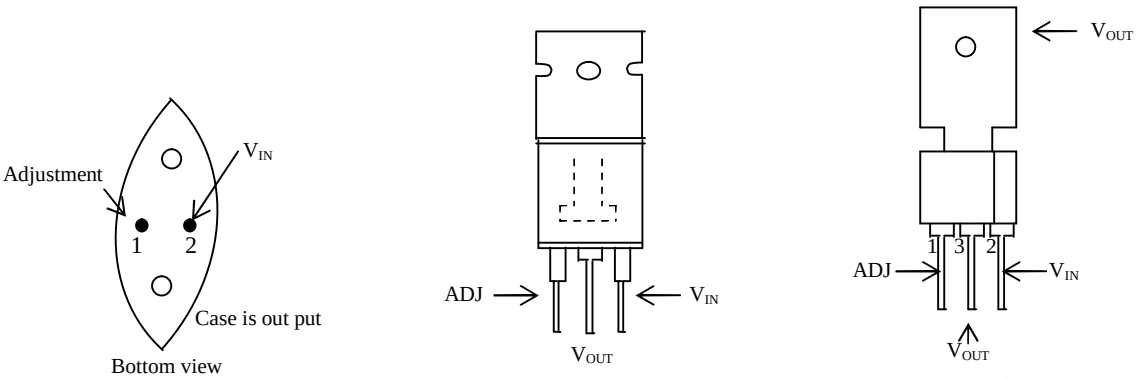
adjustment (ADJ). Fig 4.13(c) shows a typical connection diagram for the LM 317 regulator. From this diagram it is obvious that the LM 317 requires only two external resistors to set the output voltage. When configured as shown in fig 4.13(c), the LM 317 develops a nominal 1.25V, referred to as the reference voltage V_{REF}, between the output and adjustment terminal.

This reference voltage is impressed across resistor R_1 and, since the voltage is constant, the current I_1 is also constant for a given value of R_1. Because resistor R_1 sets current I_1, it is called the current set or program resistor. In addition to the current I_1, the current I_{ADJ} from the adjustment terminal also flows through the output set resistor R_2.

The LM 317 is designed such that I_{ADJ} is very small and constant with line and load changes. The maximum value of adjustment pin current I_{ADJ} is 100μA.

Device	Available V_0 (V)	Output Current (A)	V_{in} max (V)	Ripple rejection (dB)	Package
LM317	1.2 to 37	1.5	40	80	To-39
LM317 H	1.2 to 37	0.5	40	80	To-39
LM317 HV	1.2 to 57	1.5	60	80	To-3
LM317 HVH	1.2 to 37	0.5	40	80	To-39
LM317 L	1.2 to 37	0.10	40	65	To-92
LM317 M	1.2 to 37	0.50	40	80	To-202

Fig 4.13(a) LM317 typical ratings.



TO-3 steel metal can

TO-220 Plastic

Front View TO-220 Plastic Package

Fig 4.13(b) LM 317 standard package types.

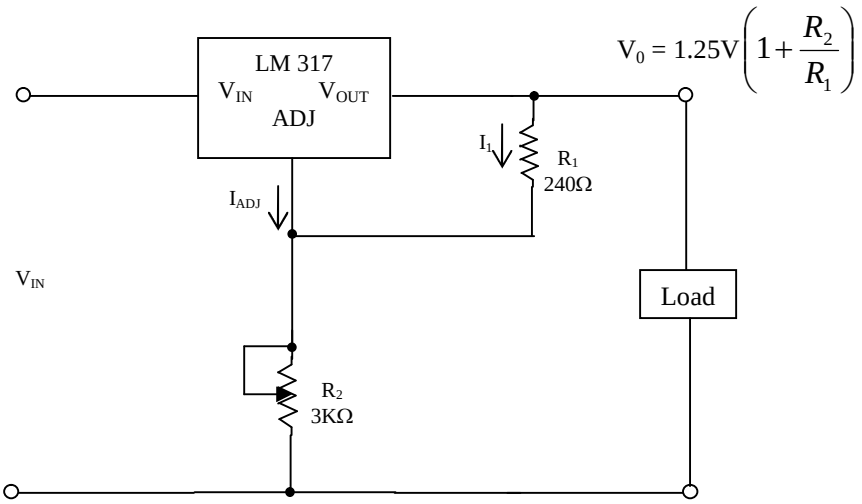


Fig 4.13(c) LM 317A typical connection diagram.

Thus, referring to fig 4.13(c), the output voltage V_0 is

$$V_0 = R_1 I_1 + R_2 (I_1 + I_{ADJ}) \quad \text{----------(4.26)}$$

Where $I_1 = \frac{V_{REF}}{R_1}$; R_1 = Current (I_1) set resistor R_2 = output (V_0) set resistor

I_{ADJ} = adjustment pin current.

Substituting the value of I_1 in equation (4.26) and rearranging, we get

$$V_0 = V_{REF}\left(1+\frac{R_2}{R_1}\right) + I_{ADJ} R_2 \quad \text{----------} \quad (4.27)$$

Where V_{REF} = 1.25V = reference voltage between the output and adjustment terminals. However, the current I_{ADJ} is very small (100μA) and constant. Therefore, the voltage drop across R_2 due to I_{ADJ} is also very small and can be neglected. In short

$$V_0 = 1.25\left(1+\frac{R_2}{R_1}\right) \quad \text{----------} \quad (4.28)$$

Equation (4.28) indicates that the output voltage V_0 is a function of R_2 for a given value of R_1 and can be varied by adjusting the value of R_2 . The current set resister R_1 is usually 240Ω, and to active good load regulation it should be tied directly to the output of the regulator rather than near the load.

4.3 Summary

When large input signals are to be amplified for the operation of output device such as speakers and motors, the amplifier must be capable of handling large amount of power and its efficiency of converting

input dc power to output ac power must be high. Such an amplifier is known as a power amplifier. Power amplifiers are classified on the basis of their operating point as class A, B, AB, C etc.

In class A amplifiers, there is always collector current regardless of the time in the cycle of the applied signals. As a result, power losses many be increased. Since the collector current flows only for one half of the input signal in class B amplifier and in order to get output power for full cycle, two class B amplifiers are used in a combination known as push-pull. Therefore, the power losses may be minimized and simultaneously the collector efficiency also increases in class B push-pull power amplifier:

A voltage regulator is a circuit that supplies a constant voltage regardless of changes in load currents. Monolithic voltage regulators are available on a variety of different output voltage ratings and also quicker and easier to use. The 7800 is a fixed positive voltage regulator series with seven voltage options while the 7900 is a fixed negative voltage regulator series with nine voltage options. Adjustable voltage regulator such as the LM 317 is more popular because of its versatility, performance and reliability.

For adjustable negative voltage regulators see reference Text book.

4.4 Key Terminology:

1. Quiescent (Q) point: The operating point in the absence of signal is called Quiescent point or simply the Q point of the device. The particular Q point at which the device will operate depends on the base current(I_B) and V_{CE}..
2. Direct Coupling: To amplify all the frequencies in the signal, signal source is coupled to the amplifier stage directly without using DC blocking capacitors and isolation transformers. In the same manner amplifier output is connected to the load. Direct coupling may be used to connect to the amplifier stage to the other. However precautions must be taken to provide level shifting so that satisfactory biases are established at various points in the circuit. This type of coupling is preferred in the fabrication of the integrated circuits(ICs).
3. Push-Pull Configuration : Two transistors amplifiers are said to be connected in push-pull configuration when the amplifiers are connected such that as current in one stage pushes forward, in the other, current pulls back. This configuration avoids the power losses in the transformer core and results in higher efficiency than a two transistor parallel coupled amplifier.
4. Regulation : Keeping the output voltage of the DC Power supply constant, irrespective of the variation of the supply voltage and load current is called regulation.
5. Ripple : After rectification of the AC signal using a full-wave rectifier to obtain a DC voltage we find that the output is not a pure DC , but contains a small amount of AC voltage which has to be removed by filtering or by incorporating the regulation circuit.

4.5 Self-assessment questions :

1. Describe the various classes of power amplifier in terms of these operating point.
2. Define efficiency of power amplifier and compare efficiencies in case of A, B and C classes of amplifiers.
3. Obtain an expression for the collector efficiency of transformer coupled class-A transistor power amplifier.
4. Describe a class-B push-pull power amplifier and obtain an expression for its efficiency. How the class-B push-pull amplifier is an advantageous over the class-A amplifier?
5. What do you mean by voltage regulator? List four different types of voltage regulators.
6. What are the advantages of adjustable voltage regulator over the fixed voltage regulators?

4.6 Reference Books:

(1) Operational Amplifiers and Linear Integrated Circuit Technology by Ramakanth A. Gaykwad Prentice Hall Inc.,

(2) Basic Electronics by DC Tayal, Himalaya Publish Co.,

(3) Semi Conductor Electronics by A K. Sharma
New Age International Publishers

(4) Foundations of Electronics
By D. Chattopadhyay, PC Rakshit, B. Saha, M.N. Purkait
Second Edition, Wiley Eastern Ltd.,

Name of Author :

Dr D.Krishna Rao, M.Sc; M.Phil; Ph.D.,
Associate Professor, Physics Department,
Acharya Nagarjuna University,
Nagarjuna Nagar 522 510.

Unit – II

Lesson – 5

Microwaves – wave guides

Objectives:

In this lesson you will learn

(i) Microwaves and its applications

(ii) TE modes in Rectangular Wave-guide

(iii) TM modes in Rectangular wave guide

Structure:

5.1 Introduction:

Microwaves are a part of the electromagnetic spectrum, i.e., radio waves, in the frequency band between 300MHz and 30 GHz (see in fig 5.1).

Referring to fig 5.1 the frequency of microwaves is placed in between radio frequencies and light wave frequencies. From this we can conclude that terms and theories from both these fields are used to describe microwaves. Further the wavelength is in the same magnitude as the physical dimensions of the components. This compares with alternating currents, where a wavelength is 6000Km. Fig 5.2a gives an idea about electromagnetic waves.

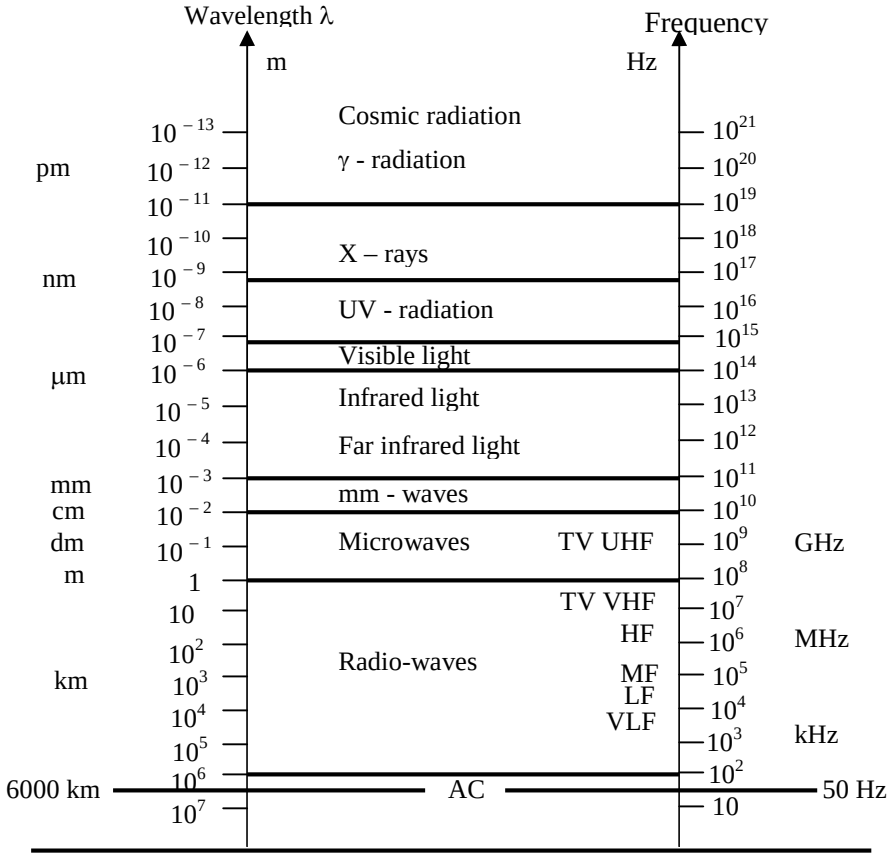


Fig 5.1 Electromagnetic spectrum

These ideas can be summarized as follows.

(a) Electromagnetic waves (EM –Waves) are generated by charged particles, e.g. electrons, in constant motion. The wave can propagate through free space, i.e., it does not need any medium for its propagation.

(b) A charged particle has one electric field (E – field).

(c) A change in the E – field generate a magnetic field (H – field)

A change in the H – field causes a further change in E – field.

(d) The E – and H – fields are perpendicular to each other and to the propagation direction.

Referring to fig 5.2(b) the wavelength λ_0 is defined as the distance between two successive points in the propagation direction having the same phase. The EM-wave can also propagate in transmission lines, for instance in a co-axial cable or a wave-guide and is then associated with voltages and currents in the conductors. With suitable antenna arrangement the wave will radiate into free space.

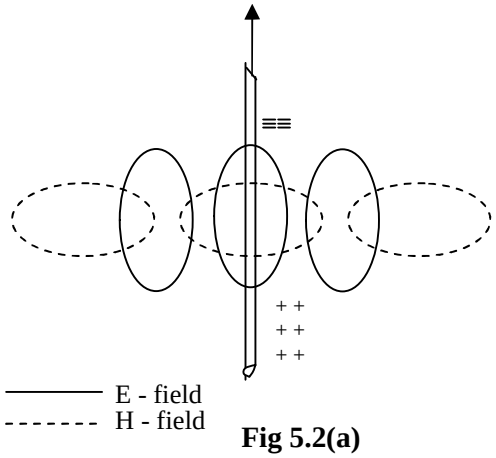


Fig 5.2(a)

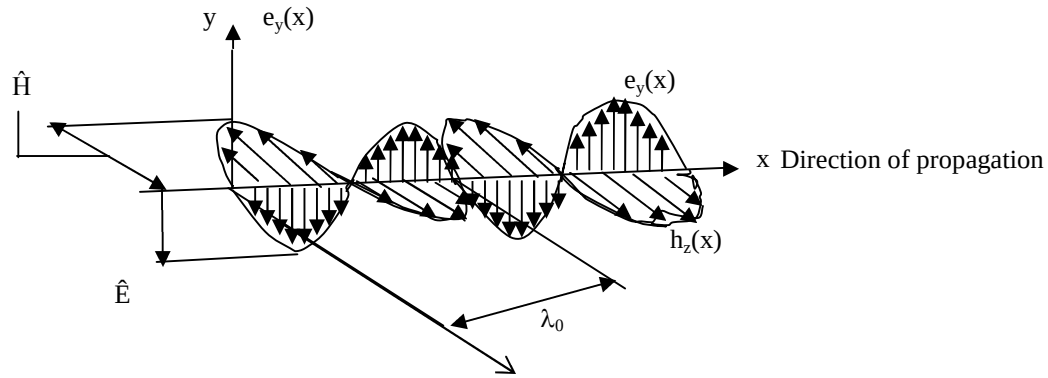


Fig 5.2(b)

5.1.1 Applications of Microwaves:

There are a number of areas where microwave techniques are used for practical applications. Some of them are

i) In aviations surveillance and navigation, microwave radars systems are being used. In meteorology, radars are used in weather forecasting. Shipyards and airports use radar for navigation and to pilot traffic.

ii) Telecommunications use microwaves to convey telephone and TV signals between

continents via satellites and modern communications systems use microwave links.

iii) Microwave radiation is used for heating materials in industrial applications and in food processing.

iv) Radio astronomy reveals existence of distant galaxies with microwaves and space-vehicles land and dock using microwave navigational aids.

v) Microwaves are used in measuring moisture contents in food, paper etc.

vi) Automatic door openers and bugler alarms often use microwaves.

vii) Microwaves are also used for speed control, short distance measurement, short distance directed communications e.g, in dirty and smoky areas.

viii) Microwaves have also scientific and medical applications.

5.2 Maxwell's equations:

The Ampere's law, faraday Law and Guass's Law lead to Maxwell equation.

Ampere's law:

In its basic form, it is given by

$$\int \overline{H} \cdot dl = I \quad \text{----------} \quad (5.1)$$

where dl is the incremental length of the closed loop, $\overline{H}$ in the magnetic field vector, and $\bar{I}$ is the current enclosed by the loop.

The current $\bar{I}$ can be replaced by the surface integral of the conduction current density $\overline{J_T}$ over the area bounded by the path of integration.

The total current density is given by

$$\overline{J_T} = \bar{J} + \frac{\partial \overline{D}}{\partial t} \quad \text{----------} \quad (5.2)$$

Where $\bar{J} = \sigma \overline{E}$ in the conduction current density and $\frac{\partial \overline{D}}{\partial t}$ is the displacement current density.

Displacement current is the current flowing in pure capacitor dielectric to which an ac voltage is applied. Current does not flow through the dielectric but the external effect is as though it did and so we imagine a current there, which we call as displacement current in the die electric. Using eqn 5.2 in eqn 5.1 we have

$$\int \overline{H} \cdot dl = \int_s \left(\bar{J} + \frac{\partial \overline{D}}{\partial t} \right) . ds \quad \text{----------} \quad (5.3)$$

Where 'S' stands for surface integration. Equation (3) is the integral form of Maxwell's 1st equation

By applying Stokes theorem

$$\int \overline{H} \,.\, dl = \int_s \left(\nabla \times \overline{H}\right) ds$$

$$\int_s \left(\nabla \times \overline{H}\right) ds = \int_s \left(\bar{J} + \frac{\partial \overline{D}}{\partial t} \right) .ds$$

Therefore, the differential form of Maxwell's 1st equation can be written as.

$$\nabla \times \overline{H} = \bar{J} + \frac{\partial \overline{D}}{\partial t} \quad \text{----------} \quad (5.4)$$

Faraday's Law:

In its basic form it is given by

$$\overline{V} = -\frac{d\phi}{dt}$$

$$\text{Since } \phi = \int_s \overline{B}.ds \text{ and } V = \int \overline{E} \,.\, dl$$

$$\int \overline{E} \,.\, dl = - \frac{\partial}{\partial t} \int_s \overline{B}.ds \quad \text{----------} \quad (5.5)$$

Equation (5) is the integral form of Maxwell's 2nd equation.

Applying Stokes theorem we get the differential form of Maxwell's 2nd equation.

$$\nabla \times \overline{E} = -\frac{\partial \overline{B}}{\partial t} \quad \text{----------} \quad (5.6)$$

Gauss's Law:

In its basic form it is given by

$$\int_S \overline{D} \,.\, ds = \int_V \rho \, dv \quad \text{----------} \quad (5.7)$$

Where, S = surface of integration

V = volume enclosed by it,

ρ = charge density

$$\lim_{dv \to 0} \int \frac{\overline{D}.ds}{dv} = \nabla.\overline{D} \quad \text{(Diversence of } \overline{D}\text{)}$$

The 3rd Maxwell's equation is therefore given by

$$\nabla.\overline{D} = \rho \quad \text{----------} \quad (5.8)$$

In the case of magnetic fields the surface integral of '$\overline{B}$' over a closed surface is zero. This is because a magnetic field has neither a source nor a sink.

The fourth Maxwell's equation is given by

$$\nabla . \overline{B} = 0 \text{ ---------- } (5.9)$$

5.3 Rectangular wave-guides and wave propagation:

A rectangular wave guide is a hollow metallic tube with a rectangular cross section. The conducting walls of the guide confine the electro magnetic fields and thereby guide the electromagnetic wave. A number of distinct field configurations or modes can exist in a wave -guide. When the wave travel longitudinally down the guide, the plane waves are reflected from wall to wall. This process results in a component of either electric or magnetic field in the direction of propagation of the resultant wave; therefore the wave is no longer a transverse electromagnetic (TEM) wave.

The electromagnetic wave inside a waveguide can have an infinite number of patterns, which are called modes. There are two kinds of modes in a waveguide. They are transverse electric mode (TE) and Transverse Magnetic mode(TM).

In the first type, the electric field is always transverse to the direction of propagation and is called the Transverse Electric or (TE) wave.

In the second type, the magnetic field is always transverse to the direction of propagation and is called the Transverse Magnetic or (TM) wave.

Thus in a TE mode, no electric line is in direction of propagation i.e. $\overline{E}_Z = 0$, if Z is the direction of the propagation. But $\overline{H}_z \neq 0$.

In a TM mode, number of magnetic lines is in direction of propagation i.e., $\overline{H}_z = 0$ but $\overline{E}_Z \neq 0$.

5.4 TE modes in Rectangular wave-guides :

Figure 5.3 shows the coordinates of a rectangular waveguide.

The TE_{mn} modes in a rectangular guide are characterized by $E_z = 0$. In other words, the z component of the magnetic field, H_z, must exist in order to have energy transmission in the guide.

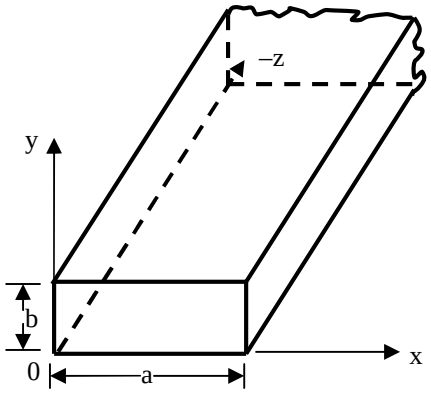


Fig 5.3 coordinates of a rectangular guide.

Consequently, from Helmholtz equation.

$$\nabla^2 H_Z = \gamma^2 H_Z \quad \text{----------} \quad (5.10)$$

a solution in the form of

$$H_Z = \left[A_m \sin\left(\frac{m\pi x}{a}\right) + B_m \cos\left(\frac{m\pi x}{a}\right)\right] \times \left[C_n \sin\left(\frac{n\pi y}{b}\right) + D_n \cos\left(\frac{n\pi y}{b}\right)\right] e^{-j\beta_g z} \quad \text{----- (5.10(a))}$$

will be determined in accordance with the given boundary conditions, where $k_x = \frac{m\pi}{a}$ and $k_y = \frac{m\pi}{b}$ are replaced. For a loss less dielectric, Maxwell's curl equations in frequency domain are

$$\nabla \times E = - j\omega\mu H \quad \text{----------} \quad (5.11)$$

$$\nabla \times H = j\omega \in E \quad \text{----------} \quad (5.12)$$

In rectangular coordinates, their components are

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = - j\omega\mu H_x \quad \text{----------} \quad (5.13)$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = - j\omega\mu H_y \quad \text{----------} \quad (5.14)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = - j\omega\mu H_z \quad \text{----------} \quad (5.15)$$

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = j\omega \in E_x \quad \text{----------} \quad (5.16)$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = j\omega \in E_y \quad \text{----------} \quad (5.17)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega \in E_z \quad \text{----------} \quad (5.18)$$

With the substitution $\frac{\partial}{\partial z} = -j\beta_g$ and $E_z = 0$, foregoing equations are simplified to

$$\beta_g E_y = -\omega\mu H_z \quad \text{----------} \quad (5.19)$$

$$\beta_g E_x = -\omega\mu H_y \quad \text{----------} \quad (5.20)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = - j\omega\mu H_z \quad \text{----------} \quad (5.21)$$

$$\frac{\partial H_z}{\partial y} - j\beta_g H_y = j\omega \in E_x \quad \text{----------} \quad (5.22)$$

$$-j\beta_g H_x - \frac{\partial H_z}{\partial x} = j\omega \in E_y \quad \text{----------} \quad (5.23)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = 0 \quad \text{----------} \quad (5.24)$$

Solving these six equations for E_{xy} E_{yz} H_{xz} and H_y in terms of H_z will give the TE – mode field equations in rectangular wave-guide as

$$E_x = \frac{-j\omega\mu}{k_c^2}\frac{\partial H_z}{\partial y} \quad \text{----------} \quad (5.25)$$

$$E_y = \frac{j\omega\mu}{k_c^2}\frac{\partial H_z}{\partial y} \quad \text{----------} \quad (5.26)$$

$$E_z = 0 \quad \text{----------} \quad (5.27)$$

$$H_x = \frac{-j\beta_g}{k_c^2}\frac{\partial H_z}{\partial x} \quad \text{----------} \quad (5.28)$$

$$H_y = \frac{-j\beta_g}{k_c^2}\frac{\partial H_z}{\partial y} \quad \text{----------} \quad (5.29)$$

$$H_z = \text{Eq. (5.10(a))} \quad \text{----------} \quad (5.30)$$

Where $k_c^2 = \omega^2\mu\in - \beta_g^2$ has been replaced.

Differentiating Eq. (5.10(a)) with respect to x and y and then substituting the results in Eqs. (5.25) through (5.29) yield a set of field equations. The boundary conditions are applied to the newly found field equations in such a manner that either the tangent E field or the normal H field vanishes at the surface of the conductor. Since $E_x = 0$, then $\frac{\partial H_z}{\partial y} = 0$ at y = 0, b. Hence $C_n = 0$. since $E_y = 0$, then $\frac{\partial H_z}{\partial x} = 0$ at x = 0, a. Hence $A_m = 0$.

It is generally concluded that the normal derivative of H_z must vanish at the conducting surfaces – that is,

$$\frac{\partial H_z}{\partial n} = 0 \quad \text{----------} \quad (5.31)$$

at the guide walls. Therefore the magnetic field in the positive z direction is given by

$$H_z = H_{0z}\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-j\beta_g z} \quad \text{----------} \quad (5.32)$$

Where H_{0z} is the amplitude constant.

Substitution of Eq. (5.32) in Eqs. (5.25) through (5.29)yields the TE_{mn} field equations in rectangular waveguide as

$$E_x = E_{ox} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \quad \text{---------- (5.33)}$$

$$E_y = E_{oy} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \quad \text{---------- (5.34)}$$

$$E_z = 0 \quad \text{---------- (5.35)}$$

$$H_x = H_{ox} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-j\beta g Z} \quad \text{---------- (5.36)}$$

$$H_y = H_{oy} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \quad \text{---------- (5.37)}$$

$$H_z = \text{Eq.(5.32)} \quad \text{---------- (5.38)}$$

Where m = 0, 1, 2, ……

n = 0, 1, 2, ……

m = n = 0 excepted

The cutoff wave number k_c, for the TE_{mn} modes, is given by

$$k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} = \omega_c \sqrt{\mu \in} \quad \text{---------- (5.39)}$$

where ‘a’ and ‘b’ are in meters. The cutoff frequency for the TE_{mn} modes, is given by

$$f_c = \frac{1}{2\sqrt{\mu \in}} \sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}} \quad \text{---------- (5.40)}$$

The propagation constant (or the phase constant here)β_g, is expressed by

$$\beta_g = \omega \sqrt{\mu \in} \sqrt{1 - \left(\frac{f_c}{f}\right)^2} \quad \text{---------- (5.41)}$$

The phase velocity in the positive z direction for the TE_{mn} modes is shown as

$$v_g = \frac{\omega}{\beta_g} = \frac{v_p}{\sqrt{1 - (f_c / f)^2}} \quad \text{---------- (5.42)}$$

Where $v_p = \frac{1}{\sqrt{\mu \in}}$ is the phase velocity in an unbounded dielectric.

The characteristic wave impedance of TE_{mn} modes in the guide can be shown to be:

$$Z_g = \frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{\omega\mu}{\beta_g} = \frac{\eta}{\sqrt{1-(f_c/f)^2}} \quad \text{---------- (5.43)}$$

Where $\eta = \sqrt{\frac{\mu}{\in}}$ is the intrinsic impedance in an unbounded dielectric. The wavelength 'λ_g' in the guide for the TE_{mn} modes is given by

$$\lambda_g = \frac{\lambda}{\sqrt{1-(f_c/f)^2}} \quad \text{---------- (5.44)}$$

Where $\lambda = \frac{v_p}{f}$ is the wavelength in an unbounded dielectric.

Whenever two or more modes have the same cutoff frequency, they are said to be degenerate modes. In a rectangular guide the corresponding TE_{mn} and TM_{mn} modes are always degenerate. In a square guide the TE_{mn}, TE_{nm}, TM_{mn}, and TM_{nm} modes form a foursome of degeneracy. Rectangular guides ordinarily have dimensions of a = 2b ratio. The mode with the lowest cutoff frequency in a particular guide is called the dominant mode. The dominant mode in a rectangular guide with a > b is the TE_{10} mode. Each mode has a specific mode pattern (or field pattern).

5.7 TM modes in Rectangular wave-guides:

The TM_{mn} modes in a rectangular guide are characterized by $H_z = 0$. In other words, the z component of an electric field E must exist in order to have energy transmission in the guide. Consequently, the Helmholtz equation for E in the rectangular coordinates is given by

$$\nabla^2 E_z = \gamma^2 E_z \quad \text{---------- (5.45)}$$

A solution of the Helmholtz equation is in form of

$$E_z = \left[A_m \sin\left(\frac{m\pi x}{a}\right) + B_m \cos\left(\frac{m\pi x}{a}\right)\right]\left[C_n \sin\left(\frac{n\pi y}{b}\right) + D_n \cos\left(\frac{n\pi y}{b}\right)\right] e^{-j\beta_g z} \quad \text{------- (5.46)}$$

Which must be determined according to the given boundary conditions. The procedure for doing so are similar to those used in the case of TE – mode wave.

The boundary conditions on E_z require that the field vanish at the waveguide walls, since the tangent component of the electric field E_z is zero on the conducting surface. The requirement is that for $E_z = 0$ at x = 0 and a, $B_m = 0$, and for $E_z = 0$ at y = 0 and b, $D_n = 0$. Thus the solution as shown in Eq (5.46) reduces to

$$E_z = E_{oz} \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \quad \text{----------} \quad (5.47)$$

Where m = 1, 2, 3, …..

n = 1, 2, 3, …..

If either m = 0 or n = 0, the field intensities all vanish. So there is no TM_{01} or TM_{10} mode in a rectangular wave guide, which means that TE_{10} is the dominant mode in a rectangular waveguide for a > b, for $H_z = 0$, the field equations, after expanding $\nabla \times H = j\omega \in E$, are given by

$$\frac{\partial E_z}{\partial y} + j\beta_g E_y = -j\omega\mu H_x \quad \text{----------} \quad (5.48)$$

$$j\beta_g E_x + \frac{\partial E_z}{\partial x} = j\omega\mu H_y \quad \text{----------} \quad (5.49)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0 \quad \text{----------} \quad (5.50)$$

$$\beta_g H_y = \omega \in E_x \quad \text{----------} \quad (5.51)$$

$$-\beta_g H_x = \omega \in E_y \quad \text{----------} \quad (5.52)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega \in E_z \quad \text{----------} \quad (5.53)$$

These equations can be solved simultaneously for E_x, E_y, H_x, and H_y in terms of E_z.

The resultant field equations for TM modes are

$$E_x = \frac{-j\beta_g}{k_c^2} \frac{\partial E_z}{\partial x} \quad \text{----------(5.54)}$$

$$E_y = \frac{-j\beta_g}{k_c^2} \frac{\partial E_z}{\partial y} \quad \text{----------} \quad (5.55)$$

$$E_z = E_{oz} \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \quad \text{----------} \quad (5.56)$$

$$H_x = \frac{j\omega \in}{k_c^2} \frac{\partial E_z}{\partial y} \quad \text{----------} \quad (5.57)$$

$$H_y = \frac{-j\omega \in}{k_c^2} \frac{\partial E_z}{\partial x} \quad \text{----------} \quad (5.58)$$

$$H_z = 0 \quad \text{----------} \quad (5.59)$$

Where $\beta_c^2 - \omega^2\mu\in = -K_c^2$ is replaced.

Differentiating Eq.(5.47) with respect to 'x' or 'y' and substituting the results in Eqs.(5.54) through (5.59) yield a new set of field equations the TM_{mn} mode field equations in rectangular waveguides are

$$E_x = E_{ox} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \quad \text{---------- (5.60)}$$

$$E_y = E_{oy} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \quad \text{---------- (5.61)}$$

$$E_z = \text{Eqs.(5.47)} \quad \text{---------- (5.62)}$$

$$H_x = H_{ox} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \quad \text{---------- (5.63)}$$

$$H_y = H_{oy} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \quad \text{---------- (5.64)}$$

$$H_z = 0 \quad \text{---------- (5.65)}$$

Some of the TM – mode characteristic equations are identical to those of the TE – modes, but some are different, For convenience, all are shown here:

$$f_c = \frac{1}{2\sqrt{\mu \in}} \sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}} \quad \text{---------- (5.66)}$$

$$\beta_g = \omega\sqrt{\mu \in} \sqrt{1 - \left(\frac{f_c}{f}\right)^2} \quad \text{---------- (5.67)}$$

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (f_c / f)^2}} \quad \text{---------- (5.68)}$$

$$v_g = \frac{v_p}{\sqrt{1 - (f_c / f)^2}} \quad \text{---------- (5.69)}$$

$$Z_g = \frac{\beta_g}{\omega \in} = \frac{\eta}{\sqrt{1 - (f_c / f)^2}} \quad \text{---------- (5.70)}$$

5.8 Coaxial Lines:

The coaxial transmission line is the most widely used transmission line for high frequency applications. A cross – sectional view of a typical coaxial cable is shown in Fig 5.4a and the TEM mode field supported by it is shown in Fig 5.4b.

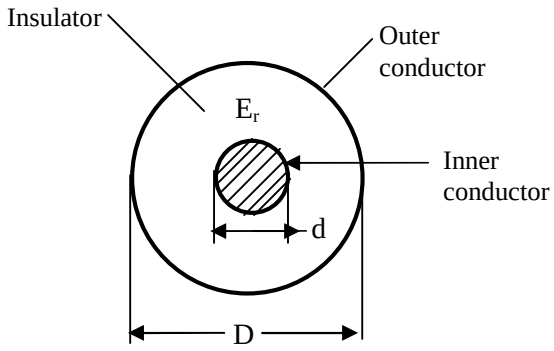


Fig 5.4(a) Cross – sectional view

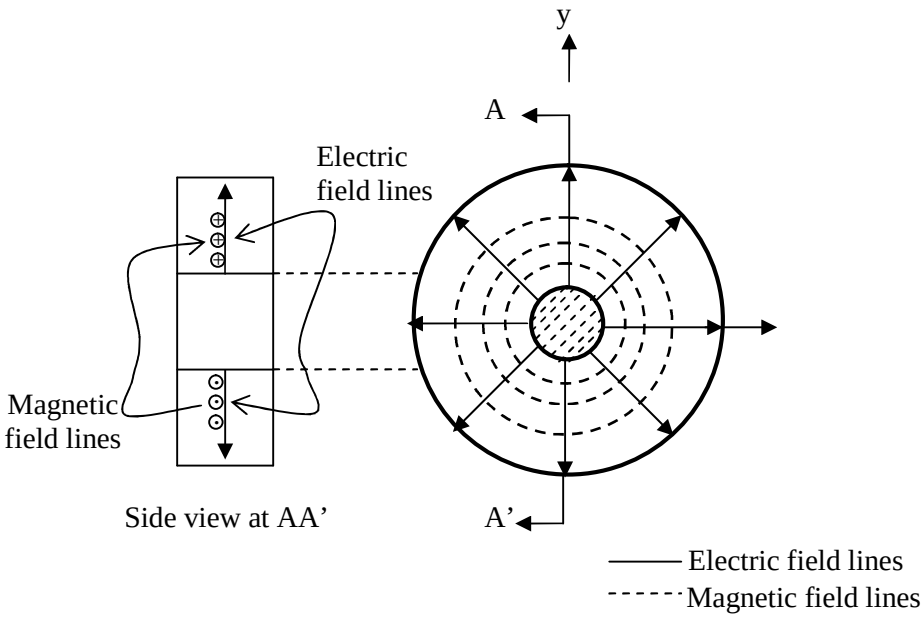


Fig 5.4(b) TEM wave in coaxial lines

A coaxial line consists of an inner conductor with diameter `d' which is surrounded by a concentric cylinder of an insulating material with a dielectric constant of $\in$,. The outer conductor is a concentric cylinder with inner diameter D. The coaxial cable is an unbalanced transmission line since the outer conductor is normally at ground potential. The dominant mode in a coaxial line is the TEM mode although higher modes do exist (Fig 5.4b) at high frequencies. The lowest order and highest order – modes in coaxial cables are TE_{11} and TM_{01}. The cut off wavelengths of these modes are $\lambda_c\,(TE_{11}) = \pi(D + d)$ and $\lambda_c\,(TM_{01}) = 2(D - d)$. Hence the average circumference of the inner and outer conductors in the propagating cross-section of the coaxial line should be less than the operating wavelength to prevent higher order mode interference.

As shown in the mode patterns, coaxial lines are non-radiating since the e.m. fields are confined in the insulator between the inner and outer conductors. Coaxial lines can operate well up to 40GHz due to development of precision connectors for smaller diameter coaxial cables.

The inductance and capacitance per unit length of a coaxial cable are given by

$$L = \frac{\mu}{2\pi} \ln \frac{D}{d} \text{ H/m; } C = \frac{2\pi \in}{\ln(D/d)} \text{ F/m}$$

Since $\mu = 1$ and $\mu_0 = 4\pi \times 10^{-7}$ H/m and $\in = \in_r \in_0$ and $\in_0 = 8.854 \times 10^{-12}$ F/m.

$$L = 2 \times 10^{-7} \ln \frac{D}{d} \text{ H/m and } C = 55.56 \times 10^{-12} \frac{\in_r}{\ln(D/d)} \text{ F/m. ---------- (5.71)}$$

Hence the characteristic impedance of coaxial lines will be,

$$R_0 = \frac{1}{2\pi} \sqrt{\frac{\mu}{\in}} \ln \frac{D}{d} \Omega$$

Since $\mu = 1$ for non-magnetic materials, substituting for $\in_0$ and μ_0, we get

$$R_0 = \frac{60}{\sqrt{\in_r}} \ln \frac{D}{d} \Omega \text{ ---------- (5.72)}$$

The velocity of propagation for the coaxial cable is given by

$$v = \frac{1}{\sqrt{\mu \in}} = \frac{c}{\sqrt{\mu \in}} \text{ m/s}$$

Since $c = \frac{1}{\sqrt{\mu_0 \in_0}}$ and $\mu_0 = 1$ for coaxial cables,

$$v = \frac{c}{\sqrt{\in_r}} \text{ m/s ---------- (5.73)}$$

Also since TEM mode does not have a cut off frequency, a coaxial cable / line is a broad band device. The electric and magnetic field for a wave propagating in the z direction are given by

$$E = E_t = E_0 (\rho, \phi) e^{-j\beta Z} = \frac{V_0}{\ln(D/d)} \frac{\hat{\rho}}{\rho} e^{-j\beta Z} \text{ ---------- (5.74)}$$

$$H = H_t = \pm \frac{\hat{z} E_0 (\rho, \phi)}{\eta} e^{-j\beta Z} = \frac{V_0}{\eta \ln(D/d)} \frac{\hat{\theta}}{\rho} e^{-j\beta Z} \text{ ---------- (5.75)}$$

Where $\beta = \omega\sqrt{\mu \in}$ and $\eta = \sqrt{\frac{\mu}{\in}}$, the wave impedance of the TEM wave.

The current density on the outer surface of the inner core is given by

$$J_s = \hat{n} \times H = \hat{\rho} \times H_t = \frac{V_0 e^{-\beta Z}}{\eta \ln(D/d) d} \quad \text{---------- (5.76)}$$

The power flow through the coaxial line is given by

$$P = \frac{1}{2} \mathrm{Re} \int_d^D \int_0^{2x} E \times H \, ds = \frac{\pi V_0^2}{\eta \ln(D/d)} \quad \text{---------- (5.77)}$$

Where $ds = \rho \, d\rho \, d\phi$

The power loss in the coaxial cable is due to finite conductivity (σ) of the conductor and the dielectric loss in the insulator between the inner and outer conductors.

For low loss lines,

$$\alpha = \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}} \right) \quad \text{---------- (5.78)}$$

where $R = \frac{R_s}{2\pi}\left(\frac{1}{d} + \frac{1}{D}\right)$

$G = \frac{\omega \in'' C}{\in'}$

and 'L' and 'C' are given by Eq.(5.71) with $G \to 0$, the attenuation 'α' is minimum for D/d = 3.6. With an air dielectric ($\in_r = 1$) and D/d = 3.6 gives a characteristic impedance (Z_0) of 76.86Ω. With solid polyethylene dielectric ($\in_r = 2.3$) and D/d = 3.6 gives a Z_0 of 50.67Ω. Generally 50Ω coaxial lines are preferred for minimum attenuation. For low power transmission, coaxial cables can operate up to 40 GHz whereas for high power transmission they can be employed only up to 34 GHz.

5.9 Summary:

- In wave guides the electric and magnetic fields are confined to the space within the guide.
- It is possible to propagate several modes of electromagnetic waves within a wave guide. These modes correspond to solution of Maxwell's equations for particular wave guide.
- The process of solving the wave guide problems may involve three steps.
- The desired wave equations are written in the form of either rectangular or cylindrical coordinate systems suitable to the problem at hand.
- The boundary conditions are then applied to the wave equations set up in step 1.

- The resultant equations usually are in the form of partial differential equations in either time or frequency domain. They can be solved by using suitable methods.

5.10 Self – Assessment Questions:

1. What are microwaves? Discuss the applications of microwaves?
2. What are TE, TM and TEM modes? Explain?
3. Derive the wave equation for a TM wave and obtain the field components in a rectangular wave guide?
4. Derive the wave equation for a TE wave and obtain all the field components in a rectangular wave guide?
5. Derive the expression for the wave impedance for the TEM wave in coaxial line?

5.11 Key Terminology:-

Coaxial line: A transmission line in which one conductor completely surrounds the other, the **t**wo being coaxial and separated by a continuous solid dielectric or by dielectric spacers. Such a line is characterized by no external field and by having no susceptibility to external fields from other sources.

Giga cycle: 10^9 cycles. Common term for expressing microwave frequencies.

Guide wavelength: The length of waveguide corresponding to one cycle of variation in the axial direction.

Mode: A form of propagation of guided waves that is characterized by a particular field pattern in a plane transverse to the direction of propagation. The field pattern is independent of the position along the axis of the waveguide and, for uni-conductor wave guide, independent of frequency.

Microwave region: That portion of the electromagnetic spectrum lying between the far infrared and conventional RF portion. Commonly regarded as extending from 1,000 megacycles to 300,000 megacycles.

Cut-off frequency: The lowest frequency at which loss-less waveguide will propagate energy in some particular mode without attenuation.

Attenuation: Decrease in magnitude of current, voltage, or power of a signal in transmission between points.

Attenuation constant: For a traveling plane wave of a given frequency, the rate of exponential decrease of the amplitude of a field component in the direction of propagation. Expressed in neperes or db per unit length.

5.12 References:

1. Microwave Engineering – Rajeswari Chatterjee
   Affiliated East – West Press Pvt. Ltd. – New Delhi.
2. Microwave Principles – Herbert. J. Reich
   Affiliated East – West Press Pvt. Ltd. – New Delhi.
3. Microwave Propagation and Techniques – D.C. Sarkar – S. Chand & Company Ltd.
4. Foundation for Microwave Engineering – Collin, R.E., Mc Graw – Hill, New York.

Name of the Author:

S.V. RangaNayakulu, M.Sc., M.Phil., M.Ed.,
Dept of Physics & Electronics,
T.J.P.S. COLLEGE (P.G.COURSES)
GÜNTUR-522 006

Unit – II

Lesson – 6

Micro wave Tubes, Components and Resonators

In this lesson you are able to explain the

1. Principle and working of klystron
2. Reflex klystron
3. Magic Tee
4. Microwave Resonators

Structure of the Lesson:

6.1 Introduction:

At microwave frequencies, the size of electronic devices required for generation of microwave energy becomes smaller and smaller.

Conventional Triodes, Tetrodes and Pentodes are useful only at low microwave frequencies. Special tubes would be required even at UHF frequencies (300 – 3000MHz) as conventional tubes have certain limitations

at microwave frequencies. However all tubes including UHF tubes are bound to fail at higher microwave frequencies due to their limitations at these frequencies.

In this chapter we shall discuss the most important microwave tubes such as two cavity klystron, reflex klystron, and magnetrons. In addition to the tubes, we shall discuss the importance and applications of microwave Resonators and Magic Tee's.

6.2 Klystron:

A klystron is a velocity-modulated tube, in which the velocity modulation process produces a density-modulated stream of electrons. The earliest form of velocity variation device is the "Two cavity klystron amplifier", whose schematic diagram is shown in fig 6.1. It is seen that a high velocity electron beam is formed, focused and sent down along glass tube to a collector electrode, which is a high positive potential with respect to the cathode. Magnetic focusing is used here, but the arrangement was not shown in the figure for the sake of simplicity. As it is clear from the schematic fig 6.1, a two-cavity klystron amplifier consists of a cathode, focusing electrodes; two buncher grids separated by a very small distance forming a gap A, two catcher grids with a small gap B followed by a collector. The significance of buncher and catcher grids will be clear from the following discussion.

The input and output are taken from the tube via resonant cavities. The separation between buncher grids and catcher grids is called drift space. The electron beam passes gap A in the buncher cavity to which RF signal to be amplified is applied and is then allowed to drift freely without any influence from RF. fields until it reaches gap B in the output or catcher cavity.

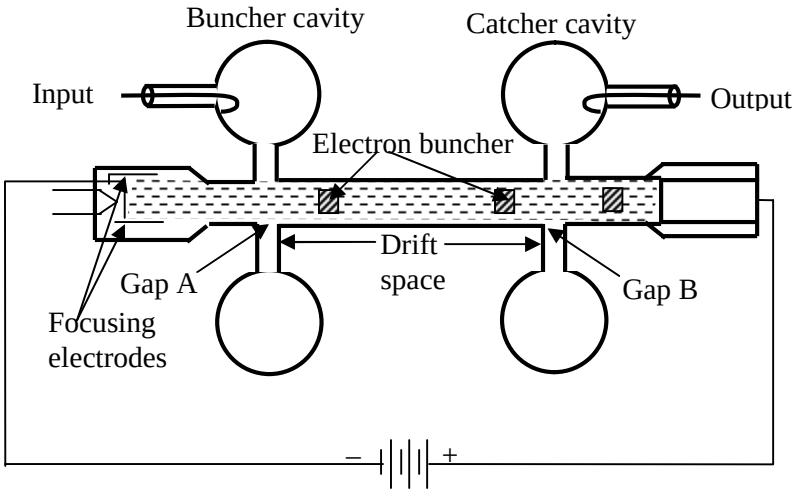


Fig 6.1 Schematic diagram of a klystron amplifier.

The first grid (focusing grid) controls the number of electrons in the electron beam and serves to focus the beam. The beam accelerating potential determines velocities of electrons in the beam. On leaving the region of first grid, the electrons pass through the grids of buncher cavity. The grids of the cavity allow the electrons to pass through, but confine the magnetic fields within the cavity. The space between the grids is referred to as interaction space. When electrons travel through this space they are subjected to RF potentials at a frequency determined by the cavity resonant frequency or the input frequency. The amplitude of this RF potential between the grids is determined by the amplitude of the incoming signal in case of the amplifier, or by the amplitude feedback signal from the second cavity if used as an oscillator. If all goes well, oscillations will be excited in the second cavity, which are of a power much higher than the buncher cavity, so that a large output can be taken.

The cavities are re-entrant type and are also tunable. Further more, they may be integral or demountable, in the latter case the wire grid meshes may be connected to the rings external to the glass envelope, and cavities may be attached to the rings. Although the drift space is quite long and the transit time in it is put to use, the gaps must be short that the voltage across them does not change significantly during the passage of a particular bunch of electrons, having a high collector voltage helps in this regard.

It has already been said that when electrons pass through gap A, they are influenced by the RF voltage across this gap. However, the extent of this effect on any electron will depend on the voltage across the gap, at the time the electron passes this gap. It thus becomes necessary, to investigate the effect of the gap voltage upon passing electrons individually.

6.2.1 Velocity Modulation:

Consider the situation when there is no voltage across the gap, electrons passing it are unaffected and continue on to the collector with the same constant velocities they had, before approaching the gap (fig.6.2). Sometimes later, after an input has been fed to the buncher grid, an electron will pass gap A at the time when the voltage across this gap is zero and going positive, let this be the reference electron ‘y’. This reference electron is unaffected by the gap, as evidenced by the fact that it has the same slope on the “Applegate diagram” of Fig. 6.2, as electrons passing the gap before any signal was applied to the buncher cavity.

Another electron ‘z’ passes the gap slightly later than ‘y’, as shown in fig 6.2. In the absence of gap voltage, both electrons would have continued past the gap with unchanged velocity and, therefore, neither would have caught up with the other. In presence of positive voltage across gap A, however electron ‘z’ is accelerated slightly and given enough time, will catch up with the reference electron ‘y’ easily before gap B is approached. Similarly, electron ‘x’ passes gap A slightly before the reference electron, and is retarded by the negative voltage, at that instant across the gap, since electron ‘y’ was not so retarded, it has an excellent chance of catching electron ‘x’, before gap B, and this is done as shown in Fig 6.2.

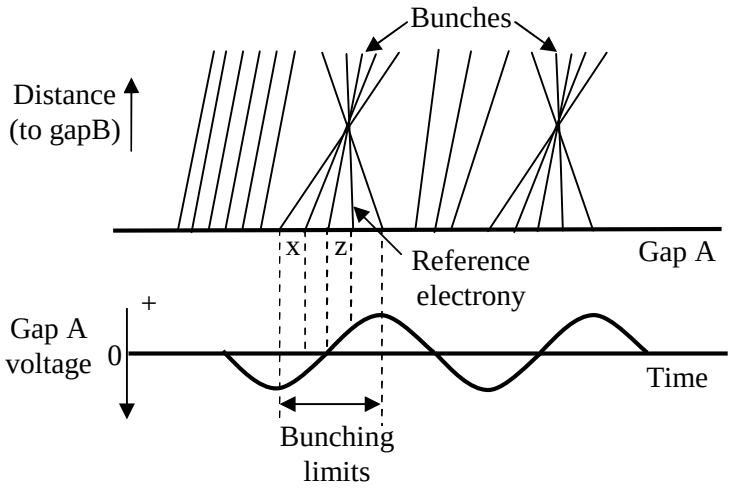


Fig 6.2 Applegate diagram for klystron amplifier

As electrons pass the buncher gap, they are velocity modulated by the RF voltage existing across this gap; such velocity modulation is not sufficient in itself for amplification, by the klystron. However, as explained with reference to the Applegate diagram, the electrons are given an opportunity to bunch in the drift space. When an electron catches up with another one, it may simply pass it and forge ahead. On the other hand, it may exchange energy with the slower electron, giving it some excess energy, and the two bunch together and move on with the average velocity of the beam. As the beam progresses further down along the drift space, the bunching becomes more complete, as more and more of the faster electrons catch up with bunches ahead. Eventually, the current passes the catcher gap with quite pronounced bunches and therefore, varies cyclically with time, and this variation in current density enables the klystron to have a significant gain.

It is noted that bunching can occur only once per cycle, centering on the reference electron. The limits of bunching are also shown, any electrons arriving slightly after the second limit are not accelerated sufficiently to catch the reference electron and that the reference electron cannot catch up with any electron passing through the gap A just before the first limit. Bunches, therefore, arrive at the catcher grid, once per cycle and then deliver energy to this cavity. The catcher cavity is excited into oscillations at its resonant frequency (which is same as the input frequency) and a large sinusoidal output can be obtained because of the flywheel effect of the output resonator. Bunching therefore is dependent upon the following parameters.

1. Drift space should be properly adjusted.
2. Signal amplitude should be such that proper bunching takes place.
3. D.C. anode voltage.

Above three factors when properly adjusted gives maximum efficiency

6.3 Reflex Klystron:

A reflex klystron is a low power, low efficiency microwave oscillator, illustrated schematically in fig.6.3. It consists of an electron gun similar to that of a multi cavity klystron, a filament surrounded by a cathode and a focusing electrode at the cathode potential. The suitably formed electron beam is accelerated towards the cavity, which has a high positive voltage applied to it and thus acts as an anode. After passing the gap in the cavity, electrons travel towards repeller electrode, which is at a high negative potential.

The electrons never reach this electrode because of the negative field.

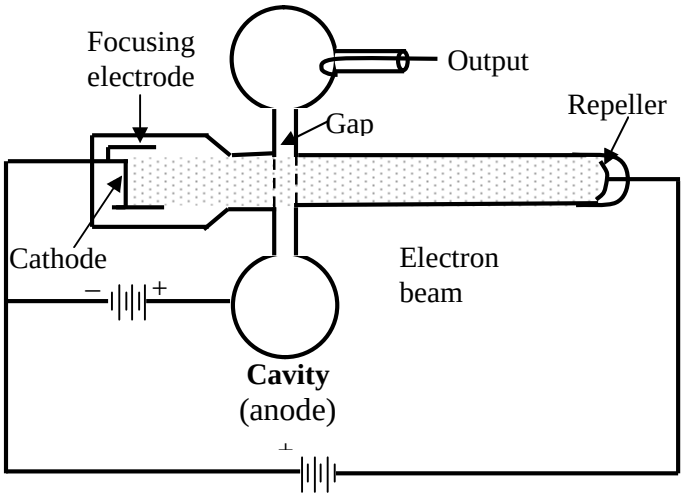


Fig 6.3 Reflex Klystron schematic.

The electrons are repelled back from midway of the repeller space by the repeller grid towards the anode, if conditions are properly adjusted, then the returning electrons give more energy to the gap than they took from it on forward journey, and thus oscillations are sustained.

6.3.1 Operation: -

It is assumed here as also is the case with two-cavity klystron oscillator those oscillations are capable of starting the through noise or switching transients. In order to understand the operation, it is advantageous to consider specific electrons, passing the gap for the first time at selected times. Here reference electron taken is 'y' and it passes the gap without being affected by the voltage across it and travels towards to repeller, and is returned back to anode without touching the repeller. The fast electrons come nearer to the repeller than slow ones, and therefore take a longer time to return to the resonator gap than electrons, which did not approach the repeller so closely.

Fig 6.4 shows the Applegate diagram for a reflex klystron. It shows the path of reference electron 'y', rather the position of 'y' at any instant in the repeller space; its path of course is straight out and then straight back along the same line. Consider an electron 'x' which passes the resonator gap (on its way out) slightly before the reference electron. Had there been no gap voltage, 'x' would have returned before the reference electron. However, due to presence of RF voltage the electrons will be velocity modulated. Electron 'x', as seen is in fact accelerated due to the positive field available to it and so comes closer to the repeller as compared to reference electron 'y'. As seen from fig 6.4, it is quite possible that 'y' will catch up with electron 'x' as they enter back into resonator gap. Similarly electron 'z' leaves the gap slightly after the reference electron 'y' and experiences a negative field, which slows it; therefore electron 'z' does not reach as close to repeller compared to electron 'y', before it is returned back to resonator gap. There exists every possibility that electron 'z' also catches up with electron 'y' as it enters back into the gap. Thus all the three electrons 'x', 'y' and 'z' return back nearly at the same time into the resonator gap.

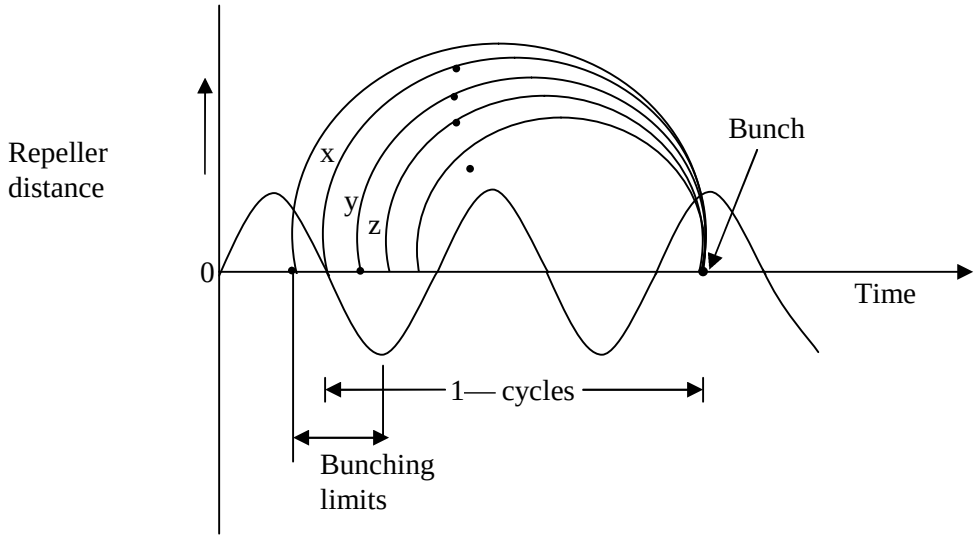


Fig 6.4 Applegate diagram for Reflex klystron in 1— cycle mode.

The system now becomes analogous to a two-cavity klystron where velocity modulation is converted into current modulation after the electrons have left the gap on their outward journey. The bunching in reflex klystron is not as complete as in a multi cavity klystron, because there are quite a few electrons arriving at the gap out of phase and contributing to the high noise and low efficiency of the device. However, bunching is sufficient between the bunching limits, to make operation possible.

It is to be noted here that one bunch forms per cycle of oscillations around the reference electron, and that these bunches deliver energy to the gap.

6.3.2 Transit time:

For oscillations to be maintained, the time taken by the electron to travel into the repeller space and back to the gap (called transit time) must have an optimum value. This factor is not so important in a klystron amplifier, but it assumes a great importance here. Here we will see what should be an optimum correct time for the oscillations to be sustained.

The most optimum departure time is obviously centered on the reference electron, which is at 180^0 point of sine wave voltage across the resonator gap. Actually no energy goes into velocity modulating of the electron beam. It takes some energy to accelerate electrons, but just as much energy is gained from retarding electrons; since there are as many retarded electrons as accelerated, the total energy outlay is nil. From the above discussion it is evident that the best possible time for electrons to return to the gap is at a time at which the voltage then existing across the gap will apply maximum retardation to them; this is when the gap voltage is positive maximum. This causes electrons to fall through the maximum negative voltage between the gap grids, thus giving up the maximum amount of energy to the gap. Thus it appears that, the best time for electrons to return to the gap is at the 90^0 point of the sine wave gap voltage. Referring to fig 6.4, returning of electrons after $1\frac{3}{4}$ cycle satisfies the above requirements; when generalized the transit time in repeller space is given by

$$T = n + \frac{3}{4}$$

Where n = any integer.

Applications: Followings are the four major fields of application of reflex klystron oscillator. Reflex klystron is used as

1) Local oscillator in microwave receiver.
2) Signal source in Microwave generators.
3) Frequency modulated oscillator in low power portable microwave links.
4) Pump oscillator for parametric amplifiers.

6.4 Magnetron:

Magnetron forms one of the various high power microwave oscillator and it forms the basis of many a microwave radar transmitter system even now a days. This microwave generator makes use of magnetic field for producing oscillations at microwave frequencies. Microwave magnetrons now find many applications apart from radar. They are capable of furnishing megawatts of peak power in the centimeter wavelength range and may be operated at wavelengths extending down to millimeter range.

All magnetrons consist of some form of anode and cathode operated in a magnetic field that is normal to the electric field between the cathode and the anode. The magnetic field causes electrons emitted from the cathode to move in curved paths. In case the flux density of the magnetic field is sufficiently great, the electrons do not strike the anode but return to the cathode and the anode current is cut off

Magnetrons form three categories namely

(a) Negative resistance magnetrons.

(b) Cyclotron frequency magnetrons.

(c) Traveling wave (or) Cavity type magnetrons.

Negative resistance Magnetrons make use of static negative resistance present between two anode plates, but have low efficiency and are useful only at low frequencies (< 500 MHz).

Cyclotron frequency magnetrons depend on synchronism between an alternating component of electric field and periodic oscillation of electrons in a direction parallel to this electric field. These are useful only for frequencies greater than 100MHz.

Cavity magnetrons depend upon the interaction of electrons with a rotating electro-magnetic field of constant angular velocity. These provide oscillations of very high peak power and hence are very useful in radar applications.

6.4.1 Cavity Magnetron:

It is a diode usually of cylindrical configuration with a thick cylindrical cathode at the center and a co-axial cylindrical block of copper as anode. In the anode block are cut a number of holes and slots, which act as resonant anode cavities. The space between the anode and cathode is the interaction space and to one of the cavities is connected a co-axial line or wave guide for extracting the output. It is a cross-field device as the electric field between anode and cathode is radial where as the magnetic field produced by a permanent magnet is axial. The permanent magnet is placed such that the magnetic lines are parallel to the vertical cathode and perpendicular to the electric field between cathode and anode. The construction is shown in fig 6.5 a and b.

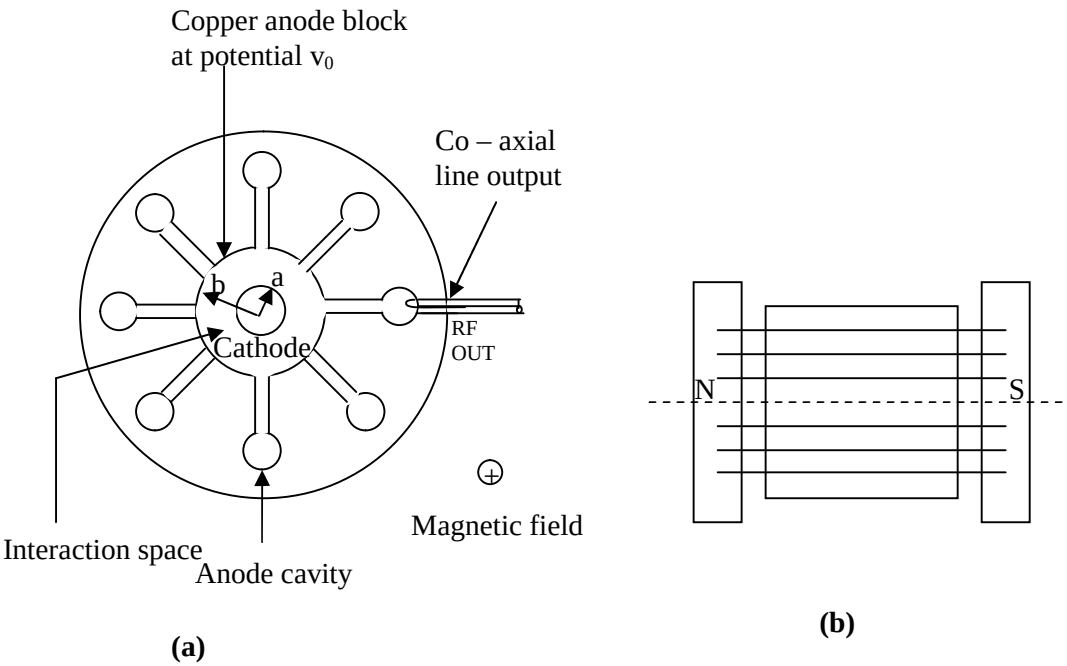


Fig 6.5 (a) Constructional detail of cavity magnetron.
(b) Magnetic flux lines in magnetron.

Operation: The cavity magnetron shown in Fig 6.5(a) has 8 cavities that are tightly coupled to each other. We know, in general that a N-cavity tightly coupled system will have N-modes of operation each of which is uniquely characterized by a combination of frequency and phase of oscillation relative to the adjacent cavity. In addition, these modes must be self-consistent so that the total phase shift around the ring of cavity resonators is $2n\pi$ where 'n' is an integer. For example, a phase shift should be 40^0 between cavities of a 8-cavity magnetron will mean that the first cavity is out of phase with itself by 320^0 The correct minimum phase shift should be 45^0($45 \times 8 = 360^0$). Therefore if ϕ represents the relative phase change of the ac electric field across adjacent cavities, then

$$\phi_v = \frac{2n\pi}{N} \text{ where } n = 0, \pm1, \pm2, \pm\left(\frac{N}{2}-1\right), \pm\frac{N}{2} \quad \text{----------} \quad (6.1)$$

i.e., $\frac{N}{2}$ mode of resonance can exist if 'N' is an even number.

$$\text{If } n = \frac{N}{2}, \phi_v = \pi$$

This mode of resonance is called the π - mode.

If n = 0, $\phi_v = 0$.

This is the zero mode, meaning there will be no RF electric field between anode and cathode (called the fringing field) and is of no use in magnetron operation.

To understand the operation of cavity magnetron, we must first look at how the electrons behave in the presence of closed electric and magnetic fields.

Depending on the relative strengths of the electric and magnetic fields the electrons emitted from the cathode and moving towards the anode will traverse through the interaction space as shown in fig 6.6.

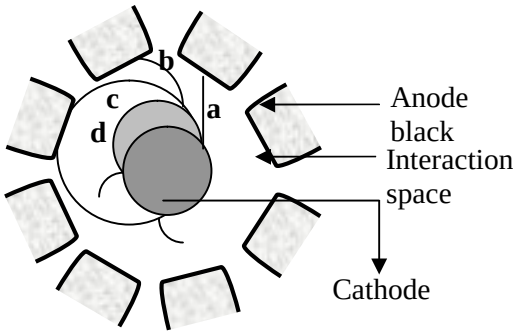


Fig 6.6 Electron trajectories in the presence of crossed electrical and magnetic fields. (a) No magnetic field (b) Small magnetic field (c) Magnetic field =B_c (d) Excessive magnetic field.

In the absence of magnetic field (B = 0), the electron travels straight from the cathode to the anode due to the radial electric field force acting on it (indicated by the trajectory 'a' in fig 6.6). if the magnetic field strength is increased slightly (i.e., for moderate value of B) it will exert a lateral force bending the path of the electron as shown by path 'b' in fig 6.6. The radius of the path is given by R = $\frac{mv}{eB}$, that varies directly with electron velocity and inversely as the magnetic field strength.

If the strength of the magnetic field is made sufficiently high so as to prevent the electrons from reaching the anode (as shown by path 'c' and those inside in Fig.6.6) the anode current becomes zero. The magnetic field required to return electrons back to cathode just grazing the surface of the anode is called the critical magnetic field (B_c) or the cut-off magnetic field. If the magnetic field is made larger than the critical field ($B>B_c$), the electron experiences a greater rotational force and may return back to cathode quite faster. All such electrons may cause back heating of the cathode. This can be avoided by switching off the heater supply after commencement of oscillation. This is done to avoid fall in the emitting efficiency of the cathode.

All the above explanation is for a static case in the absence of the RF field in the cavity of magnetron.

Assuming RF oscillations to have been initiated due to some noise transient within the magnetron, the oscillations will be sustained by device operation. As pointed out earlier self-consistent oscillations will be obtained if the phase difference between adjacent anode poles is $n\pi/4$, where n is an integer; n = 4 results in π-mode of operation which is shown in Fig.6.6. Here the anode poles are π radians apart in phase. The dotted electron paths refer to the case of static fields with no RF field. The solid paths refer to the electron trajectories in the presence of RF oscillations in the interaction space. The electron 'a' is slowed down in presence of oscillations thus transferring energy to the oscillations, during its longer journey from cathode to anode. Such electrons, which participate in transferring energy to the RF field are called favored electrons and are responsible for bunching-effect. An electron 'b' is accelerated by the RF field and instead of imparting energy to the oscillations takes energy from oscillations resulting in increased velocity. Hence bends more sharply, spends very little time in the interaction space and is returned back to the cathode. Such electrons are called unfavoured electrons which do not participate in the bunching process rather they are harmful in the sense they cause back heating. Similarly an electron 'c' which is emitted a little later to be in correct position moves faster and tries to catch up with electron 'a' and an electron emitted at 'd' will be slowed down to fall back in step with electron 'a'. This results in all favored electrons like a, c, d to form a bunch and are confined to spokes or electron clouds. The process is called the phase focusing effect corresponding to a bunch of favored electrons around the reference electrons 'a '. The spokes so formed in the π-mode rotate with an angular velocity corresponding to two poles per cycle.

The phase focusing effect of these favored electrons imparts enough energy to the RF oscillations so that they are sustained.

6.5 Magic Tee:

In some applications of wave guide systems need arises for a hybrid circuit, shown in general form in fig 6.7. The characteristics of a hybrid circuit are such that if power enters the circuit through arm A or arm C, the power is delivered entirely to arms B and D, with no direct transmission from A to C or C to A. It may also be desirable that power entering through arm B or arm D, is delivered entirely to arms A and C with no direct transmission between B and D. The most commonly used type of hybrid wave guide junction is the magic tee, shown in fig 6.7(a), which is a combination of an E-plane tee and an H-plane tee. The characteristics of series and shunt tees are such that, when waves of equal amplitude and phase enter the P and S arms, the E fields cancel in one of the side arms and add in the other. By the reciprocity principle, therefore, energy applied to arm 1 or 2 is divided equally between the P and S arms, none emerging from the opposite side arm. If arms 1 and 2 are terminated in matched loads and no reflections take place within the Junction, the entrance of power through either arm S or arm P results in equal power delivery to arms 1 and 2.

Reflections take place because of severe discontinuities within the Junctions and because at the junction each arm is effectively terminated by two other arms, of equal impedance, in series of parallel. Reflections must be prevented by the use of reactive tuning elements such as screws or diaphragms.

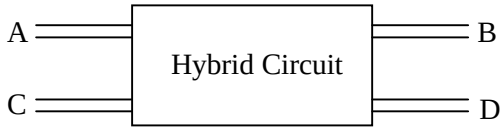


Fig 6.7 General form of a hybrid circuit.

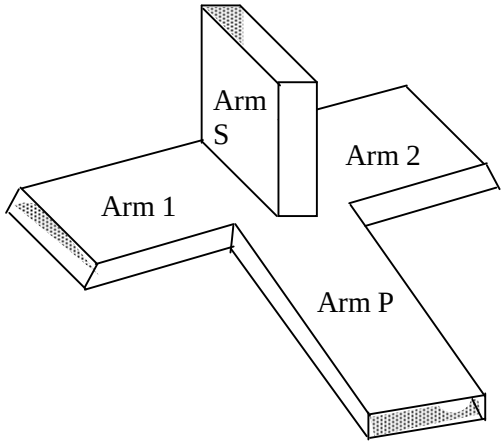


Fig 6.7(a) One form of a microwave magic tee.

6.5.1 Applications of Magic Tee:

A magic Tee has several applications. A few of them have been discussed here.

(a) Measurement of Impedance: A magic Tee has been used in the form of a bridge, as shown in Fig.6.8 for measuring impedance.

Microwave source is connected in arm 3 and a null detector in arm 4. The unknown impedance is connected in arm 2 and a standard variable known impedance in arm 1. Using the properties of Magic Tee, the power from microwave source a_3 gets divided equally between arms 1 and 2 by an amount $\left(\frac{a_3}{\sqrt{2}}\right)$(to the unknown impedance and standard variable impedances). These impedances are not equal to characteristic impedance Z_0 and hence there will be reflections from arms 1 and 2. If ρ_1 and ρ_2 are the reflection

coefficients, powers $\left(\frac{\rho_1 a_3}{\sqrt{2}}\right)$ and $\left(\frac{\rho_2 a_3}{\sqrt{2}}\right)$ enter the Magic Tee junction from arms 1 and 2 as shown in fig.6.8.

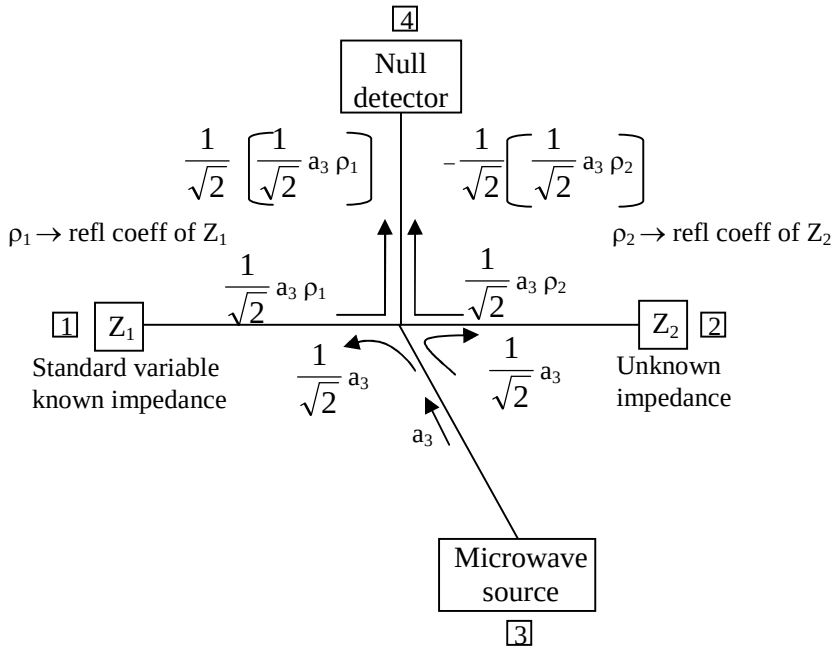


Fig 6.8 Magic Tee for measurement of impedances.

The resultant wave into arm 4 i.e., the null detector can be calculated as follows:-

The net wave reaching the null detector (Refer Fig 6.8)

$$\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}} a_3 \rho_1\right) - \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}} a_3 \rho_2\right) = \frac{1}{2} a_3 (\rho_1 - \rho_2) \quad \text{---------- 6.2}$$

For perfect balancing of the bridge (null detection) Eq.6.2 is equated to zero.

i.e., $\frac{1}{2} a_3(\rho_1 - \rho_2) = 0$

or $\rho_1 - \rho_2 = 0$ or $\rho_1 = \rho_2$

or $\frac{Z_1 - Z_z}{Z_1 + Z_z} = \frac{Z_2 - Z_z}{Z_2 + Z_z}$

$\therefore$ $Z_1 = Z_2$

i.e., $R_1 + jX_1 = R_2 + jX_2$

or $R_1 = R_2$ and $X_1 = X_2$

Thus the unknown impedance can be measured by adjusting the standard variable impedance till the bridge is balanced and both impedances become equal.

(b) Magic Tee as a Duplexer: The transmitter and receiver are connected in ports (2) and (1) respectively, antenna in the E-arm or port (4), and port (3) of Magic Tee is terminated in a matched load as shown in Fig 6.9. During transmission half the power reaches the antenna from where it is radiated into space. Other half reaches the matched load where it is absorbed without reflections. No transmitter power reaches the receiver since port (1) and (2) are isolated ports in Magic Tee. During reception, half of the received power goes to the receiver and the other half to the transmitter are isolated during reception as well as during transmission.

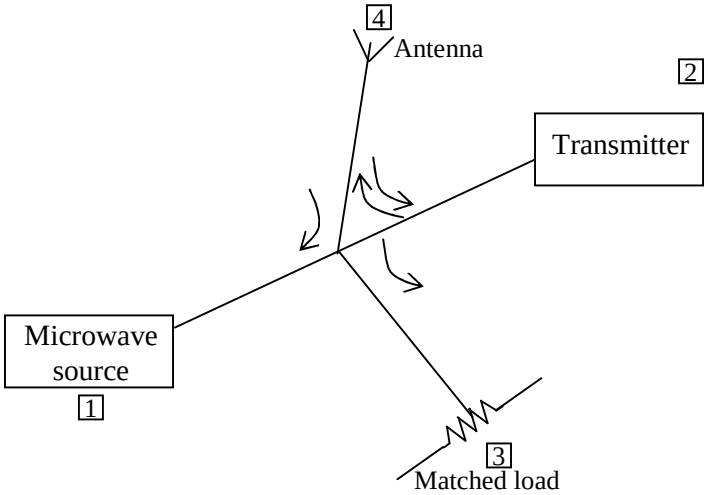


Fig 6.9 Magic Tee as a duplexer

(a) **Magic Tee as a Mixer:** A Magic Tee can also be used in microwave receivers as a mixer where the signal and local oscillator are fed into the E and H arms as shown in Fig. 6.10.

Half of the local oscillator power and half of the received power from antenna goes to the mixer where they are mixed to generate the IF frequency.

$$IF = f_{in} \sim f_0$$

Magic tee has many other applications such as a microwave discriminator, microwave bridge etc.,

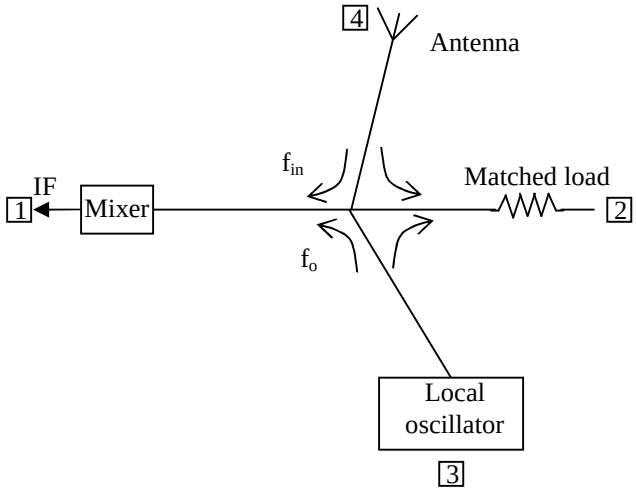
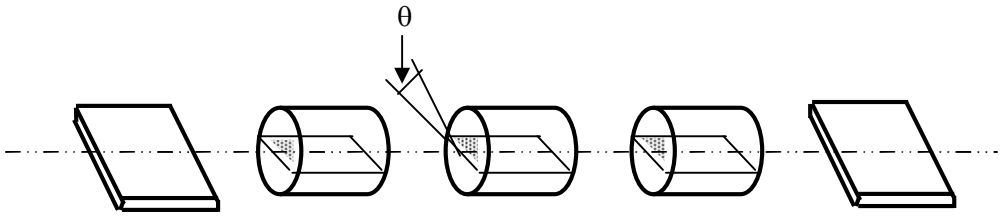


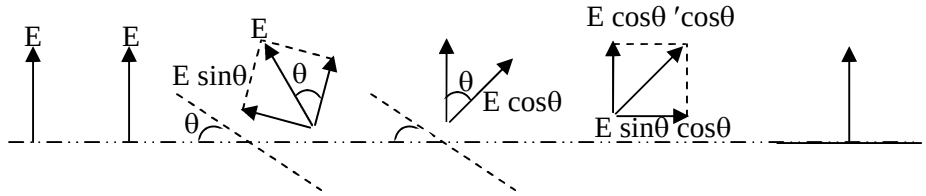
Fig 6.10 Magic Tee as a mixer

6.6 Attenuators:

There are two classes of attenuators. The one that provides a fixed amount of attenuation is called the fixed type of attenuator, and the one that provides a variable amount of attenuation is called the variable type attenuator. The latter type is most commonly used in Bridge set ups used to measure transmission coefficients. There are a number of ways by which a variable attenuator can be realized of which one type, the rotary attenuator, is considered in detail.



(a)



(b)

Fig 6.11 Functional diagram including operating principle of the Rotary Vane Attenuator.

The rotary vane attenuator is a direct reading precision attenuator and obeys a simple mathematical law. A functional diagram indicating the operating principle of this type of attenuator is shown in fig 6.11 (a) and (b).

Basically, it consists of three sections of Wave Guide in tandem as shown in Fig.6.12(a). A rectangular to circular wave guide transition containing a horizontal attenuator strip is connected to a rotatable circular wave guide containing an attenuator strip and this in turn is connected to a circular to rectangular wave guide transition containing a horizontal attenuator strip. The attenuator strip for the Rotary Vane attenuator is shown in Fig.6.12(b).

The incoming TE_{10} mode is transformed into TE_{11} mode in the circular wave guide by the rectangular to circular wave guide transition with negligible reflections. The polarization of the

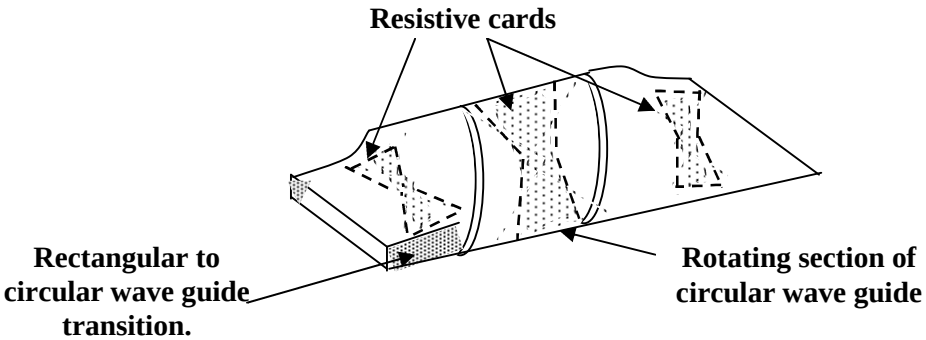


Fig 6.12(a) Basic construction of a rotary attenuator.

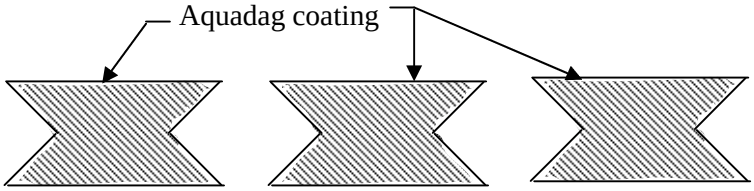


Fig 6.12(b) The attenuator strips for the Rotary Vane Attenuator.

TE_{11} mode is such that the electric field is perpendicular to the thin resistive card in the transition section. As such, this resistive card has a negligible effect on the TE_{11} mode. Since the resistive card in the center section can be rotated, its orientation relative to the electric field of the incoming TE_{11} mode can be varied so that the amount by which this mode is attenuated is adjustable.

With reference to fig 6.12(b), when all strips are aligned, the electric field of the applied wave is normal to the strips and hence no current flows in the attenuation strips and therefore no attenuation occurs. In a position where the central attenuation strip is rotated by angle θ, the electric field of the applied wave can be resolved into two orthogonally polarized modes; one perpendicular and one parallel to the resistive card. That

portion which is parallel to the resistive slab will be absorbed, whereas the portion which is polarized perpendicular to the slab, will be transmitted. It can be proved mathematically, by considering the analytic expression for the TE_{11} mode electric field that the attenuation produced is given in decibels by

$$\alpha = - 20 \log (\sin^2 \theta) = - 40 \log (\sin \theta).$$

6.7Microwave Resonators:

Resonators are an essential part of most microwave circuits. A knowledge of the structure, characteristics, and design of microwave resonators is therefore necessary to a clear understanding of microwave amplifiers and oscillators. Because of the difficulty of predicting in advance, the shunt the reactance of tubes and their associated mountings as they are to be used in particular resonators, the equations and curves presented in this chapter cannot be used to predetermine with exactness the characteristics of a resonator of known dimensions incorporating a tube.

Commonly used microwave resonators are of the butterfly, parallel-line, coaxial line, waveguide, and cavity types.

Parallel-line and Coaxial-line Resonators:

A type of resonator that may be used to advantage in certain types of microwave oscillators at frequencies up to approximately 3000 MHz consists of a section of parallel-wire transmission line. Most commonly, one end of the line terminates in the electrodes and the other end is short-circuited. Such a line may be considered as the limiting configuration of a lumped-parameter resonator obtained when the inductance and capacitance are reduced until the inductance consists of a single turn of wire and the only lumped capacitance is that of the tube electrodes. Another frequently used parallel-wire resonator is one that is short-circuited at both ends, the tube electrodes being connected to any point of the line at which the alternating voltage is high at the resonance frequency, preferably the midpoint of the line. This type of resonator has advantage that, the upper limit of the frequency range can be made high by terminating one end of the line within the tube envelope, close to the electrodes. Parallel-wire resonators are ordinarily tuned by means of one or more movable short-circuiting bridges or by means of lumped resistance shunting the resonator at a point where the voltage is high.

The principal advantages of parallel-wire resonators in their application to amplifiers or oscillators are their simplicity and their adaptability to use with tubes in which the electrode leads have the form of wires or pins, such as split-anode magnetrons and some types of conventional triodes. Because of relatively high losses, however, particularly by radiation, parallel-wire resonators are not ordinarily used at frequencies above a few hundred mega cycles. Although they may be tuned by means of a wire bridging the line, such an arrangement is not very satisfactory, since much of the electromagnetic energy passes the bridge. A second bridge, a

quarter-wavelength beyond the first, greatly improves the effectiveness of the quasi-short-circuit. A better type of short-circuiting bridge is a conducting plate of radius several times the conductor spacing and of thickness greater than the skin depth, having two holes through which the conductors pass.

Space, requirements have been reduced in a modified form of parallel-wire resonator in which the wires forming the line are coiled to form spirals or double helices. Such structures also simplify the mechanical problems of tuning, since the short-circuiting bridge can be mounted on an arm and caused to move along the line by rotating the arm or the spirals or helices.

Although the characteristics of coaxial-line resonators are in many respects similar to those of parallel-wire resonators, coaxial-line resonators have the advantage over parallel-line resonators that the electromagnetic field may be entirely confined within the outer conductor. Radiation losses and undesired coupling to other circuit elements are therefore avoided, and the resonant Q is higher than in parallel-line resonators. The frequency of resonance can be changed by the use of one or more movable short-circuiting plungers to changing the resonator length or by means of a lumped capacitance shunting the resonator at a point where the electric field is high.

The absence of a lower limit to the frequency at which propagation is obtained in the TEM mode makes possible a wide tuning range in coaxial and parallel-line resonators without interference from resonances associated with higher-order modes of propagation. Waveguide and cavity resonators do not share this advantage.

6.8 Summary:

- A Klystron is a vacuum tube that can be used either as a generator or as an amplifier of power at microwave frequencies.
- The Reflex Klystron is a single cavity variable frequency microwave generator of low power and low efficiency.
- Reflex Klystron is most widely used in applications where variable frequency is derived as in radar receivers, local oscillator in microwave receivers and pump oscillator in parametric amplifier.
- Magnetron form one of the various high power microwave oscillators and it forms the basic of many a microwave radar transmitter systems.
- Magnetrons form three categories namely (i) Negative resistance magnetron (ii) Cyclotron frequency magnetron (iii) Traveling wave magnetron.
- Resonators are an essential part of most microwave circuits.
- A knowledge of the structure, characteristics, and design of microwave resonators is therefore necessary to a clear understanding of microwave amplifier and oscillators.

- Commonly used microwave resonators are of the butterfly, parallel-line, coaxial line, wave guide and cavity types.
- A T–junction is an intersection of three wave guides in the form of English alphabet 'T'.
- There are several types of Tee Junctions. (i) H–plane Tee Junction (ii) E–Plane Tee Junction. (iii) E–H plane Tee Junction (iv) Magic–T Junction (v) Rate race Junction.
- Attenuators are commonly used for measuring power gain or loss in dBS, for reducing the power input to a particular stage to prevent over loading and also for providing the signal generators with a means of calibrating their outputs accurately so that precise measurement could be made.
- Attenuators can be classified as fixed or variable types.

6.9 Key Terminology:

Attenuator (flap): A device designed to introduce attenuation into a wave guide circuit by means of a resistive material moved into the guide.

Attenuator (rotary vane): A device designed to introduce attenuation into a wave guide circuit by means of varying the angular position of a resistive material in the guide.

Magnetron: A high – power microwave oscillator tube with a fixed or limited frequency range. Frequency, and power depend on magnetic field strength and anode voltage.

Reflex Klystron: A low-power microwave oscillator tube which depends primarily on the physical size of a cavity resonator for its frequency. Normally has a wider frequency range than a magnetron.

Noise figure: A figure of merit for microwave amplifiers. A ratio in db between actual output noise power and the output noise power which would come from a noiseless amplifier with identical gain and bandwidth.

6.10 Self-Assessment Questions:

1. Discuss the theory of klystron amplifier.
2. By means of an Applegate diagram explain the operation of a reflex klystron.
3. Give a brief note on (i) attenuators and (ii) Magic Tee
4. Discuss the applications of Magic Tee.
5. Discuss the performance of Magnetrons and list the important applications.

6. Give a brief note on Microwave Resonators.
7. Discuss the working of Coaxial line cavity Resonator.

6.11 References:

1. Microwave circuits and passive devices – M L Sisodia & G.S. Raghuvamshe
   New Age International Publishers Ltd.
2. Microwave and Radar Engineering – M. Kulkarni – Umesh Publications.
3. Microwave devices and circuits – Samuel Y. Liao – Prentice – Hall of India Pvt.Ltd.
4. Microwave Propagation and Techniques – D.C. Sarkar – S. Chand & Company Ltd.

Name of the Author:

S.V. RangaNayakulu, M.Sc., M.Phil., M.Ed.,
Dept of Physics & Electronics
T.J.P.S. COLLEGE (P.G.COURSES)
GÜNTUR-522 006

UNIT II:

LESSON 7:

Modulation

Objectives:

In this chapter, we study about the

(i) Processes of amplitude and frequency modulation

(ii) Types of amplitude modulation

(iii) Process of detection of AM and FM waves

Structure:

Section I

7.1 Introduction.

7.2 Need for modulation.

7.3 Amplitude modulation.

7.4 Mathematical analysis of an A.M. wave.

7.5 Generation of A.M. waves

7.6 Demodulation of A.M. waves.

7.7 Super heterodyne receiver.

Section II

7.8 Frequency modulation.

7.9 Analysis of F.M. waves

7.10 Production of F.M. waves.

7.12 F.M. Detection.

7.13 Foster-Seeley discriminator.

Section I

Amplitude modulation and A.M. detection

7.1 Introduction:

For successful transmission and reception of intelligence (code, voice, music etc) by the use of radio waves, two processes are essential. They are (i) Modulation (ii) De-Modulation or detection. To modulate means to regulate or adjust. Speech and music etc. are sent thousands of kilometers away by a radio transmitter. Similarly, a scene in front of a television camera is also sent many kilometers away to viewers. In such cases the carrier is the high frequency radio wave. The intelligence i.e. video, sound and other data is impressed on the carrier wave and is carried along with it to the destination.

Modulation is the process of combining the low frequency signal with a very high frequency radio wave called carrier wave (c.w.). The resultant wave is called modulated carrier wave. This job is done at the transmitting station.

During modulation some characteristic of the carrier wave is varied in time with the modulating signal. Accordingly there are three types of sine wave modulations known as amplitude Modulation, Frequency Modulation and Phase Modulation.

The A.M. wave contains three frequency components, a carrier frequency (f_c), one component above the carrier frequency ($f_c + f_s$) and the other component 'below' the carrier frequency ($f_c - f_s$) where 'f_c' and 'f_s' are the carrier and signal frequencies respectively. At the receiver the signal is extracted from the modulated carrier. This process is called A.M. detection or de-modulation.

7.2 Need for Modulation:

Sometimes, beginners question the necessity of modulation is using a carrier wave to carry the low frequency signal from one place to another. Why not transmit the signals directly and save lot of difficulty? Unfortunately there are the many hurdles in the process of such direct transmission of audio frequency signals. These are

(i) Voice, data and video signals have low frequencies. These waves have relatively short range.

(ii) If everybody transmits these low frequency signals directly, mutual interference will render all of them ineffective.

(iii) Size of antennas required for their efficient radiation would be large, i.e., about 75Km!

Hence, the solution lies in modulation, which enables a low frequency signal to travel very large distances through space with the help of a high frequency carrier wave (f_c). These carrier waves need reasonably sized antennas and care is taken to select the carrier frequencies to avoid interference with other transmitters operating in the same area.

7.3 Amplitude Modulation:

Definition: When the amplitude of high frequency carrier wave is changed in accordance with the instantaneous value of the amplitude of the signal, it is called amplitude modulation. However the frequency of the modulated wave remains the same i.e at carrier frequency. The following fig (1) shows the principle of amplitude modulation. Fig 1(a) shows the audio electrical signal where as fig 1(b) shows a carrier wave.
Fig 1(c) shows the amplitude modulated (AM) wave.

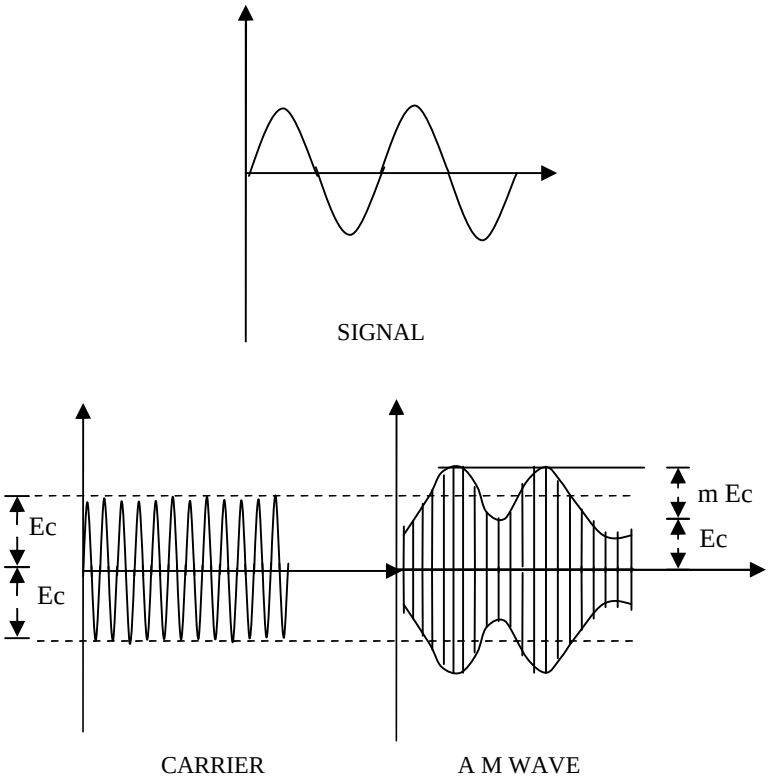


Fig 7.1.1 Amplitude Modulation

Note that the amplitudes of both positive and negative half cycles of carrier wave are changed in accordance with the signal. Amplitude modulation is done by, an electronic circuit called modulator.

Modulation factor:

An important consideration in A.M. is to describe the depth of modulation, i.e the extent to which the Amplitude of carrier wave is changed by the signal. This is described by a factor called modulation index, which may be defined as under.

Definition of Modulation index / factor:

The ratio of change of carrier wave amplitude due to modulation to the amplitude of un- modulated carrier wave is called the modulation factor m.

7.4 Mathematical analysis of Wave:

A carrier wave may be represented by $e_c = E_c \cos \omega_c t$, where

e_c = instantaneous voltage of carrier

E_c = amplitude of carrier

$\omega_c = 2\Pi f_c$ = angular velocity of carrier of frequency 'f_c'.

The modulating signal can be represented by

$e_s = Es \cos \omega_s t$

where e_s is the instantaneous value of modulating signal. Es is its amplitude.

$\omega_s = 2\Pi f_s$ = angular frequency 'f_s'.

In amplitude modulation the amplitude E_c of the carrier wave is varied in accordance with the instantaneous value of the signal as shown in fig (7.1.1).

the amplitude of modulated signal can be represented by

$E = Ec + k_a E_m \cos \omega_s t$

Where k_a is called coefficient of modulation

The carrier wave amplitude is varied at signal frequency f_s. Therefore

$$e = E_c . (1 + m \cos \omega_s t) \cos \omega_c t.$$

Where $m = k_a Es/Ec$ is called modulation index or modulation factor

The expression for A.M. wave can be expanded as

$$= E_c . \cos \omega_c t + m . E_c . \cos \omega_s t . \cos \omega_c t.$$

$$= E_c . \cos \omega_c t + m . \frac{E_c}{2} . \{2 \cos \omega_s t . \cos \omega_c t\}$$

$$= E_c . \cos \omega_c t + m . \frac{E_c}{2} \{\cos (\omega_c+\omega_s)t + \cos (\omega_c-\omega_s)t\}$$

$$= E_c . \cos \omega_c t + . \frac{mE_c}{2} . \cos (\omega_c+\omega_s).t + \frac{mE_c}{2} . \cos (\omega_c -\omega_s) t.$$

The following points may be noted from the equation of Amplitude modulated wave.

The A.M. wave is equivalent to the summation of three sinusoidal waves: one having amplitude E_c and frequency 'f_c', the second having amplitude $\frac{mE_c}{2}$ and frequency (f_c + f_s) and the third having amplitude $\frac{mE_c}{2}$ and frequency (f_c - f_s).

The A.M. wave contains three frequencies. They are f_c, f_c + f_s and (f_c-f_s). Thus, the process of amplitude modulation does not change the original carrier frequency but produces two new frequencies (f_c + f_s) and (f_c - f_s), which are called side band frequencies. The sum of carrier frequency and signal frequency i.e.,(f_c +f_s) is called upper side band frequency. The lower side band frequency is (f_c - f_s) i.e., the difference between carrier and signal frequencies. The amplitudes of the side bands are equal and proportional to depth of modulation. It can be shown that the maximum power in each side band occurs when m = 1 and is equal to one fourth of carrier power .

Upper and Lower Sidebands:

In the above discussion, it was assumed that the modulating signal was composed of one frequency component only. However, in a broadcasting station, the modulating signal is the

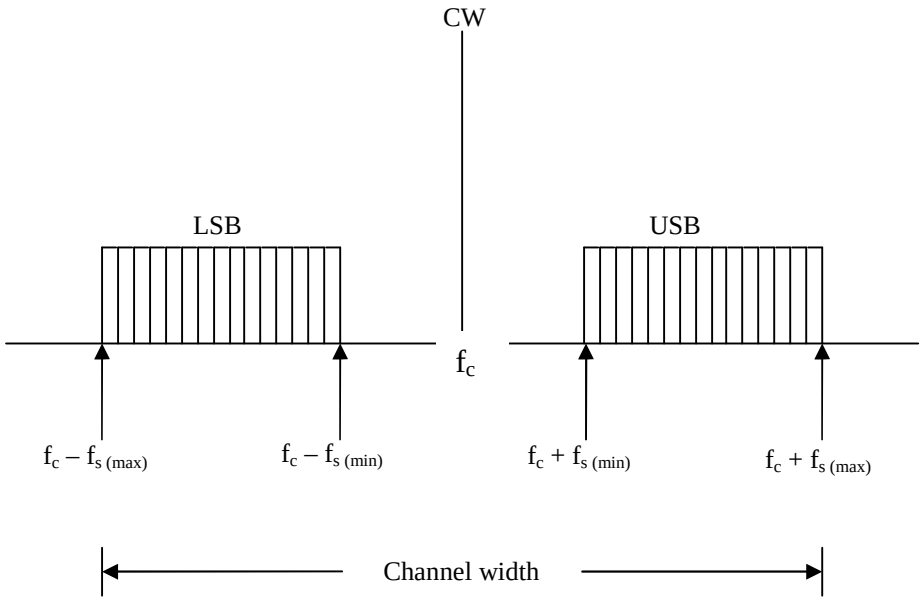


Fig 7.1.2

human voice (or music), which contains waves with a frequency range of 20-4000 Hz. Each of these waves have its own LSF and USF. When combined together, they give rise to an Upper-side band (USB) and a Lower-side band (LSB) as Shown in fig.7.1.2

The USB, in fact, contains all sum components of the signal and carrier frequency whereas LSB contains their difference components. The channel width (or bandwidth) is given by the difference between extreme frequencies i.e., between maximum frequency of USB and minimum frequency of LSB. As seen

Channel width = 2 x maximum frequency of modulating signal = 2 x f_s(max).

Example: An audio signal given by 15 sin 2Π (2000 t) amplitude-modulates a sinusoidal carrier wave 60 sin 2Π (100,000) t.. Assuming ka=1 determine a) modulation index, (b) percent modulation, (c) frequencies of signal and carrier, (d) frequency spectrum of the modulated wave.

Solution. Here carrier amplitude, A = 60 and modulating signal amplitude B = 15. Therefore

a) Modulation index, $m = \frac{B}{A} = \frac{15}{60} = 0.25$

b) b) Percent modulation, M = m x 100 = 0.25 x 100 = 25%

c) f_s = 2000 Hz --------by inspection of the given equation

f_c = 100,000 Hz --------by inspection of the given equation

d) The three frequencies present in the modulated carrier wave are

(i) 100,000 Hz = 100 kHz,(ii) 120,000 or 120 kHz (iii) 80,000 or 80kHz

Experimentally by measuring the maximum and minimum value of the modulated carrier amplitude one can determine the depth of modulation from the relation

$m = (E_{max} - E_{min}) / (E_{max} + E_{min})$

From this relation we see that , when $E_{max} = E_{min}$, m = 0 or there is no modulation and when $E_{min} = 0$ there is 100 % modulation. In commercial radio broadcasting the depth of modulation is maintained around 40%.

7.5 Generation of A.M. waves:

Amplitude modulation is produced by combining the carrier and the signal frequencies using a non-linear device. Diodes are non-linear devices but they are not used as they do not offer any gain. Transistors behave as non-linear elements and offer gain as such they are suitable for this applications. Fig 7.1.3 shows a simple amplitude – modulated amplifier.

The supply V_{CC} in combination with the resistors R_1, R_2, R_C & R_E sets the quiescent point for the transistor. The carrier e_C is the input to the CE amplifier. The circuit amplifies the carrier by a factor A_V where A_V is the voltage gain, so that the amplifier output is A_V e_C. The modulating signal e_m is applied in emitter circuit and hence forms a part of the biasing. It produces variations in emitter current at modulating frequency, which in turn produces variations in base emitter resistance and gain A_V. Thus, the amplitude of the carrier varies in accordance with the strength of the signal there by producing Amplitude modulated output across 'R_L'.

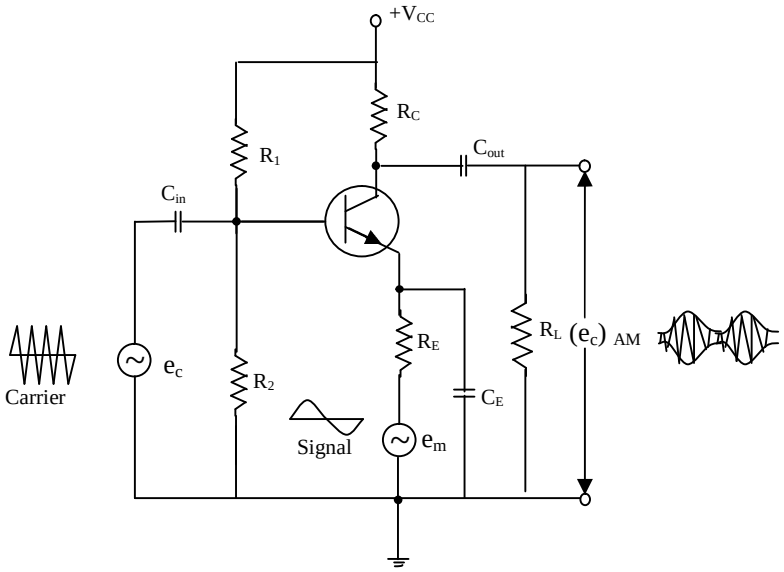


Fig 7.1.3 An amplitude-modulated amplifier.

Forms of amplitude modulation:

We know now that one carrier and two side bands are produced in A.M. generation. It is not necessary to transmit all these signals to enable the receiver to reconstruct the original signal. Accordingly, we may attenuate or altogether remove the carrier or any one of the side bands with out affecting the communication process. The advantages are,

(1) Less transmitted power and

(2) Small bandwidth required.

The different suppressed component systems are: (a) DSB-SC (b) SSB-TC (C) SSB-SC,

(a)DSB-SC: It stands for double side band suppressed carrier system. Here carrier component is suppressed there by saving enormous amount of power.

(b)SSB-TC: It stands for single side band transmitted carrier system. In this case one sideband is suppressed but the other sideband and carrier are transmitted.

(c) SSB-SC: It stands for single side band suppressed carrier system. It suppresses one side band and the carrier and transmits only the remaining side bond.

Out of the above three systems, the SSB-SC (or simply SSB) is more preferable because of its advantage over other.

Generation of SSB-SC or SSB systems:

The SSB-SC or simply SSB is generated mainly through the following three methods:

(1) Filter method.

(2) Phase cancellation (shift) method.

(3) The third method.

1. Filter method: In this method, DSB-SC signal is generated by, the balanced modulator. This signal is allowed to pass through side band filters, which are narrow band pass filters that only allow to pass the desired side band of frequencies. The filter may be a LC, mechanical or a crystal filter. All these filters have the disadvantage that their operating frequency is below the usual transmitting frequencies. Thus a balanced mixer and crystal oscillator are used to provide up conversion to the final transmitter frequency.

2. Phase shift method: In this method, two balanced modulators and two phase shifting networks are used. One of the balanced modulator receives the carrier voltage shifted by 90^0 and the modulating voltage signal while the other balanced modulator receives the carrier voltage and the modulating voltage shifted by 90^0. The output of both modulators consists of sidebands only. The signal is applied directly to the modulator M_1, therefore the modulator puts out two side bands, each one is shifted in phase by 90^0. The signal is shifted 90^0 before it is applied to the modulator, therefore, the modulator also puts two side bands, but in this case the upper and lower side bands and shifted by $+90^0$ and -90^0 respectively. The two lower side bands, which are out of phase, when combined in the adder cancel each other. The upper side hands, which are in phase add directly in the adder. Thus SSB in which lower sideband is removed is produced. The block diagram of this method is shown in the fig.

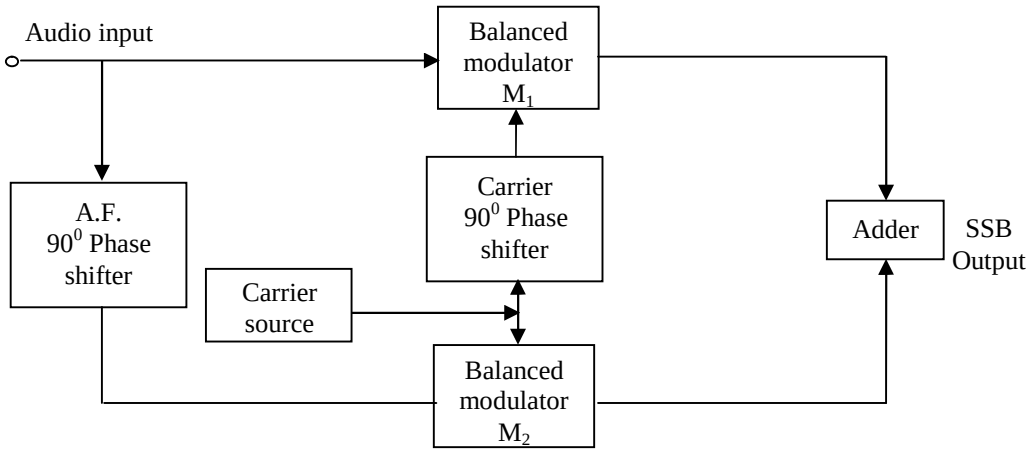


Fig 7.1.4 SSB – SC transmitter using phase shift.

The output of first balanced modulator M_1

$v_1 = 2\ v_c .\ m_a \cos \omega_c t\ .\ \sin \omega_m\ t\ .$ ---------- (1)

The output from the balanced modulator M_2

$v_2 = 2\ v_c\ m_a$ co sin $\omega_c t\ .\ \cos \omega_m\ .\ t$ ----------(2)

The summing amplifier, the resultant output

$v = v_1 + v_2 = 2\, v_c \,.\, m_a \sin(\omega_c + \omega_m)t$ ---------- (3)

(By adding (1) & (2))

Thus the output contains only upper sideband, similarly, the output with lower side band can be obtained by passing the signal carrier directly to balanced modulator M_1 and each through 90^0 phase shifting networks to the modulator M_2.

(iii) The third method:-

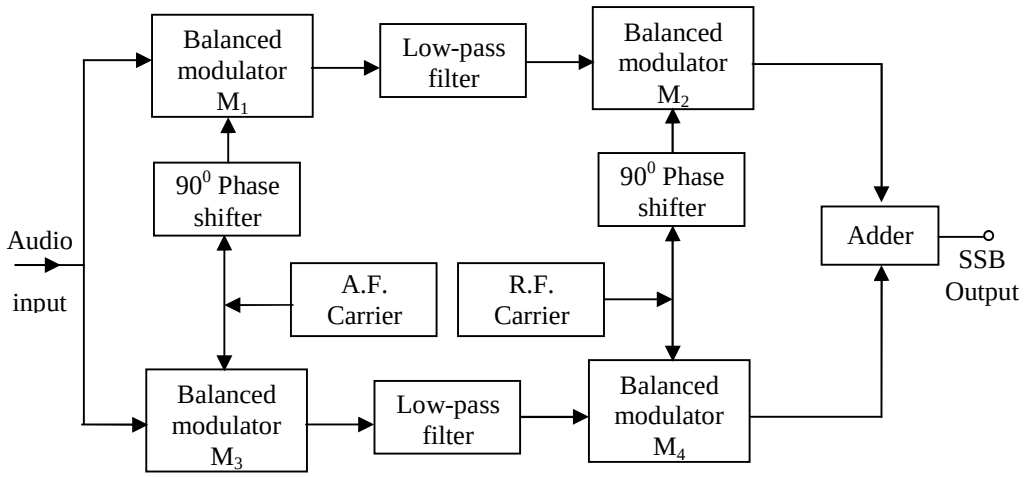


Fig 7.1.5 Third method for SSB generation.

This method has the ability to generate SSB at any frequency and the low audio frequencies. It neither refuses a filter circuit nor a wide band, audio phase shift network. The circuit is identical to that of the phase shift method, but the way in which voltages are fed to the two balanced modulators has been changed. Due to this reason, this method is called modified phase shift method the block diagram of the method is shown in the following fig.7.1.4.

7.6 Demodulation of A.M. Waves:

Definition of Demodulation:

The process of recovering the audio signal from the modulated wave is known as demodulation of detection.

At the broadcasting station, modulation is done to transmit the audio signal over larger distances to a receiver. When the modulated wave is picked up by the radio receiver, it is necessary to recover the audio signal from it. This process is accomplished in the radio receiver and is called demodulation.

Necessity of Demodulation:

We know, that an A.M. wave consists of carrier and side band frequencies. The audio signal is in side band frequencies, which are radio frequencies. If the modulated wave after amplification is directly fed to the

speaker, no sound will be heard because of the inertia of the diaphragm as it is not able to respond to such high frequencies.

Essentials of demodulation:

For a modulated wave to be audible, it is necessary to change the nature of modulated wave. This is done by a circuit called detector. A detector circuit performs the following two functions

It rectifies the modulated wave.

It separates the audio signal from the carrier.

The demodulation can be accomplished with the help of a diode detector.

A.M. Diode detector:

Diode detection is also known as envelope detection or linear detection.

In appearance it looks like an ordinary half wave rectifier circuit with capacitor input. As shown in the below fig--

It is called envelope detection because it recovers the A.F. signal envelope from the composite signal. Similarly, diode detector is called linear detector because its output is proportional to the voltage of the input signal.

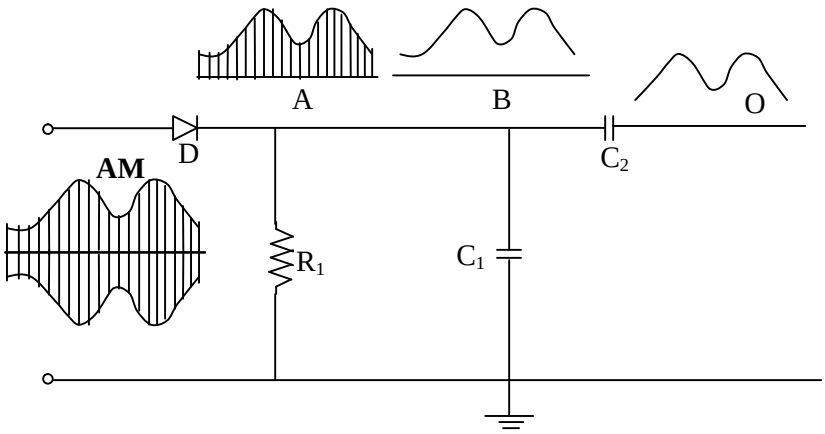


Fig 7.1.6 Simple detector
AM – Amplitude Modulated Wave

Circuit Action:

This circuit involves the rectification of the signal, and filtering the r.f. and dc components. The signal is rectified by diode 'D'. The rectified wave is shown at A in the fig. The combination $C_1 - R_1$ removes the high frequency component and provide the signal super posed over a dc component. The dc component is eliminated by capacitor 'C_2' to give the output shown in the figure 7.1.5

7.7 Super heterodyne receiver:

The shortcomings of a radio receiver were overcome by the invention of super heterodyne receiver.

Principle of super - heterodyning: In this type of radio receiver the selected radio frequency is converted to a fixed lower value called intermediate frequency (IF). This is achieved by a special electronic circuit called mixer circuit. There is a local oscillator in the radio receiver itself. This oscillator produces high frequency waves. The selected radio frequency is mixed with the high frequency wave by the mixer circuit. In this process, beats are produced and the mixer produces a frequency equal to the difference between local oscillator and radio wave frequency. The circuit is so designed that oscillator always produces a frequency 455KHz above the selected & R.F. Therefore the mixer always produces an IF of 455KHz regardless of the station to which the receiver is tuned. The production of mixed IF is the salient feature of a super heterodyne circuit. At this fixed IF, the amplifier circuit operates with a maximum stability, selectivity and sensitivity. As the conversion of incoming RF to IF is achieved by heterodyning or beating the local oscillator against R.F., therefore this circuit is called super heterodyne circuit.

Stages of super heterodyne receiver

The following fig 7.1.6 shows the block diagram of a super hetero dyne receiver. It consists of 5 stages. There are 1) RF amplifier stage 2) Mixer stage 3) IF amplifier stage 4) Detector stage 5)AF amplifier stage

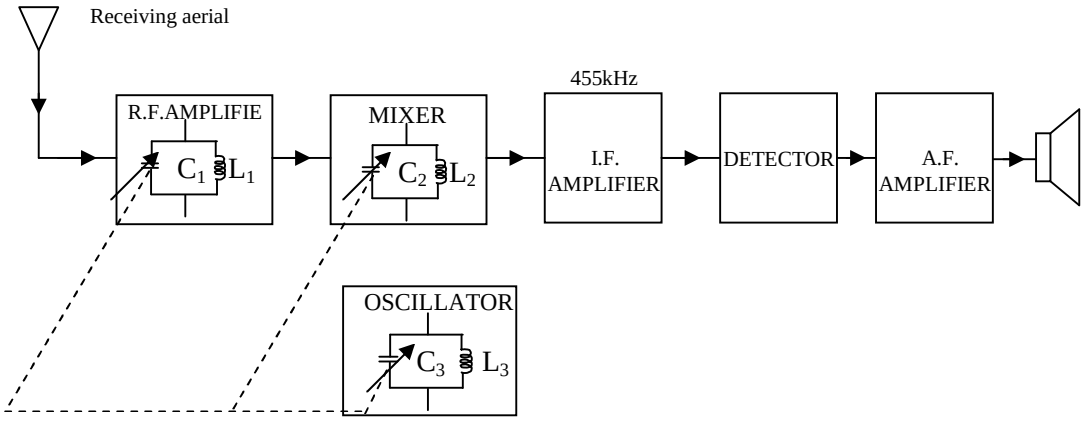


Fig 7.1.7

RF amplifier stage: This stage uses a tuned parallel circuit L_1 C_1 with a variable capacitor 'C_1'. The radio waves from various broadcasting stations are intercepted by the receiving aerial and are coupled to this stage. This stage selects the desired radio wave and raises the strength of the wave to the desired level.

Mixer stage: The amplified output of RF amplifier is fed to the mixer stage. Hence it is combined with the output of local oscillator. The two frequencies beat together and produce an intermediate frequency (IF).

IF = Oscillator frequency – Radio frequency.

= 455 KHz (always)

IF amplifier stage: The output of mixer is always 455KHz and is fed to fixed tuned IF amplifiers. These amplifiers are tuned to one frequency and give nice amplification.

Detector stage: The output from the last IF amplifier stage is coupled to the input of the detector stage. Here, the audio signal is extracted from the input.

AM amplifier stage: The audio signal output of the detector stage is fed to multistage audio amplifier. Here, the signal is amplified until it is sufficiently strong to drive the speaker. The speaker converts the audio signal into sound waves corresponding to the original sound of the broadcasting station.

Advantages of super heterodyne circuit:

The super heterodyne principle has the following advantages.

High RF Amplification.

Improved selectivity.

Low cost.

UNIT II

LESSON – 7

SECTION II

Frequency modulation and detection

Introduction:- Amplitude modulated signals are influenced by noise. Most noise appears as additional amplitude modulation on the signal. The effect of noise is minimized in frequency modulation.

Def: When the frequency of carrier wave is changed in accordance with the instantaneous value of signal, it is called frequency modulation.

In frequency modulation, only the frequency of the carrier wave is changed in accordance with the signal. However, the amplitude of the modulated wave remains the same i.e. carrier wave amplitude. The amount of change in frequency is determined by the amplitude of the modulating signal where as rate of change is determined by the frequency of the modulating signal as shown in fig.

In F.M., information or intelligence is carried as variations in carrier frequency.

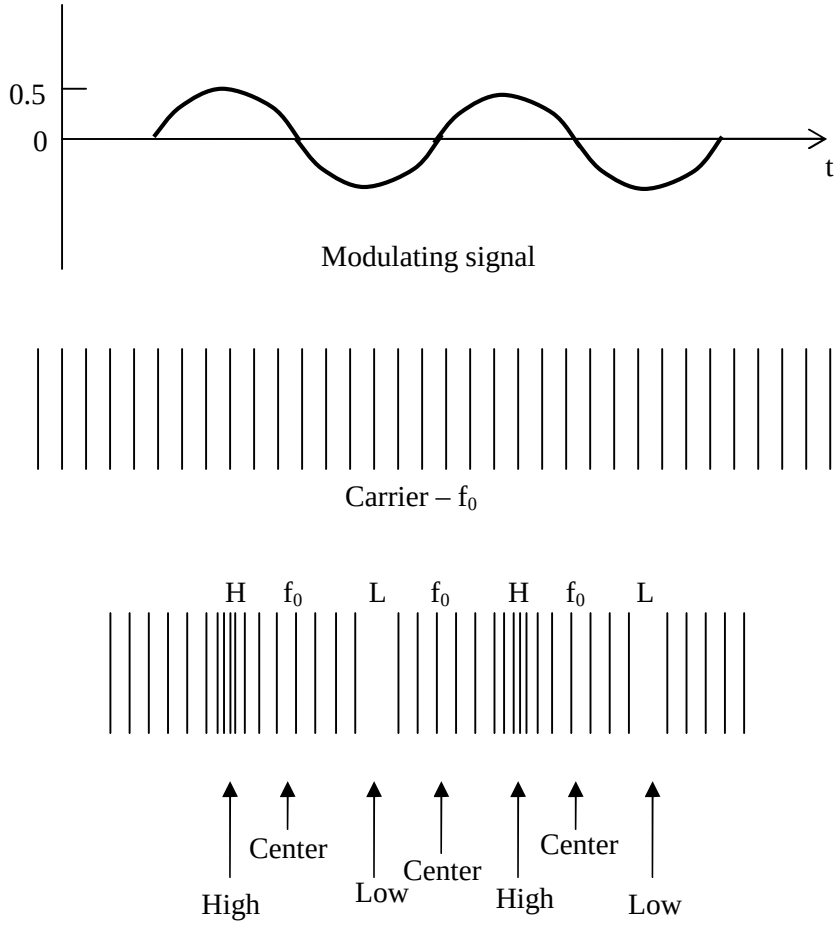


Fig 7.2.1

Mathematical expression for F.M. wave:

The unmodulated carrier is given by

$e_c = A \sin 2\prod f_0 t$ --------(1)

The modulating signal frequency is given by

$e_m = B \sin 2\prod f_m t$ --------(2)

The modulated carrier frequency 'f' swings around the resting frequency f_0. Thus

$f = f_0 + \Delta f \sin 2\prod f_m t$ --------(3)

Hence equation for the frequency-modulated wave becomes

$e = A \sin 2\prod f t$ --------(4)

$= A \sin \{2\prod (f_0 + \Delta f .\sin 2\prod f_m t) t\}$

(Substituting (3) in (4));

$$= A \sin (2\prod f_0 t + \frac{\Delta f}{f_m} .\cos 2\prod f_m t)$$

$e = A \sin (2\prod f_0 t + m_f .\cos 2\prod f_m t)$

$e= A \sin(\omega_0 t + m_f \cos\omega_m t)$ since $\omega = 2\prod f$

$e= A \{\sin(\omega_0 t) \cos(m_f \cos\omega_m t) + \cos(\omega_0 t) \sin(m_f \cos\omega_m t)\}$

The $\cos(m_f \cos\omega_m t)$ and $\sin(m_f \cos\omega_m t)$ terms have to be expanded into infinite trigonometric series involving sine and cosine terms of harmonics of ω_m. When the relevant series are substituted in equation x it can be shown that an FM wave comprises of an infinite number of side frequencies on both sides of the carrier frequency. The frequencies of these components are fc +fm, fc+2fm, fc+3fm, etc. The amplitudes of these components depend on the modulation index. The amplitudes are negligible when the separation between the carrier frequency and the side-band frequencies becomes sufficiently large. Therefore, in practice only a limited number of side frequency pairs are present. Also, as in AM the information is carried by side-frequency components. Hence the bandwidth needed for transmission or reception in case of FM is 2nfm where n is the number of pairs of side frequencies. The total power in an FM wave remains constant and does not depend on the depth of modulation.

Thus, analysis of FM wave shows that unlike the AM wave, which has two side frequencies for each modulating frequency, it has an infinite number of side frequencies.

Generation of F.M. waves:-

The methods used for FM generation can be grouped into two types. They are

(i) Direct method (ii) Indirect method

Direct method: In this method, the carrier frequency is directly modulated or varies in accordance with the input modulating signal such as (a) reactance modulator and (b) varactor diode.

Indirect method: The modulating signal is first integrated and then allowed to phase modulate a carrier. It is called indirect method because phase modulator is used for frequency modulation.

Reactance Modulator:- In frequency modulation, the frequency deviation must be made proportional to the amplitude of the modulating signal. In the earlier methods, it was done by varying capacitance or inductance in the oscillator tuned circuits. A more modern technique is to insert a reactance circuit across the tuned oscillator circuit. In general series RC and RL circuits are used as reactance circuits. Phase technique can be used for the determination whether the particular circuit acts as capacitive or inductive. Fig.7.2.2 shows a typical reactance modulator.

Here RF oscillator is connected to produce carrier waves. The capacitance C is chosen so that its reactance at the oscillator frequency is much greater than R_2 in parallel with 'R_1'. Thus the current flowing through 'C' leads the voltage applied between C and ground by 90^0. The base current therefore leads the carrier voltage by 90^0. The collector current is in phase with the base current and therefore in phase with the current which passes through 'C'. The total current is leading the carrier voltage by 90^0. Thus the entire circuit appears as a capacitor.

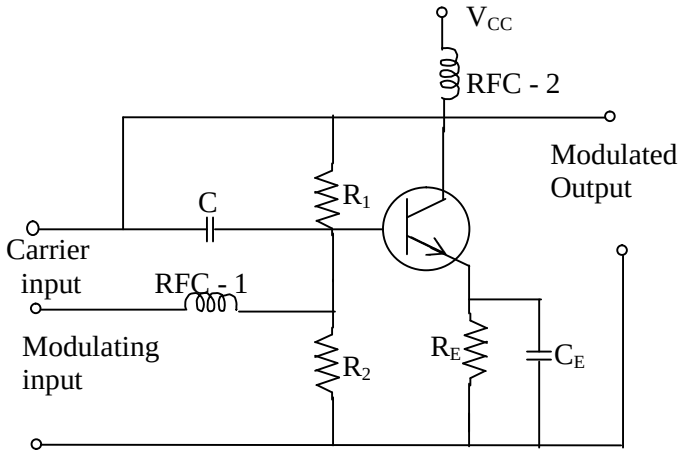


Fig 7.2.2 Transistor reactance FM Modulator

The effective or equivalent capacitance of the reactance modulator is given by

$$C_{eq} = \frac{h_{fe} R_2 C}{h_{ie} + R_2} \text{ ---------- (1)}$$

The audio signal varies the operating point, which varies h_{fe} of the transistor, and thus the equivalent capacitance of the transistor reactor modulator according to the eqn.(1). The effective capacitance C_{eq} is controlled by h_{fe}. Then the oscillator frequency across the tuned circuit

$$f = \frac{1}{2\pi\sqrt{L(C+Ceq)}} \text{ ---------- (2) where}$$

L = Inductance and

C = capacitance

F.M. Detection:

We know that an F.M. signal contains information in the form of frequency variation of the carrier signal. A simple method of converting frequency variations to voltage variations depends on the principle that reactance of the inductor or capacitor varies with the frequency. Thus the amplitude of current in the inductor or capacitor varies with the frequency of FM signal. As the frequency variations of FM signal depend on the amplitude of the AF signal, the current variations in the inductor or capacitor correspond to the AF signal. These amplitude variations in current, when made to flow through a resistor they will produce corresponding voltage variations across the resistor.

There are various types of circuits used for FM demodulations. They are

Phase discriminator or frequency discriminator

Ratio detector and

Quadrature detector.

A circuit that gives an output voltage whose amplitude is proportional to the frequency of the input signal can be used to decode or detect FM signals. Such circuits are called frequency discriminators.

Now, let us discuss in detail about the functioning of Foster-Seeley discriminator.

Foster - Seeley discriminator:

The phase shift between Primary and Secondary voltages of a tuned transformer is a function of frequency and the Foster-Seeley discriminator utilizes this frequency phase dependence for the recovery of the modulating signal. The following fig.7.2.3 shows the basic arrangement for the Foster-Seeley discriminator.

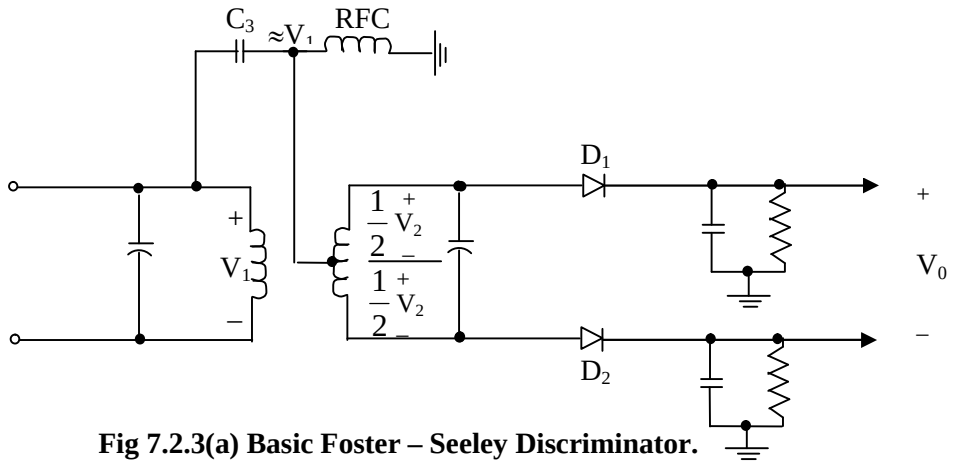


Fig 7.2.3(a) Basic Foster – Seeley Discriminator.

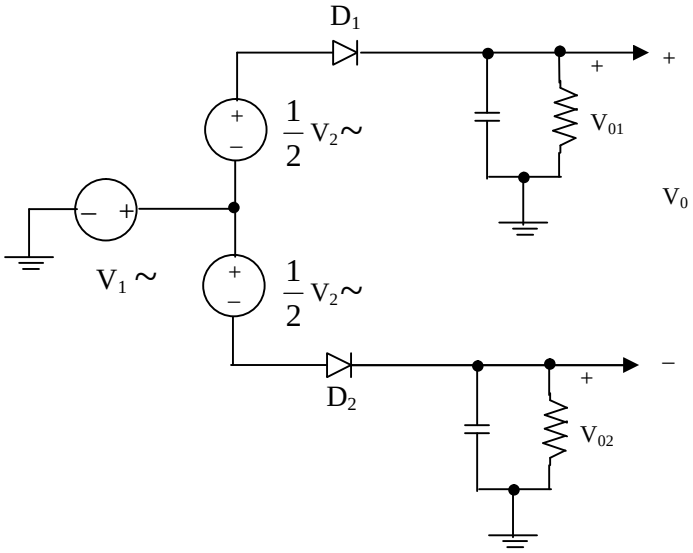


Fig 7.2.3(b) Voltage generator equivalent circuit

The primary voltage is tightly coupled through capacitor 'C_3' and the RFC to the center cap on the secondary. The coupling is tight enough that practically all the primary voltage appears between the centre top and ground.

SUMMARY OF SECTION – I

1. Amplitude modulation is a process of modulating a signal for efficient transmission. Carrier wave can be filtered using the process of demodulation by rectifying the modulated signal.
2. Frequency modulated signal does not consist of noise as in A.M. signal.
3. Modulated Broadcasting signals are received by a radio receiver with the help of antenna and resonant circuit. It is amplified, demodulated and fed to a speaker.
4. The important characteristics of a receiver are sensitivity, selectivity, fidelity, and noise figure.

SUMMARY OF SECTION – II

1. In frequency modulation the frequency of the carrier wave is modulated as a function of the instantaneous value of the modulating signal. Un like amplitude modulation there will be infinite number of sideband in the spectrum of F.M.wave. However, the bandwidth converges to 150kHz.
2. Several direct and indirect methods are available to produce FM waves.
3. F.M waves are detected by using Foster-Seeley discriminator circuit.

Self – assessment questions:

1. Draw the circuit diagram of amplitude modulator and explain its working.

2. Define modulation, define amplitude and frequency modulations and explains about them.

3. Explain how the Amplitude modulated waves are generated. Give the necessary diagrams.

4. With neat diagram and waveform, explain how the Amplitude modulated waves are detected.

5. Give the analysis of an AM wave. What are side bands? Explain Them.

6. Derive an expression for a Frequency modulated wave.

7. Discuss the principles involved in the production and detection of FM wave.

8. Give the block diagram of a super heterodyne receiver and explain the function of each block.

9. Explain a method for the generation of SSB modulation.

10. How can we detect an FM wave using Foster- Seeley discriminator? Explain with neat diagrams.

Short Answer questions.

1. Discuss the need for modulation in radio and TV Transmissions.
2. What are sidebands produced in AM Wave. Explain them.
3. Distinguish between Amplitude and frequency modulation.
4. Write in detail about AM diode detector.
5. Explain the principle of super heterodyne reception.
6. Explain now a SSB modulated system can be detected.

Key Terminology:-

Modulation: Process of combining low frequency signal with a high frequency radio wave (carrier wave).

Balanced modulator: A modulation circuit which cancels but the product term does not cancel

Name of the author: -

Dr J. Siva Rama Krishna M.Sc., Ph.D.

Department of Physics,

N.U.P.G. Center,

NUZVID

UNIT II

LESSON – 8

Basic principles of T.V. Transmission & Reception

Objectives:

To introduce the basic concepts of Television signal transmission, reception,

and radio wave propagation.

Structure of the lesson :

8.1 Section – I: T.V. Propagation and Reception

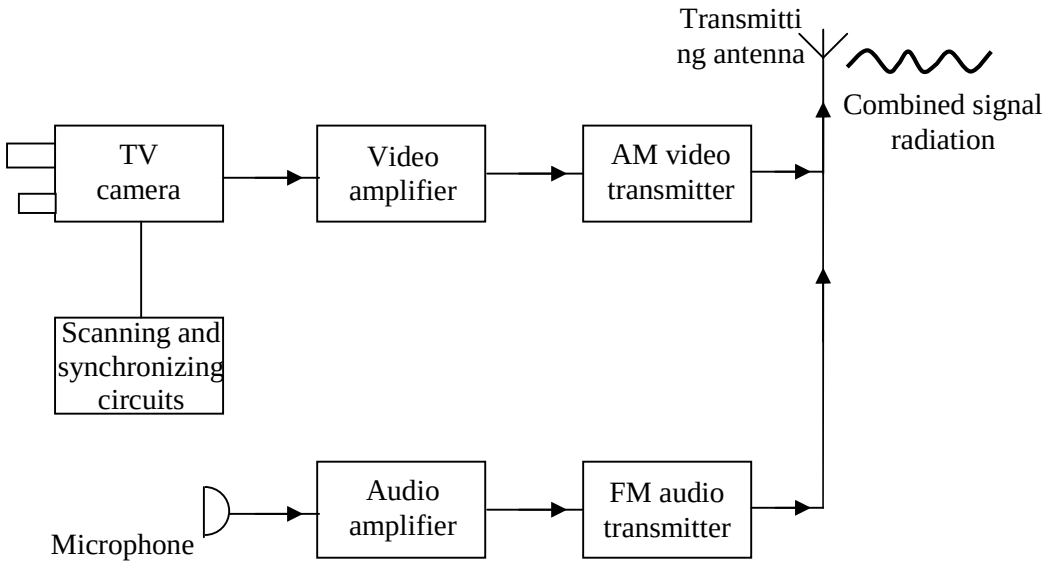
8.2 Section – II: Radio wave Propagation.

8.1 T.V Transmission and Reception

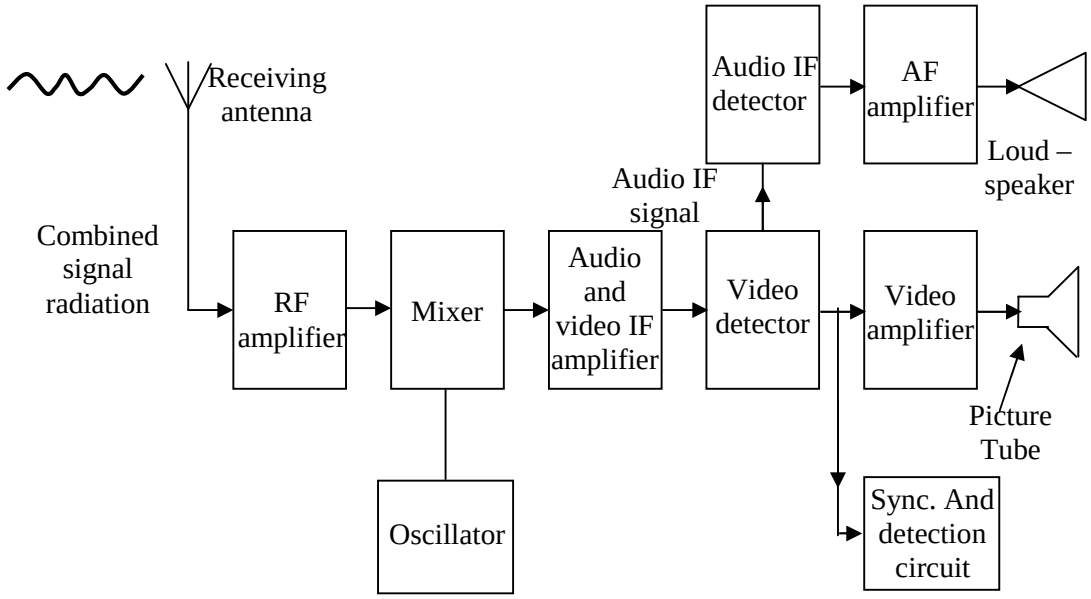
8.1.1 Television System: Introductory ideas.

The radio transmission of images of objects and scenes to far off places is one of the greatest achievements of the radio communications technology. In radio facsimile systems the video signals of the objects such as photographs, written or printed material etc are transmitted to a distant station where these are reproduced in the form of still pictures. In T.V. systems the images of a moving picture or scene are transmitted by means of rapid succession of video signals corresponding to these images.

The block diagram of a basic television system for transmission and reception of picture and sound signals is shown in the below fig 8.1.1



(a)



(b)

8.1.1 Block diagram of a TV System: (a) Transmitter (b) Receiver.

TV transmitter radiates two separate carriers by a single antenna. One of the carriers is frequency modulated by the audio or sound signal and the other is amplitude-modulated by the video or picture signals. The words audio and video in Latin mean to hear and to see respectively. The audio carrier frequency is 4.5 MHz higher than the video carrier frequency.

The audio and video signals are processed separately at a transmitting station as shown in fig (8.1.1). A T.V. camera is used to scan the different portions of a picture to be transmitted and produces a video signal proportional to the higher intensity of the portion of the picture under test. The synchronizing circuits produce synchronizing signals. These keep the reproduction of the picture at the receiver in step with the scanning at the transmitter. The composite video signal is amplified by a number of video amplifiers and is used to amplitude modulate the radio frequency (r.f.) carrier. The sound to be transmitted is picked up by a microphone and the audio signal is amplified by a number of audio amplifiers. The generated amplified audio signal is then used to frequency modulate the r.f carrier. Both the A.M. and F.M. carriers are carried to the single transmitting antenna, and are radiated out.

At the receiving point both the carriers are intercepted by the receiving antenna. The required T.V. channel is selected by the tuned circuits. The A.M. & F.M. carriers are amplified by RF

amplifiers and then heterodyned. The value of IF is usually 45.75 MHz for the picture. The audio and video signals in IF are then fed to the video detector. The video detector has two functions. (a) It demodulates the Composite video signal by using diode detector. (b) It separates the audio and video IF signals. The demodulated composite video signal is applied to the amplifier as well as synchronous signal separator. The output signal from the video amplifier interacts with an electron beam to produce a visible image on the face of the picture tube or kinescope. The synchronous separator separates the synchronizing signals from the remainder of the video signal. The synchronous signals are then applied to the beam deflection circuits to keep the electron beam that forms the image on the picture tube in step with the scanner at the transmitter.

8.1.2 Television camera tubes:

A TV camera tube sees the signal to be televised and converts its optical image into an equivalent electronic image. The picture elements of this electrical image are scanned to generate a video signal proportional to the brightness of the elements. T.V. camera tubes of various types have been developed. Let us discuss two common types namely

1) The image orthicon and
2) Videcon

The image orthicon:

A schematic diagram of the image orthicon camera tube is shown in the fig 8.1.2.

It consists of three main sections namely, (a) the image section (b) the scanning section (c) the electron multiplier section.

The image section of the tube contains a photo cathode, a target plate of low resistance thin glass and a screen positioned very near the target. The photo cathode is a thin layer of photo sensitive material and is kept at negative potential. Light from a scene to be televised is passed through an optical lens and is focused on the photo cathode. This photo cathode emits photoelectrons in proportion to the light intensity of the image. These electrons meet at common point on the target plate. On striking the target plate these electrons produces secondary electrons, which are attracted by the screen. The image signal is obtained by scanning the backside of the target using a low velocity electron beam emitted by the cathode of electron gun. Consequently, the current varies in accordance with the light intensity distribution of the optical image. In electron multiplier section

the number of electrons get multiplied. The multiplier output is an amplified version of the return beam current. This output current produces output voltage across the load resistance R_L. This voltage is called video voltage. The fluctuations in the output voltage correspond to the brightness variation in the optical image.

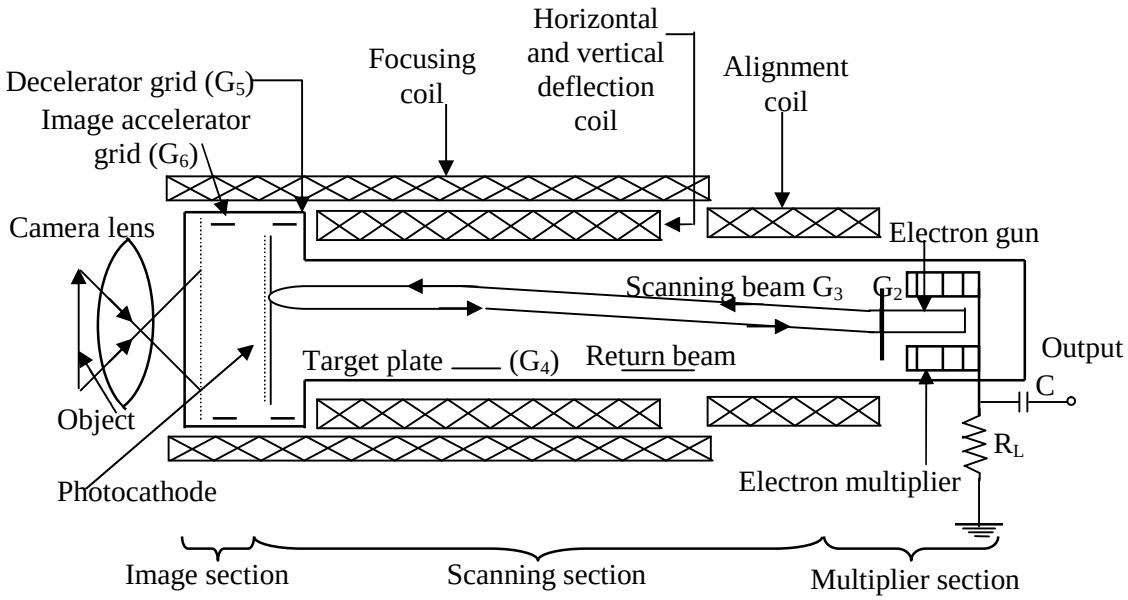


Fig 8.1.2 Schematic diagram of an image orthicon camera tube.

The vidicon:-

The vidicon camera tube employs a photo conducting material such as amorphous selenium semi conductor whose resistance decreases on exposure to light. A schematic diagram of this tube is shown in the fig 8.1.3. Here the signal plate is a transparent thin conducting film of a metal and forms the target of the tube. One side of the signal plate has a glass face plate and the other side is coated with a very thin layer of the photo conducting layer. The optical image to be televised is focused through the face plate and transparent film on to the photo conducting layer which responds in accordance with the light intensity of the image. The dark portion of the layer behaves as insulator while the bright portions become partially conducting.

Thus as the electron beam scans the surface of the photoconductive layer and it deposits charge, which varies with time in accordance with the variations in the brightness of the successive elements of the layer. Therefore, the current through the load resistance 'R_L' and hence the output voltage reproduces the variations in the light intensity of the successive

portions of the optical image. A large output voltage represents the bright portion and a small output voltage represents the dark portion of the object.

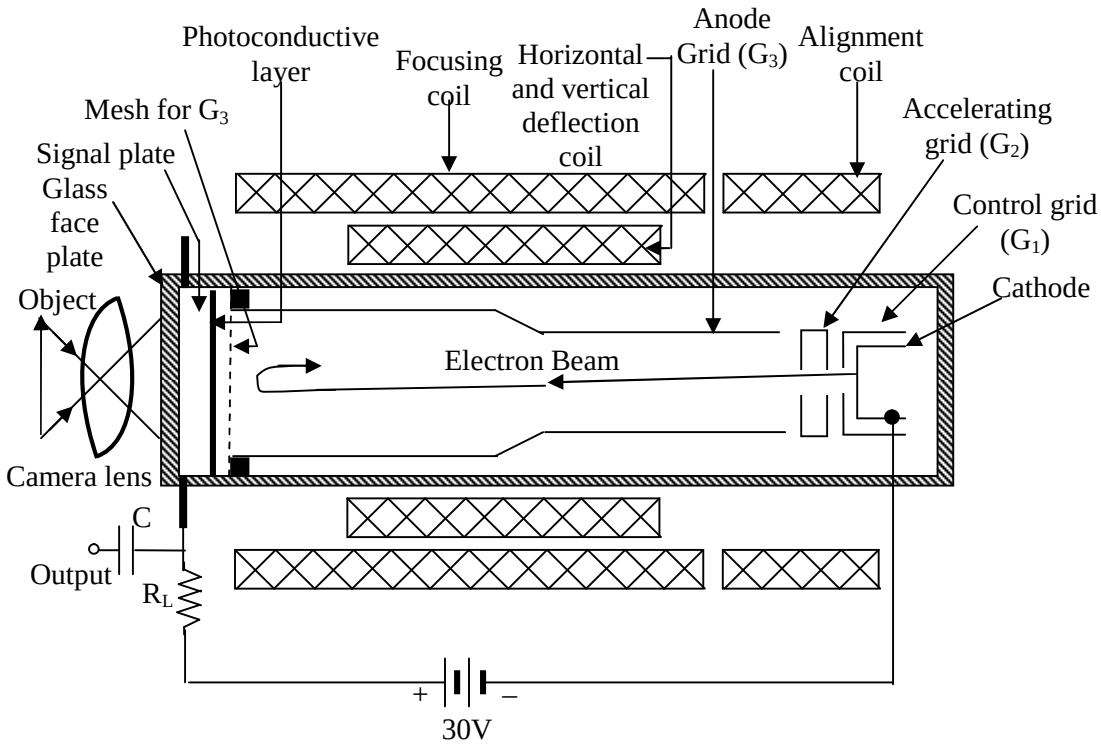


Fig 8.1.3 Schematic diagram of a vidicon picture tube.

Comparison between the Image Orthicon and Vidicon:

Both the tubes are simple, compact and equally sensitive and are widely used in TV industry. The operating principle of these tubes is different. The image orthicon works on the principle of photo-emission and electron multiplication while the Vidicon uses the principle of photo conductivity. The variation in the output voltage of each tube depends on the level of illumination of the target plates of these tubes. The sense of this variation is however opposite in the two tubes. In image orthicon a large output voltage corresponds to a dark portion of the image where as in vidicon it corresponds to a bright portion.

The Vidicon has the advantages of low cost, small size and simplicity of adjustment.

In spite of these advantages, it suffers from the following drawbacks.

(i) It has non-linear response characteristic.

(ii) Very rapid motion pictures cannot be produced very satisfactorily.

8.1.3 Scanning: The scanning of the target by the electron beam is similar to the way we read a page of print. i.e. from left to right and top to bottom. In effect, the picture is changed into a set of parallel lines called raster. Hence, scanning requires vertical and horizontal deflection systems. The one, which moves it steadily, from left to right and causes it to 'fly back' rapidly for the scan of next line, is called line scan. The other, the field scan operates simultaneously and moves the beam at a much slower rate in the down ward direction (line after line) until the picture is completely scanned and finally it restores back the beam suddenly to the top. The two pairs of coils mounted round the camera tube form the magnetic deflection system and provide the line and field scans.

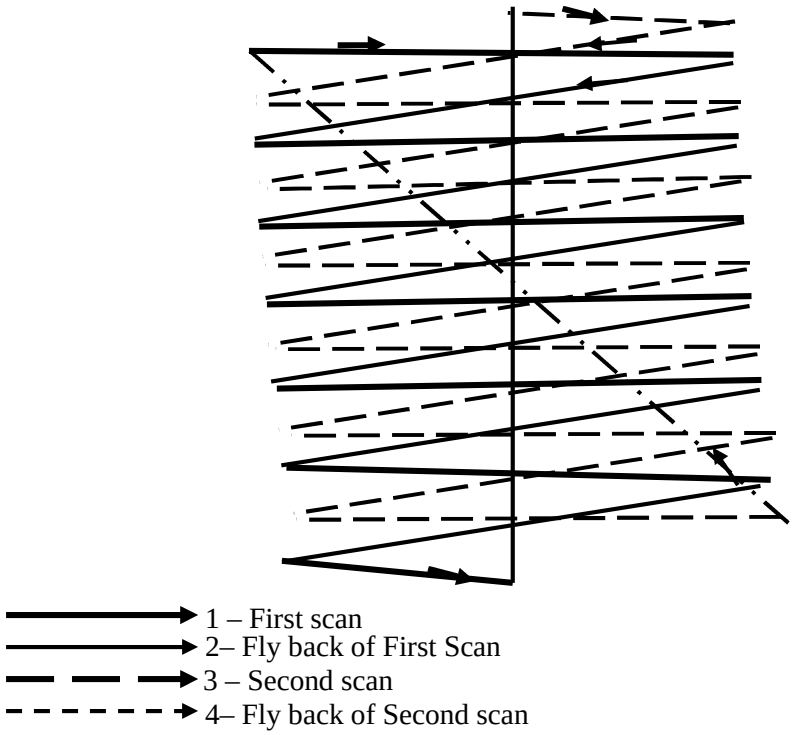


Fig 8.1.4 Interlaced scanning

The total scan of the target should occur at a rapid rate. Otherwise the continuity between successive scans is lost and causes "flicker". The band width of the video signal can be reduced to half by using interlaced scanning technique, in which alternate lines are scanned by electron beam producing half picture in every 1 /50 sec. and returns to scan the intervening lines (fig 8.1.4) due to persistence of vision, we will not notice any flicker. All TV. Systems employ interlaced scanning.

8.1.4 Synchronization:-

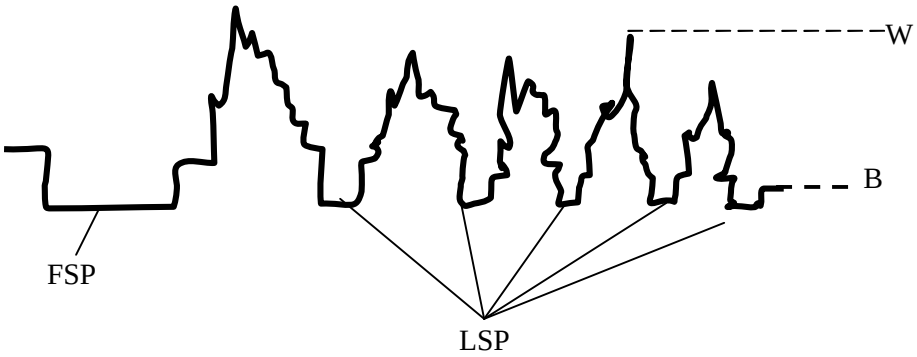


Fig 8.1.5 Composite Video signal

FSP – Field Sync . Pulses
LSP – Line Sync. Pulses
W – White (picture brightness)
B – Black (picture darkest)

When the TV camera starts scanning line '1', the receiver must also start scanning line 1. When a complete set of horizontal lines has a beam scanned, moving the electron beam from the end of the bottom line to start the top line (vertical fly back), the retrace must occur simultaneously at both transmitter and receiver. To ensure this, synchroniz-ing pulses are also sent along with picture signal. These are added to the video signal during the fly back times, when the beam is blanked out. Thus, the field synchronizing pulses will be longer and less frequent than the line sync pulses. A simplified waveform with line and field sync pulses is shown in the fig 8.1.5

Bandwidth:

The large video bandwidth necessitates the use of VHF – UHF frequencies as r.f. carrier frequencies. Amplitude modulation of the carrier with the video signal produces two side bands each 5MHz wide. This double side band vision signal will occupy a 10MHz frequency range spectrum with the carrier at the center.

8.1.5 Color television:

Color TV uses the fact that any color may be synthesized by the addition of three primary colors blue(B), green(G), and red (R) in appropriate proportions. The principles of transmission are similar to those of black & white TV. In addition, specific color information must be sent. A

practical requirement is that the color signal must produce a black & white picture on monochrome receiver.

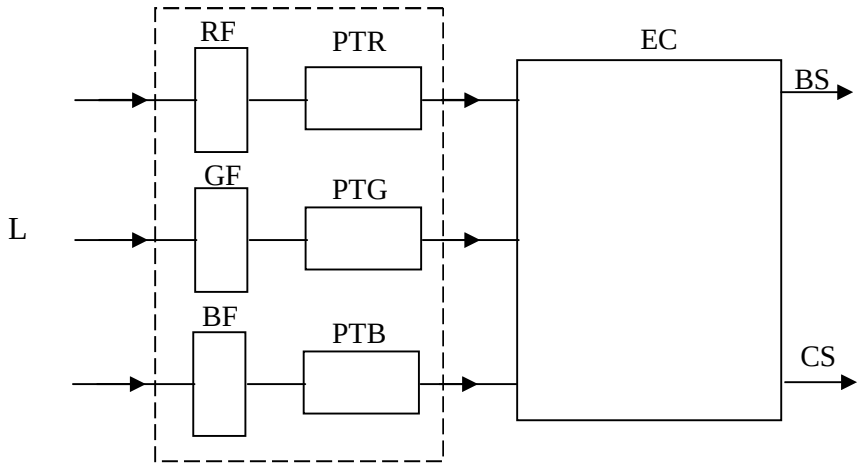


Fig 8.1.6 Block diagram of color TV Camera

L – Light from picture; RF – Red Filter
GF – Green Filter; BF – Blue Filter
PTR – Plumbicon tube for red color
PTG – Plumbicon tube for Green color
PTB – Plumbicon tube for Blue Color
EC – Encoding Circuit
BS – Brightness (Luminance) Signal
CS - Color (Chrominance) Signal

In color TV camera three Plumbicon tubes are required, each viewing picture through a different primary color filter. The red, green and blue signals so obtained provide the required color information. When added together they give brightness variations. See fig.

8.1.6 T.V. Receiver

Introduction: A T.V. Receiver utilizes the super-heterodyne principle like the radio receiver. T.V. receiver has to process audio (sound), video (picture) and synchronizing signals. The basic functions of a TV receiver are

(i) Interception of radiated picture and sound signals. Antenna does this job.

(ii) Selection of the desired channel signal from among the different signals received by the antenna.

(iii) Conversion of r.f. signal into an IF signal for amplification.

(iv) Separate detection and amplification of audio and video signals by FM and AM detectors.

(v) Separation of sync. Pulses from the video signal and application of these pulses to generate time base signals. Feeding of these signals to scan coils of the picture tube to reproduce the picture intact.

(vi) Conversion of audio signal to sound by a loud- speaker and video signal to picture by a picture tube.

The basic elements of a TV receiver for black & white reception are shown in the below fig.8.1.7

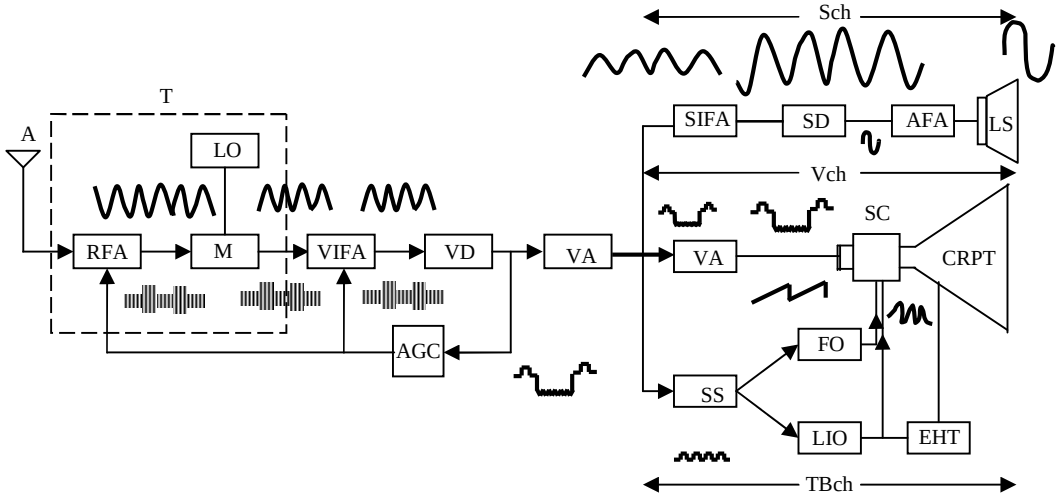


Fig 8.1.7 The basic elements of TV receiver for black & white reception

The elements are 1) Antenna (2) tuner (3) video IF amplifier (4) video detector (5) automatic gain control (6) video channel (7) cathode ray picture tube (8) time base channel and (9) sound channel.

Receiver picture tube:

The picture tube is used to reproduce the picture transmitted by the transmitting station. It is a cathode ray tube, which employs magnetic deflection instead of electro state deflection. The individual picture elements scanned at the TV Camera in the transmitter are reproduced on the fluorescent screen of the picture tube. When looked from a distance, these individual tiny picture elements give the impression of a continuous picture.

8.2 SECTION-II Radio waves Propagations

8.2.1 Introduction:-

Radio communications use electromagnetic waves propagated through the earth's atmosphere or space to carry information over long distances without the use of wires. Radio waves with frequencies ranging from about 100 Hz to above 300GHz have been used for communications purposes. More recently radiation in and near the visible range have also been used. Although the electric and magnetic fields exist simultaneously, in practice antennas are designed to work through one or other of these fields. To describe radio wave propagation sinusoidal or co-sinusoidal variations will be assumed unless stated otherwise.

8.2.2 Propagation in free space:-

Mode of propagation:- Consider first an average power P_T assumed to be radiated equally in all directions. This will spread out spherically as it moves away from the source. At distance 'd', the power density in the wave (power per unit area of wave front) will be

$$P_{di} = \frac{P_T}{4\pi d^2} \text{ watt / metre}^2 \qquad \text{---------- (8.2.1)}$$

This is so because $4\prod d^2$ is the surface area of the sphere of radius 'd', centered on the source.

P_{Di} = Isotropic Power density

It is known that all practical antennas have directional characteristics. The directivity gain is the ratio of actual power density along the main axis of radiation of the antenna to that which would be produced by an isotropic antenna at the same distance fed with the same input power.

Let G_T = Maximum directivity gain of the transmitting antenna

Then, the power density along the direction of maximum radiation will be,

$$P_D = P_{Di} \cdot G_T = \frac{P_T G_T}{4\pi d^2} \quad \text{---------- (8.2.2)}$$

A receiving antenna can be positioned in such way that it collects, maximum power from the wave. When so positioned,

Let P_R = Power delivered by antenna to the load (receiver) under matched conditions.

$$\text{Then } P_R = P_D A_{eff} = \frac{P_T G_T}{4\pi d^2} \cdot A_{eff} \quad \text{---------- (8.2.3)}$$

Where A_{eff} = Affective area of the antenna

It can be shown that for any antenna;

$$\frac{A_{eff}}{G_R} = \frac{\lambda^2}{4\pi} \quad \text{---------- (8.2.4)}$$

Here 'λ' is the wavelength of the wave being radiated.

Let G_R = Max. Directivity gain of the receiving antenna.

$$\text{Then } \frac{P_R}{P_T} = G_T \cdot G_R \cdot \left(\frac{\lambda}{4\pi d}\right)^2 \quad \text{---------- (8.2.5)}$$

This is the fundamental equation for free space transmission. Usually, it is expressed in terms of frequency 'f' in MHz, and distance 'd' in kilometers.

$$\text{We know that } \lambda f = c \quad \text{---------- (8.2.6)}$$

Substituting equation (6) in equation (5), we get,

$$\frac{P_R}{P_T} = G_T \cdot G_R \frac{\left(0.57 \times 10^{-3}\right)}{(df)^2} \quad \text{---------- (8.2.7)}$$

By expressing power ratios in decibels, equation.8.2.7 becomes

$$\left(\frac{P_R}{P_T}\right)_{dB} = (G_T)_{dB} \cdot (G_R)_{dB} - (32.5 + 20 \log_{10} d + 20 \log_{10} f) \quad \text{---------- (8.2.8)}$$

The third in parenthesis on RHS of equation.8.2.8 is the loss in dB, resulting from the spreading of the wave as its propagates outward from the source. It is the known as the transmission path Loss L.

Thus $L = (32.5 + 20 \log_{10} d + 20 \log_{10} f)$ dB ---------- (8.2.9)

Where 'd' is in Km and 'f' is in MHz

Then equation.8.2.8 becomes

$$\left(\frac{P_R}{P_T}\right)_{dB} = (G_T)_{dB} . (G_R)_{dB} - (L)_{dB} \quad \text{---------- (8.2.10)}$$

8.2.3 Propagation of radio waves:-

Factors involved in the propagating radio waves:-

There are a number of mechanisms by which radio waves may travel from a transmitting to a receiving antenna. The more important of these are designed by the terms.

(i) ground waves, (ii) Sky waves, and (iii) Space wave or tropospheric waves.

The ground wave (also sometimes called surface wave) can exist when the transmitting and receiving antennas are close to the surface of the earth when vertically polarized. This wave, supported at its lower edge by the presence of ground is of practical importance at broadcasting and lower frequencies.

The sky wave represents energy that reaches the receiving antenna as a result of a bending of the wave path introduced by ionization in the upper atmosphere. This ionized region is called the ionosphere. This begins about 80 Kms above the earth's surface and is useful for very long distance radio communication.

The space wave represents energy that travels from the transmitting to receiving antenna in the earth's Troposphere, Troposphere is the portion of the earth's atmosphere in the first ten miles adjacent to the earth's surface. A space wave consists of two components.

1) A ray that travels directly from transmitter to receiver and
2) A ray that reaches the receiver as a result of reflection from the surface of the earth.

Space wave energy may reach the receiver (i) as a result of reflection or refraction due to the variations in the electrical characteristics of the troposphere .and (ii) by diffraction around the curvature of the earth, hills etc.

Radio transmission at frequencies above about 30MHz is normally the range of space wave propagation.

Ex:-

(a) Television ,

(b) F.M.

(c) Radar use the frequencies in the rays g space wave propagation.

8.2.4 The ground wave propagation:-

In ground wave propagation, the electromagnetic waves travel along the curved surface of the earth from transmitter to receiver. In fact, they depend on the earth for a portion of their transmitting medium. These waves are influenced by the electrical characteristics of the ground, over which they travel. They are strongly absorbed by dry land. As the moisture and sea water have greater conductivity, the ground waves are much less absorbed. The table 8.2.1 shows the radio frequency spectrum.

The attenuation of ground waves depends on their direction of polarization also. These waves are always vertically polarized because the horizontal component of electric field is short circuited by the earth. But, the vertical component travels along the surface of the earth because of the electrical discontinuity between ground and atmosphere.

As the wave travel along the ground, it induces charges in the ground. These charges travel along with the wave and have induced currents. As the earth has definite resistance

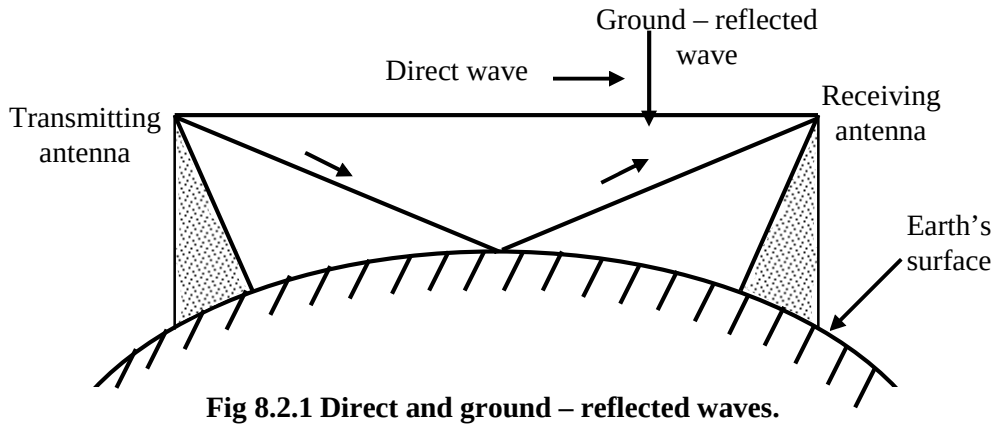


Fig 8.2.1 Direct and ground – reflected waves.

TABLE 8.2.1 Radio spectrum

Frequency band	Frequency range	Wave length range	Typical applications
Very low frequency (VLF)	10 – 30 kHz	30 – 10km	Long – range direct communication
Low frequency (LF)	30 – 300 kHz	10 – 1 km	Marine, navigational aids
Medium frequency (MF)	300 kHz – 3 MHz	1 km-100m	Broadcasting, marine
High frequency (HF)	3 – 30 MHz	100 – 10m	All types of communication
Very high frequency (VHF)	30 – 300 MHz	10 – 1m	TV, FM, radar, short – wave, air navigation
Ultra – high frequency (UHF)	300MHz – 3 GHz	1m-10cm	Radar, microwave relays, short distance communications
Super – high frequency (SHF)	3 – 30 GHz	10 – 1cm	Radar, radio relay, navigation satellite communication
Extremely high frequency (EHF)	30 – 300GHz	1cm – 1mm	Experimental

the flow of induced current results in power loss. This causes absorption of energy from E.M. Waves.

Ground waves get attenuated in another way also. As the wave travels over the ground, the wave front tilts, more and more towards the surface due to diffraction effects. This increases the

horizontal component of electric field at the cost of vertical component. This causes greater attenuation.

The absorption of ground waves increases with frequency. Therefore ground wave propagation is useful only at low frequencies.

Ground waves are useful in communication for the frequencies below 500KHz and over distances up to 1500 kms. Amplitude modulated radio broadcasts in the medium frequency band are transmitted mainly via the ground wave. But at higher frequencies the ground waves are absorbed. So, these cannot be used beyond few Kilometers around the transmitter. Ground wave transmission makes the reception extremely reliable and not subject to the seasons atmospheric conditions.

8.2.5 Space or Tropospheric waves:-

The space or tropospheric wave is that which travels from transmitting antenna to receiving antenna through the earth's troposphere. That is through the portion of the earth's atmosphere with in the first 15 kms over the surface of the earth. The space wave is usually made up of two components. They are

a) The direct or line-of-sight wave from the transmitting to receiving antenna.
b) The ground reflected wave traveling from the transmitting to the receiving antenna after reflection from the ground. These two components are shown in the fig .8.2.22

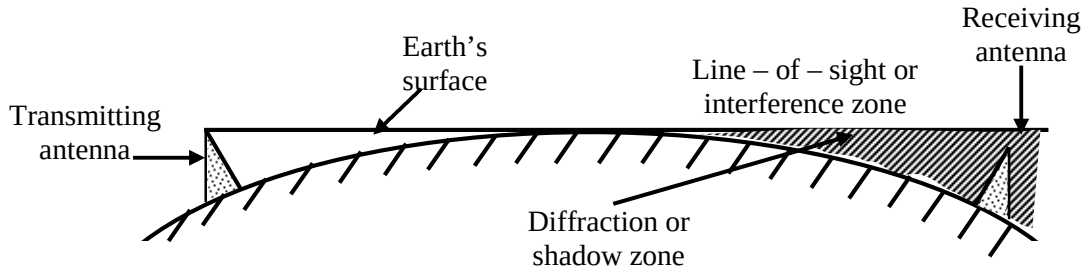


Fig 8.2.2 Interference and shadow zones.

The phase of the ground reflected wave changes by 180^0 after reflection from the ground. The strength of the signal received at the receiving end depends on the phase relationship between the direct and reflected waves.

If the path length of the direct and reflected waves differ by nλ, where 'n' is an integer and 'λ' the wavelength, then the total path difference between the two waves becomes $(n + \frac{1}{2})\lambda$.

Here, the additional path difference $\frac{\lambda}{2}$ is because of reflection. Therefore, the two waves reaching the receiving antenna will have opposite phases and will cancel each other. If the path lengths of the direct and reflected components differ by, half-integral multiple of λ the total path difference will become an integral multiple of 'λ'. Then the two waves arrive in phase and reinforce each other. Thus by varying the height of the receiving antenna the path difference between the two waves can be changed. The signal strength can be increased or decreased. This phenomenon is referred to as selective fading.

The space wave does not undergo continuous absorption by the surface of the earth. It can therefore cover large distances than the ground wave. The region beyond the line of sight is called diffraction zone or shadow zone. The signal in the shadow zone is also increased due to a phenomenon known as duct propagation.

The region outside the shadow zone where the line of sight propagation is possible is called the interference zone. In this region the two waves interfere to give the resultant field.

8.2.6 Radio horizon:-

The maximum distance over which the space wave can be transmitted is called the effective horizon or radio horizon. It is greater than the optical horizon. Optical horizon is the straight-line distance between the two points. The radio horizon of an antenna depends on its height above the ground. It is given by, an empirical relation,

$d_t = 4\sqrt{h_t}$ where

d_t = Radio horizon of the transmitting antenna of height h_t.

Similarly, $d_r = 4\sqrt{h_r}$ where

d_r = Radio horizon of the receiving antenna of height 'h_r'

Therefore the maximum limiting distance between the two antennas = $d = d_t + d_r$

$= 4(\sqrt{h_t} + \sqrt{h_r})$.

8.2.7 Sky waves:-

A sky wave is that which arrives at a receiving antenna after reflection from the ionized layers of the earth's upper atmosphere known as ionosphere.

The short wave communication around the world takes place by the sky waves, via successive reflections at the earth's surface and the ionosphere. The transmission by this mode of propagation depends on the properties of ionosphere. It is subject to fading with seasons, day and night and atmospheric conditions. It is the primary means of propagation over a wide range of frequencies extending up to about 30MHz.

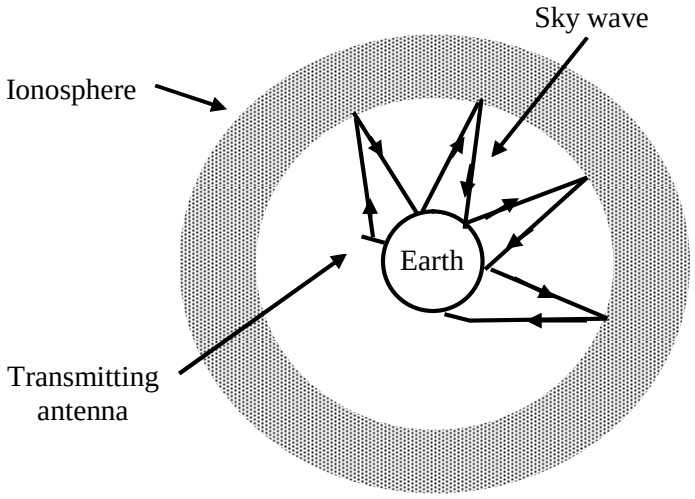


Fig 8.2.3 Around the world communication by means of sky waves.

Some terms relating to sky wave communication:

i) Skip distance:

Let a radio wave with a frequency higher than the critical frequency of layer be incident on it at various angles. The situation is illustrated in the fig.8.2.4

It is found from the figure that as the angle of incidence decreases, the distance between transmitting and receiving points on the earth's surface diminishes and then increases. Compared to the other rays, ray 2 moves through a longer horizontal distance in the layer. For the ray to be received on the earth's surface there is a minimum distance. Between the transmitting and receiving points for a given frequency. This minimum distance is known as skip distance.

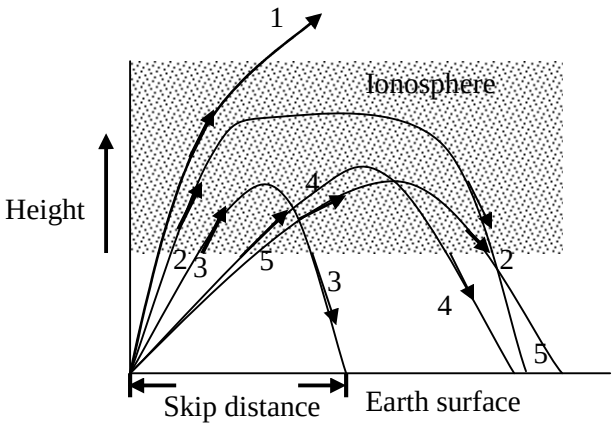


Fig 8.2.4 Ray trajectories in the ionosphere for different angels of incidence

In the above fig ray 3 is received at skip distance. Skip distance for a given frequency increases with increasing height of the layer. The skip distance is minimum at noon and increases at other times.

ii) Maximum usable frequency:-

The frequency of radio wave for which a given distance on the earth's surface is equal to the skip distance is called as maximum usable frequency (MUF). The MUF has diurinal variation. It is maximum at noon and falls on either side of the noon hour. For sky wave communication between two points the optimum frequency of the wave is taken to be about 85% of the MUF.

iii) Single and multi hop transmissions:

If a radio wave launched by a transmitter arrives at the receiver after its reflection from the ionosphere the transmission is referred to as a single hop transmission and the path of the wave is called single hop path.

Transmission using multiple reflections at the ionosphere is referred to as multi-hop transmission.

iv) Fading:

The fluctuation of the received signal existing over short period of the order of a minute or less is called as fading. Fading is an undesirable phenomenon in any communication system. It is one of the problems of radio wave communication using ionosphere reflection.

8.2.8 Summary of Lesson :

In T.V systems, images of moving pictures or scenes are transmitted as a rapid succession pf video signals corresponding to the images. A T.V. camera tube sees the picture to be televised and converts the optical image into an equivalent electronic image by using scanning techniques. Image Orthicon and vidicon are the popular T.V. Camera tubes. The electronic signals of scanned images are mixed with synchronizing signals and then amplitude modulated before transmission. Audio information is FM modulated and mixed with video signal along with synchronization signals and the composite signal is transmitted. The signals received at the T.V. receiver are separated into audio and Synchronization signals. The sync separated audio and video signals are amplified separately. The audio signals are fed to a speaker. The video signals are fed to a cathode ray (CR) tube. To the deflection coils of CR tube, sync signals are provided to see pictures without any flicker.

Summary of section II : To transmit signals to a distant receiver radio frequency waves are used as carriers. There are three modes of propagation for the radio waves.

1) Ground wave or surface wave, 2) Tropospheric wave or space wave, 3) sky wave. The choice of mode of propagation depends on the frequency of radio waves. Ground wave travels along the surface of the earth. So it can follow the earth curvature. It is vertically polarized. The attenuation of ground wave increases with frequency.

2) Space wave propagates through the earth's troposphere and consists of two components. 1) Direct or line of sight wave and the ground reflected wave.

The strength of a signal at a point depends on the interference of the two components.

A sky wave is that which arrives at a receiving antenna after reflection from the ionosphere.

It is used for short wave communication.

8.3 Key Words : 1) Vidicon 2) Orthicon 3) interlaced scanning 4) raster
5) Troposphere 6) Ionosphere 7) Sky wave 8) Ground wave
9) Skip distance 10) radio horizon

8.4 Self assessment Questions:

1. Describe the principle and working of a T.V.receiver with the help of block diagram.
2. Describe the operation of an image orthicon tube.

3. Draw the schematic diagram of a vidicon tube and describe its operation. Compare its working principle and advantages with that of image orthicon.
4. What is interlaced scanning? How is it accomplished?
5. Write notes on the following
   a) T.V.transmission and reception b) T.V.Receiver
   c) Color television d)Vestigial sideband modulation.
6. Describe the basic elements of a radio wave propagation.
7. What is fading of radio waves?
8. What is meant by duct propagation?
9. Describe the various layers of ionosphere.
10. Write notes on the following
    a)Critical frequency b) virtual height c) skip distance d) maximum usable frequency f) radio horizon

8.5 Reference Books :-

1. Electronic Communications – D. Roddy & J. Coolen.
2. Electronic Principles - Malvino.
3. Principles Electronics - V.K. Mehta
4. Basic Electronics - D.C. Tayal & Vimla Tayal
5. Basic Electronics (solid state)- B.L. Theraja
6.Electronics - D. Chattopadhyay & P.C. Rakshit
(Fundamentals & Applications)
7. Solid state Physics & Electronics – R.K.Puri & V.K. Babbar
8.Electronic Communication systems – George Kennedy
9.Electronic and Radio engineering – F.E. Terman
10.Electronic Principles - V.L. Kakani & Bhaudari

Name of author:

Dr J. Siva Rama Krishna M.Sc., Ph.D.
Department of Physics,
N.U.P.G. Center,
NUZVID

UNIT – III
LESSON: 9

DIGITAL ELECTRONICS

BASIC LOGIC GATES.

OBJECTIVES:

1) To understand what digital electronics is
2) To learn Boolean algebra
3) To understand various basic logic gates
4) To understand De Morgan's theorems
5) To know about universal gates

Structure:

9.1 Introduction:

The term digital refers to any process that is accomplished using discrete units, e.g. fingers, toes, digits etc. Each of these can be used as a unit or group of units to express a whole number. On the other-hand, analog numbers are represented as directly measurable quantities such as volts, distances, and rotations etc. Thus, in analog method, a number can be represented as an angle (in degrees) rotation of the needle on a meter. This method has been used widely in electronics to represent intensity, frequency, and time etc. The two factors namely accuracy and economy made the people to prefer digital readout devices. As examples we can quote: Digital multi-meters, Digital frequency counters, Screw gauges, Vernier calipers with digital readout facility.

Though there are many number systems, in digital electronics, binary system is used extensively. Good amount of accuracy and reliability that can be achieved with digital electronics. Because of this reason

conventional electronic circuits are being replaced with digital circuits. Already we see present day communication is based on digital techniques. Digital electronics is also extensively used in computers.

In computers information and data are processed in terms of only zeros and ones (0 and 1) which are called binary digits. In binary system those are only two digits 0 and 1. The Binary digits are called Bits. The programmer feeds instructions and data in a form that looks like English. For example ADD for addition, SUB for subtraction etc. Later computer converts these instructions and data into binary digits before processing. Processing of data in computers is performed by digital circuits. Because the data in computers is in the form of 0's and 1's, a special logic is developed by George Boole to perform the arithmetic and logic operations. Hence the algebra developed for binary systems is known as Boolean algebra.

In a digital system, we will consider inputs and outputs, as either 0's or 1's. In order to express the input and output relationship, an algebra known as switching algebra is useful. Shannon developed this in 1938 for use in telecommunication switching circuits. It is based on a more general algebra that was developed earlier by Boole. This algebra uses a limited number of operators to connect variables together to form expressions. Further, it is also easy to build electronic circuits to implement these operators. Boolean Algebra is the algebra of binary variables. Any system, which has only two states, can be expressed in terms of Boolean logic. For example a statement can be either true or false, if it is true we can represent it as logic 1 and if is false we represent it by 0. This is called positive logic.

Existence of positive voltage at a point may be taken as logic 1. Zero voltage can be represented as logic zero. A glowing bulb, a current flow, a happening of an incident can be represented by logic 1 or TRUE state. The complementary actions can be represented by a 0 or FALSE state. In fact we may represent actions that were represented by 1 by logic 0 also. It all depends on one's choice. The electronic circuits used to perform Boolean operations are called gates. The gate is a circuit with one or more input signals but only one output signal. These gates are constructed by using diodes and transistors (known as DTL logic) or using transistors only (known as TTL) Technical standards were evolved in electronics depending on the implementation of integrated circuits using various semiconductor technologies viz DTL,TTL,ECL etc. The most important and popular semiconductor logic is Transistor-Transistor Logic (TTL), in which a +5V DC is taken as logic 1 and zero volts is taken as logic 0 .

This chapter introduces the fundamentals of Boolean operations, corresponding logic circuits and gates and integrated circuits related to these circuits are discussed.

9.2 Boolean Algebra :

The fundamental operations in Boolean algebra are OR, AND and NOT, with symbols + (plus),. (dot) and bubble or bar over Boolean variable respectively.

In real life applications, one combined several logic gates and the combined effect determines a system behavior depending on the states of individual logic variables. This is usually expressed in terms of a truth table. In truth table, various inputs are listed and various combinations are worked out to determine the system behavior. As a simple example, see the OR gate truth table given in table 9.3a. Here A and B are the input variables. These two variables can have 2^2=4 combinations of 1s and 0s. As per the Boolean equation y = A + B (A+B to be red as A or B), output is true for 3 input combinations and false for one combination.

The following are the laws of Boolean algebra and complicated laws can be proved using simple Boolean laws. These laws can be verified by writing truth tables for L.H.S and R.H.S expressions.

1. $A + 0 = A$
2. $A + 1 = 1$
3. $A + A = A$
4. $A + \overline{A} = 1$

} Laws of 'OR'

5. $A . 0 = 0$
6. $A . 1 = A$
7. $A . A = A$
8. $A . \overline{A} = 0$

} Laws of 'AND'

9. $\overline{0} = 1$
10. $\overline{1} = 0$
11. of A = 0 then $\overline{A} = 1$
12. of A = 1 then $\overline{A} = 0$
13. $\overline{\overline{A}} = A$

} Laws of complementation (NOT Laws)

14. $A + B = B + A$
15. $A . B = B . A$

} Commutative Laws

16. $A + (B + C) = (A + B) + C$
17. $A .(B . C) = (A . B) . C$

} Associative Laws

18. $A . (B + C) = (A + B) . C$
19. $A + B . C = (A + B).(A + C)$

} Distributive Laws

20. $A + \overline{A} B = A + B$
21. $A + AB = A$
22. $A (A + B) = A$

23. $A(\overline{A} + B) = AB$
24. $AB + \overline{A}B = A$
25. $(A + B)(A + \overline{B}) = A$
26. $AB + \overline{A}C = (A + C)(\overline{A} + B)$
27. $(A + B)(\overline{A} + C) = AC + \overline{A}B$
28. $AB + \overline{A}C + BC = AB + \overline{A}C$
29. $(A + B)(\overline{A} + C).(B + C) = (A + B)(\overline{A} + C)$
30. $\overline{A + B + C + \text{- -}} = \overline{A}.\overline{B}.\overline{C}.$ ----
31. $\overline{A.B.C.\text{- -}} = \overline{A} + \overline{B} + \overline{C} +$ ----

} De Morgan's Laws

9.3 Basic logic gates:

Basically, a logic gate is a circuit with one or more logic input signals (each input is either 0 or 1), but with only one output signal (logically related to inputs). Logic circuits are analyzed with the help of Boolean laws. The basic logic gates are: OR, AND and NOT

9.3.1 OR Gate:

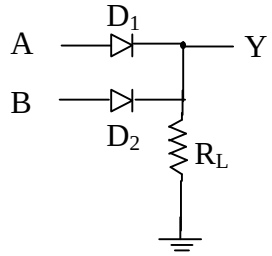
An OR gate has two or more input signals but only one output signal. If one or more input signals are high, the output signal is high.

Boolean expression for two input OR gate is Y = A + B

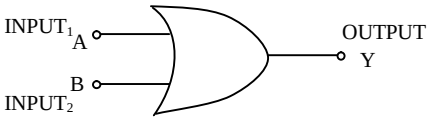
(read as Y equals A OR B).

The circuit to implement the OR function, logic symbol and its truth table(which depicts the output condition to given input is shown in fig 9.3(a).

Any one or both of the inputs A,B are high (+5V) makes the corresponding diode to forward bias and the current flows through load resistor R_L to have an output at Y.



Circuit of OR gate



Logic Symbol of OR gate

Fig 9.3(a)

TURTH TABLE

Input		Output
A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

An OR gate can have any number of inputs 1 to n. For example two input OR-gates can be used to form a 3 input OR gate as shown in fig 9.3(b)

.

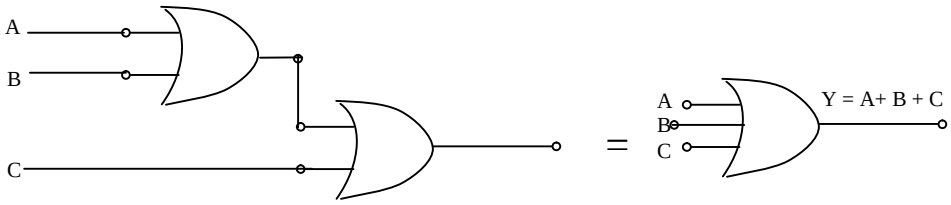


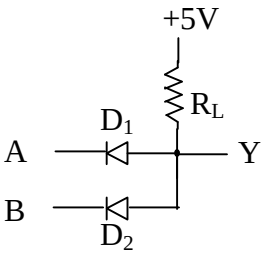
Fig 3b: Two two input OR gates connected to form a 3 input OR gate.

The OR gate circuit of Fig 9.3(a) constructed by using diodes and resistors . But the same OR gate can also be constructed by using transistors only (TTL circuit). The logic gates are available in the form of integrated circuits(ICs) to offer advantages in terms of size, cost and power. The
IC 7432 is a quad (four), two input OR gate using TTL logic and 74LS32 and 74HS32 are the low power and high speed versions of the same OR gate.

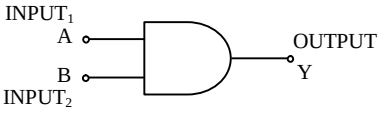
9.3.2 AND gate:

The AND gate has two or more inputs but has only one output. If all the inputs are simultaneously high, the output is high. Boolean expression for two input AND gate is

Y = A . B (read as Y equals A AND B). The circuit to implement AND function, corresponding logic diagram and its truth table is shown in Fig 9.3(c).



Circuit of AND gate



Logic Symbol of AND gate

Fig 9.3(c)

TURTH TABLE

Input		Output
A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

When any one or both of the inputs in above circuit are low, the diode of that input becomes forward biased, so the level of voltage at output point Y is at 0V i.e low state. On the other hand, if both A and B are high, both diodes are now in reverse biased having same voltages at both ends. So 5V now appears at Y, i.e high state.

As in the case of OR gate, the two input AND gates can be used to construct three input or n-input AND gate. Construction of 3-input AND gate by using two 2-input AND gates is shown in Fig 9.3(d).

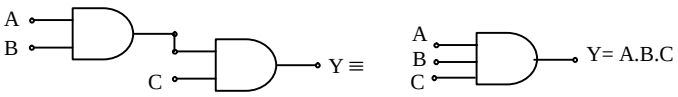
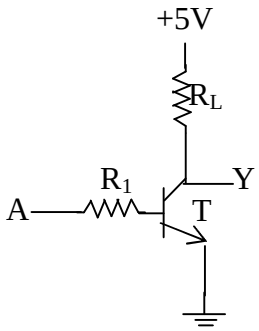


Fig 9.3(d)

IC 7408 / 74LS08 is a Quad two input AND gate whereas IC 7411 is a triple 3 input AND gate IC.

9.3.3 NOT gate: (Inverter gate):

The inverter is a gate with only one input and one output. The output state is always the opposite of the input state. It is also called as NOT gate. Boolean expression for this gate is $Y = \overline{A}$ (read as Y equals complement of A or Y equals NOT of A). The circuit, logic symbol and its truth table is shown in Fig 9.3(e).



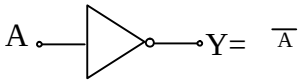


Fig 9.3(e)

TRUTH TABLE

Input A	Output Y
0	1
1	0

In the above circuit, the transistor is working either in the conduction or cut-off state. When a high state(+5V) is presented in input make to conduct transistor to produce 0V or low state at output. On the other hand, a low state at input A produces a high state at output Y by cut-off the transistor T. Input given to Two NOT gates connected in cascade results in the original Boolean variable as shown in fig 9.3(f). This is called as a buffer.

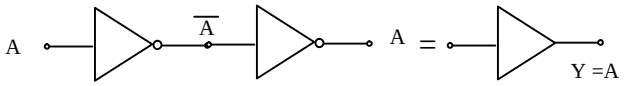


Fig 9.3(f) Two NOT gates connected to form a buffer, logic symbol of buffer and truth table

TRUTH TABLE

Input A	Output Y
0	0
1	1

Electronically buffers are very useful as they do not alter the nature of original signal but provide boosting of signal amplitude and power to standard levels. This prevents logic failure due to fall in signal strength. Tri-state buffers are also available whose output can be 0 or 1 or tri-state. In the tri-state condition the signal path is disconnected. These tri-state buffers are used in advanced logic circuits like Microprocessors and memories.

9.3.4 NOR gate:

The basic logic circuits are used to construct some more gates to perform more Boolean functions. For example, consider the NOR gate which has two or more inputs but only one output. All the inputs must be low to get a high output. Boolean expression for two input NOR gate is $Y = \overline{A + B}$.(read as Y equals not of A OR B).

The NOR gate can be constructed by connecting an OR gate with NOT gate as shown in Fig 9.3(g)

Fig 9.3(g)

TRUTH TABLE

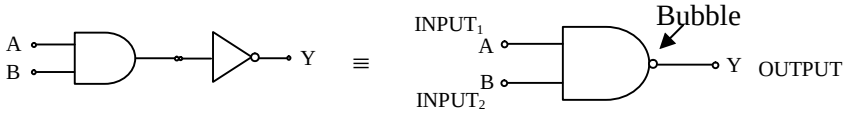
Input		Output
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

The NOR gate may have more than two inputs. Regardless of how many inputs a NOR gate has, it is still logically equivalent to one OR gate followed by an inverter. For instance, the equation for 3-input NOR gate is $Y= \overline{A+B+C}$. The 7402 is a quad 2-input NOR gate where as 7427 is a triple 3-input NOR gate.

9.3.5 NAND gate:

A NAND gate has two or more inputs but only one output. All the input are the simultaneously high to get a low output. Boolean expression for two – input NAND gate is

$Y = \overline{A.B}$ (read as Y equals Not of A AND B).



Logic Symbol of NAND gate

Fig 9.3(h)

TRUTH TABLE

Input		Output
A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

The Boolean equations for 3-input and 4-input NAND gates are $Y = \overline{ABC}$ and $Y = \overline{ABCD}$ respectively.

The IC 7400 have 4 Nos of 2- input NAND gates, whereas 7410 have three numbers of 3-input NAND gates.

9.3.6 EX – OR gate:

An OR gate recognizes words with one or more 1s, whereas the exclusive – OR gate recognizes only words that have an odd number of 1s. Boolean expression for two – input EX-OR gate is $Y = A\overline{B} + \overline{A}B = A \oplus B$. (read as Y equals A EX-OR B). The logic diagram, symbol and its truth table is shown in Fig 9.3(i).

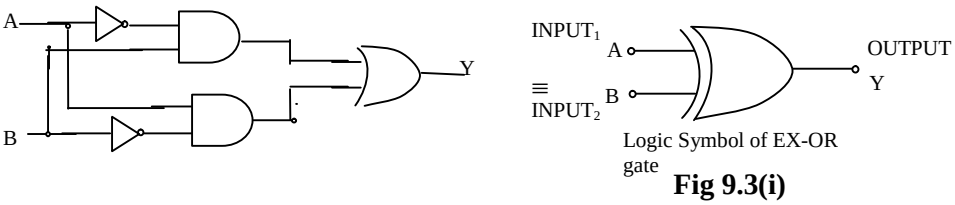


Fig 9.3(i)

TRUTH TABLE

Input		Output
A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

When both inputs A & B in the above circuit are high or low both AND gates have low outputs to have the final output zero. But when any one of the input(A or B) is high, the corresponding AND gate output is high. So the final output is one as shown in truth table. IC 7486 is the IC version of EX – OR gate.

9.4 De Morgan's theorems:

The Boolean algebra developed by George Boole was further extended by De Morgan by his famous theorems that can be used in the simplification of logic circuits.

9.4.1 De Morgan's First Theorem:

$$\overline{A + B} = \overline{A}.\overline{B}$$

Statement: Complement of sum equals the product of the complements.

Proof: The theorem can be proved by taking expressions and writing the truth tables on both sides. And finally we equate LHS to RHS.

Consider L.H.S i.e $\overline{A+B}$. It indicates a 2-input OR gate followed by a NOT gate that is a NOR gate as shown in Fig9.4(a). Fig 9.4(a) also shows its truth table. Similarly take the R.H.S of the theorem which is equal to $\overline{A}.\overline{B}$. This equation indicates that the two inputs A and B are inverted before they reach the AND gate as shown in Fig 9.4(b). Therefore we have the truth table as in Fig 9.4(b) for this circuit. Therefore the comparison of the truth table of Fig 9.4(a) with the truth table of Fig9.4(b) reveals that these two are equivalent; the circuits of Fig 9.4(a) and Fig 9.4(b) are equivalent. Hence the theorem is proved.

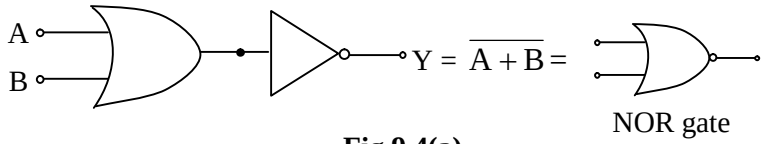


Fig 9.4(a)

A	B	A+B	$\overline{A+B}$
0	0	0	1
0	1	1	0
1	0	1	0
1	1	1	0

R.H.S = $\overline{A}.\overline{B}$

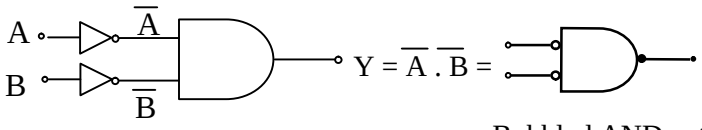


Fig 9.4(b)

A	B	$\overline{A}$	$\overline{B}$	$\overline{A}.\overline{B}$
0	0	1	1	1
0	1	1	0	0
1	0	0	1	0
1	1	0	0	0

9.4.2 De Morgan's Second Theorem

$\overline{A \cdot B} = \overline{A} + \overline{B}$

Statement: Complement of product equals the sum of the complements.

Proof: Left hand side of the theorem shown is a 2-input AND gate followed by a NOT gate i.e. a NAND gate having the circuit and truth table as shown in Fig 9.4(c).

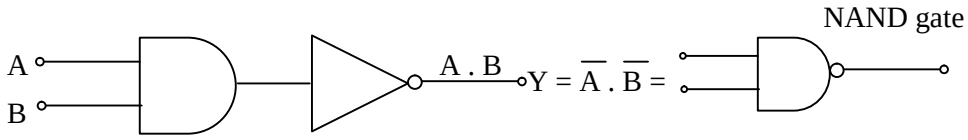


Fig 9.4(c)

A	B	A . B	$\overline{A \cdot B}$
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	0

The right hand side of the equation shows an OR gate with two inverted inputs. The circuit and truth table is shown in Fig 9.4(d).

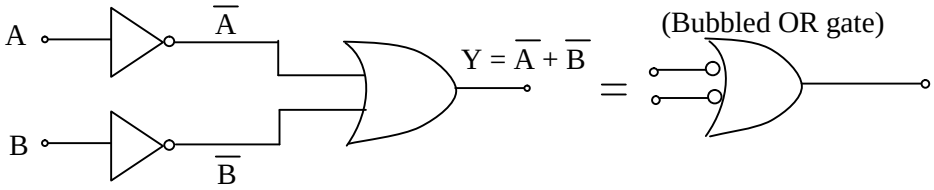


Fig 9.4(d)

A	B	$\overline{A}$	$\overline{B}$	$\overline{A} + \overline{B}$
0	0	1	1	1
0	1	1	0	1
1	0	0	1	1
1	1	0	0	0

Comparison of these two truth tables reveals that they too are identical and equal. Therefore the theorem is proved.

9.5 Universal logic gates:

The logic gates NAND and NOR are called Universal Building Blocks or Universal Logic Gates because using either NAND or NOR gates we can construct other logic gates (AND, OR, NOT, EX-OR) as well as we can implement Boolean functions.

1

9.5.1 Forming other logic gates using only NAND gates:

(i) NOT gate: From the truth table of the NAND gate given in Fig 9.3(h), we know that the same inputs of the NAND gives its complement output viz if A= B= 0 we have an output of 1 and if A= B= 1, the output is 0. Hence the circuit of Fig 9.5(a) acts as NOT gate.

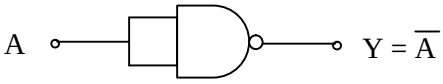


Fig 9.5(a)

ii) AND gate: The Boolean expression for NAND is Y= $\overline{A \cdot B}$. So, the first NAND gate output in the Fig 9.5(b) is $\overline{A \cdot B}$ and second NAND gate simply a NOT gate. Now the circuit is simply an AND gate circuit.

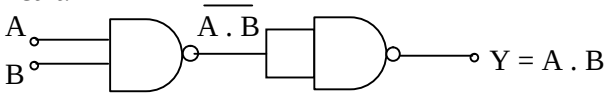


Fig 9.5(b)

(iii) OR gate: Two NOT gates of Fig 9.5(c) gives output of $\overline{A}$ and $\overline{B}$, which are the inputs of second NAND gate. Hence, second NAND gives an output of $\overline{\overline{A}.\overline{B}}$. According to second De Morgan's theorem it can be written as $\overline{\overline{A}} + \overline{\overline{B}}$ which can also be written as A+B($\overline{\overline{A}}$ =A and $\overline{\overline{B}}$ = B , refer fig 9.3(c). It is the output of an OR gate.

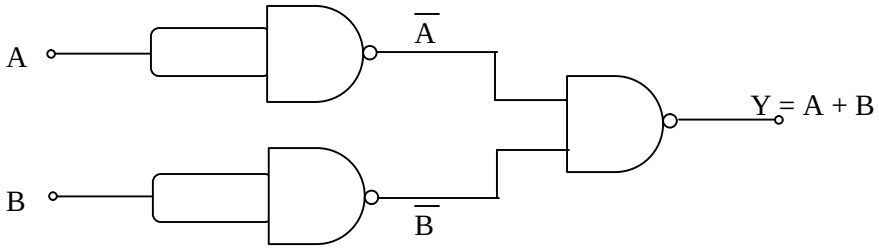


Fig 9.5(c)

(iv) NOR gate:

Addition of another NOT gate to the circuit for OR gate of Fig 9.5(c) as shown in Fig 9.5(d) gives the operation of a NOR gate.

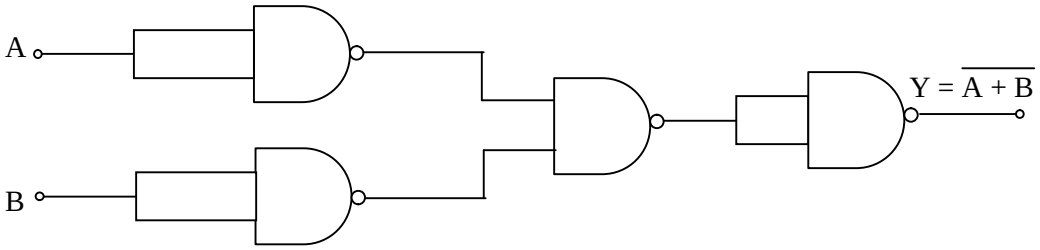


Fig 9.5(d)

(v) EX – OR gate:

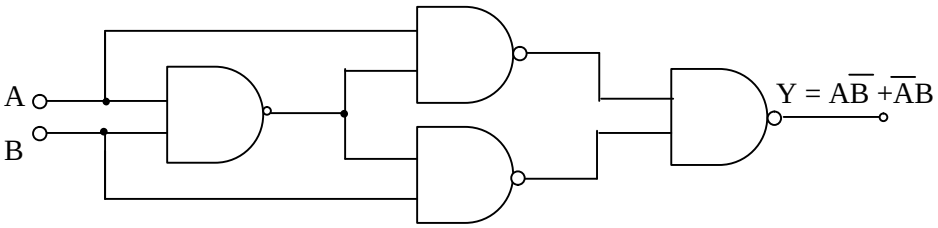


Fig 9.5(e)

9.5.2 Other Logic gates using only NOR gates :

(i) NOT gate :

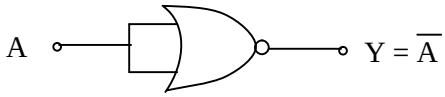


Fig 9.5(f)

(ii) OR gate:

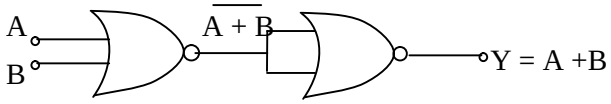


Fig 9.5(g)

(iii) AND gate:

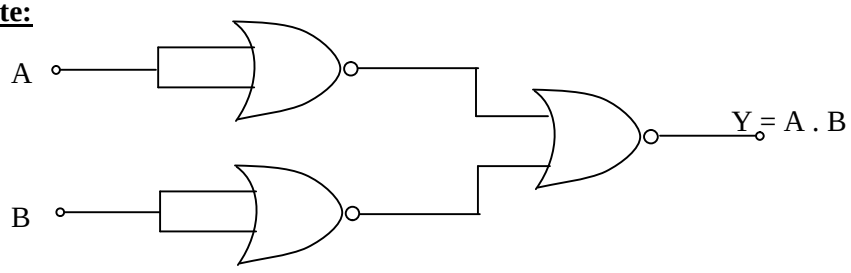


Fig 9.5(h)

(iii) NAND gate:

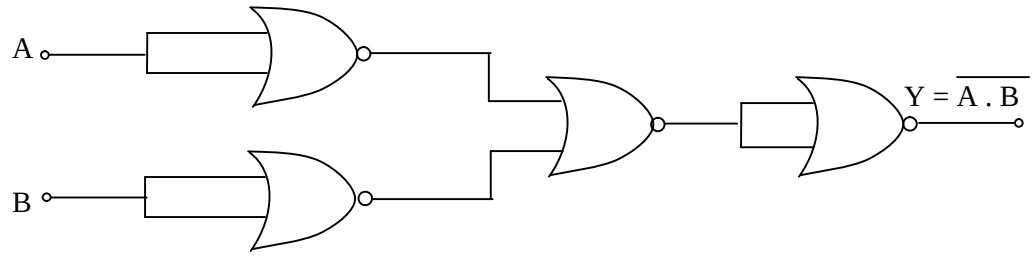


Fig 9.5(i)

(v) EX - OR gate:

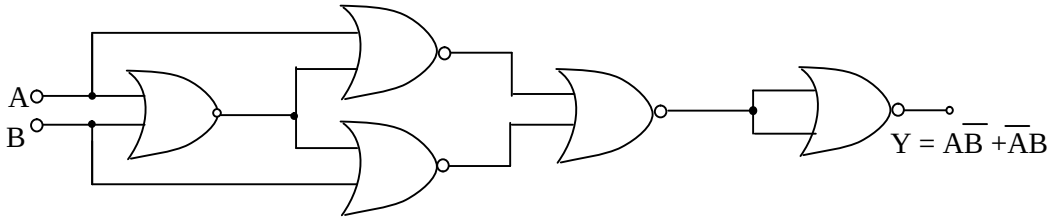


Fig 9.5(j)

Of the several technologies available Transistor–Transistor Logic is widely used where logic 1 is represented by +5V DC. And Logic zero by 0V DC. Circuits using other types of logic are provided with TTL logic compatibility to facilitate easy interfacing of logic circuits.

When a logic gate is connected to another logic gate, depending on the logic status it is supposed to supply or take currents. Supplying current is called sourcing and taking the current is called sinking. Capability of a circuit depends on how much current a circuit can sink and source.

This is expressed in terms of how many logic gates can be connected to it (fan-in) and how many circuits can be driven by it (fan-out). For standard TTL fan – out is 10.

Logic circuits are supposed to change their states from 0 to 1 or 1 to 0 instantaneously, but, electronic circuits have inherent delays and because of it, there will be some definite rise time, fall time and propagation delay as a signal passes from input to output. Further there will be signal attenuation and induction effects and noise, which affect signal amplitude. For a standard TTL circuit if a logic 1 signal amplitude falls below 2.4V DC it cannot be recognized as Logic1. Like wise, if logic 0 signal amplitude is greater than 0.4V DC it cannot be identified as logic 0. This results in failure of logic circuits. Crossing the fan-in and fan-out limits also changes logic voltages. So logic circuits are to be buffered to bring back, the signal levels to TTL levels

before they are deteriated. A signal in its transmission path may develop glitches, slopes and other types of distortion. Schmitt trigger circuit is used to produce rectangular wave shape regardless of the input waveform.

9.6 Summary:

Digital electronics mainly deals with binary number system. Binary number system has a base of 2. It has two numerals 0 and 1. Any big number can be expressed in the form of 0s and 1s. In a digital system, logic 1 means high positive signal value, logic 0 means low positive signal value in a positive logic system. But in a negative logic system, zero value represents logic 1 and a negative value represents logic '0'. Digital electronics mainly depends on Boolean algebra and logic gates. Using the gates, one can construct bit circuits that can perform complex operations. These gates possess, in general, 2 or more inputs and only one output. NOT, OR and AND are the basic logics and circuits used to implement higher logic. However semiconductor technology prefers NAND and NOR logic gates. Using only NAND or using only NOR gates other logic gates can be constructed. Because of this reason, NAND and NOR are called Universal logic gates.

De Morgan's first theorem says that a NOR gate and bubbled input AND gate are equivalent. Similarly De Morgan's second theorem says that a NAND gate and bubbled input OR gate are equivalent. If Boolean expressions are given, one can construct logic circuits. On the other hand if logic circuits are given, one can write the Boolean expressions.

Logic gates can be fabricated using various Semiconductor technologies like TTL, NMOS, and CMOS etc. Each technology has its own advantages.

9.7 Key Terminology:

(i) Gate: An electronic circuit with one or more inputs but with only one output.

(ii) Logic Gate: As the gates simulate mental processes, these are often called Logic Gates.

(iii) Universal Gate: The gate with which other logic circuits / gates can be constructed (Ex: NAND, NOR)

(iv) Boolean Algebra: The modern algebra that uses the set of numbers 0 and 1.

(v) Complement: The output of an inverter gate.

(vi) Truth table: A table that shows all input and output possibilities for a logic circuit.

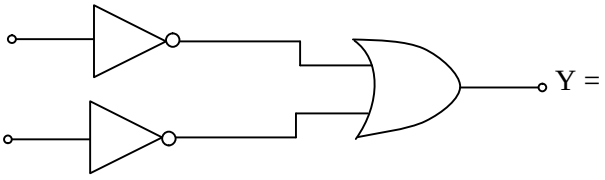
(vii) Word: A string of bits that represent a coded instruction or data.

(viii) Fan – in: The maximum number of inputs that can be applied to a logic gate.

(ix) Fan – out: The number of gates that can be driven by a logic gate.

9.8 Self – assessment questions:

1. Prove the De Morgan's Theorems.
2. Draw the truth table of 3 – input, EX – OR gate.
3. Draw the truth table of 4 – input NOR gate.
4. What is the Boolean equation of the figure given below?



5. Draw the logic circuits of the Boolean expressions given below.

(i) $Y = A \cdot B + B \cdot C + \overline{A \cdot C}$

(ii) $Y = A \cdot B \cdot C + \overline{A \cdot B} + \overline{B \cdot C}$

(iii) Y = $\overline{(A + C + B)} \cdot \overline{(ACB)}$

6. Apply De Morgan's Laws to the following expressions.

$\overline{(A + B)}.\overline{B}$, $\overline{(A + B)}.\overline{C}$, $\overline{\left(\overline{A \cdot B} + \overline{C \cdot D}\right)}$

9.9 Reference books:

1. Principles of Digital Electronics – Malvino & Leach (Tata - McGraw Hill Publishers)
2. Digital Electronics – William H. Gothman. (P.H.I publishers)
3. Digital Computer Electronics – Albert Paul Malvino. (Tata - McGraw Hill publishers)

Name of the Author

Dr B.V.S Goud,

Associate Professor and Head,

Department of Electronics and Instrumentation Technology

Acharya Nagarjuna University

Nagarjuna Nagar.

UNIT III
Lesson : 10

Combinational Logic Circuits

Objectives:

(1) To understand binary addition

(2) To understand the working of various combinational logic circuits like adders, decoders, de-multiplexers, data selectors, multiplexers .

Structure:

Introduction:

Basic logic gates are the building blocks of logic systems like digital computers. In digital computers arithmetical operations are performed using binary number system. Further any arithmetic operations like subtraction, multiplication and division can be expressed in the terms of binary addition. In this lesson we learn basic concepts of binary additions using logic gates. Varies logic gates are combined to generate a complex logical operation and if these circuits are made to work independent of time. Such circuits are called simply combinational logic circuits. If these logic circuits are made to work with time as a parameter these circuits are called sequential logic circuits. A general logic circuits may be combinational and sequential logic circuits. In this lesson learn about some combinational logic circuits.

Binary codes are used to specify instructions, addresses and coded data. One binary digit (bit) can stand for two codes 0 or 1. with two bits we can have four codes namely 00, 01, 10, 11. Similarly with n bits we can form 2^n codes. Circuits, that generate these codes are called encoders. It become necessary not only the code information but also to decode the codes whenever necessary. Circuits with decode are called decoders.

Information coming on several channels may have to be rooted on a single channel this is called multiplexing and circuits with do this multiplexers. Multiplexer information coming on a single channel have to be separated and sent and varies channels and this is called demultiplexers. In this lesson we learn about these combinational logic circuits also.

10.1 Binary addition:

When we perform binary addition and subtraction, the following rules must be followed.

Binary addition	Binary subtraction
$0 + 0 = 0 \rightarrow$ (1)	0 – 0 = 0 borrow 0
$0 + 1 = 1 \rightarrow$ (2)	0 – 1 = 1 borrow 1
$1 + 0 = 1 \rightarrow$ (3)	1 – 0 = 1 borrow 0
$1 + 1 = 10 \rightarrow$ (4)	1 – 1 = 0 borrow 0
$1 + 1 + 1 = 11 \rightarrow$ (5)	

In binary addition the first three operations produce a sum with only one digit, but when both augend and added bits are equal to 1, the binary sum consists of two digits. The higher significant bit of this result is called a carry.

10.2 Half Adder:

A combinational circuit that performs the addition of two – bits is called a Half – adder. The Half – adder has two binary inputs and two binary outputs.

This is 1 – bit adder. This circuit is called half – adder because it cannot accept a carry – in from previous additions. The inputs to the circuits are the addend and augend bits. The outputs produced by it are sum (s) and carry (c). The half adder can be constructed by using one AND gate and an EX – OR gate. Fig 10.1(a) shows logic circuit that adds 2 bits namely A and B.

Fig 10.1(b) shows the truth table of it. The carry output is 0 unless both inputs are 1. The S output represents the least significant bit of the sum.

Boolean equation for sum (s) = $A \oplus B = A\overline{B} + \bar{A}B$.

= A EX – OR B

Carry (c) = A AND B = AB

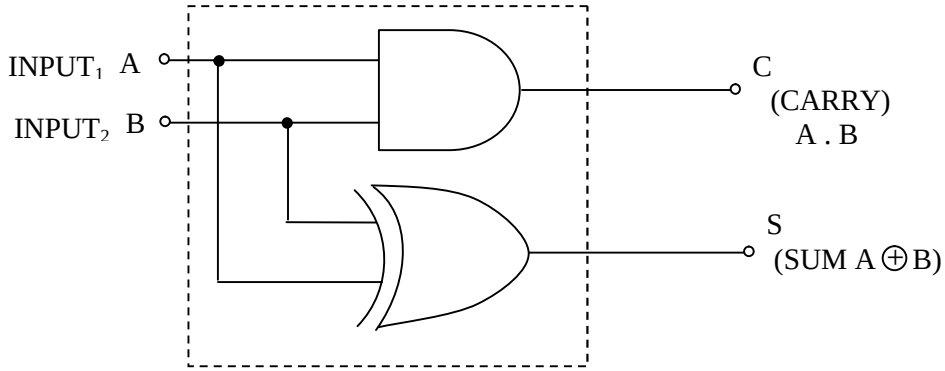


Fig 10.1(a) Half-Adder Logic circuit

Fig 10.1 (b) Half – adder Logic symbol

Half – adder Truth table

A	B	Carry C	Sum S
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

Fig 10.1 (c)

10.3 Full adder:

A full adder is combinational circuit that forms the arithmetic sum of three input bits. The two significant bits to be added are denoted as A and B, where as the third input C represents the carry from the previous lower significant position. The full adder can be constructed from two half adders and one OR gate as shown in fig 10.2(a). and its symbol is shown in fig 10.2(b).

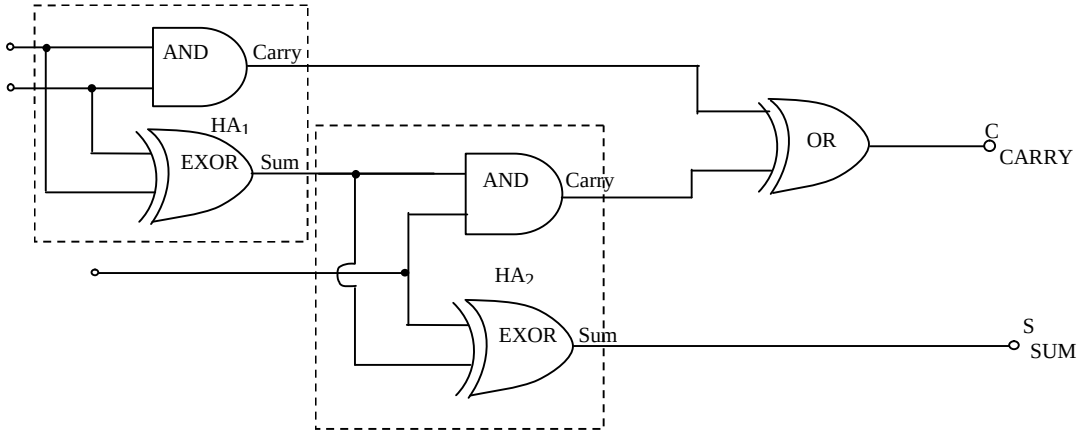


Fig 10.2(a) : Full – adder Logic diagram

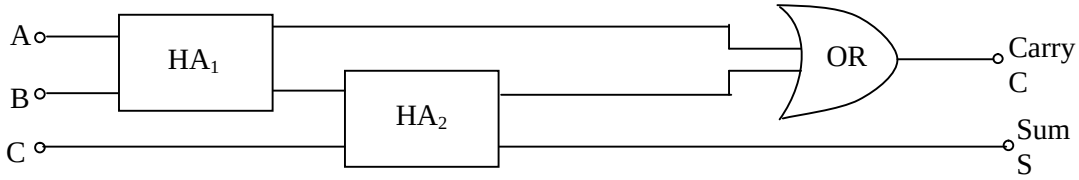


Fig 10.2 (b): Full – adder Logic Symbol

INPUTS			OUTPUTS	
A	B	C	Carry	Sum
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

Fig 10.2 (c) Full – adder Truth Table

10.4 Binary four bit parallel Adder:

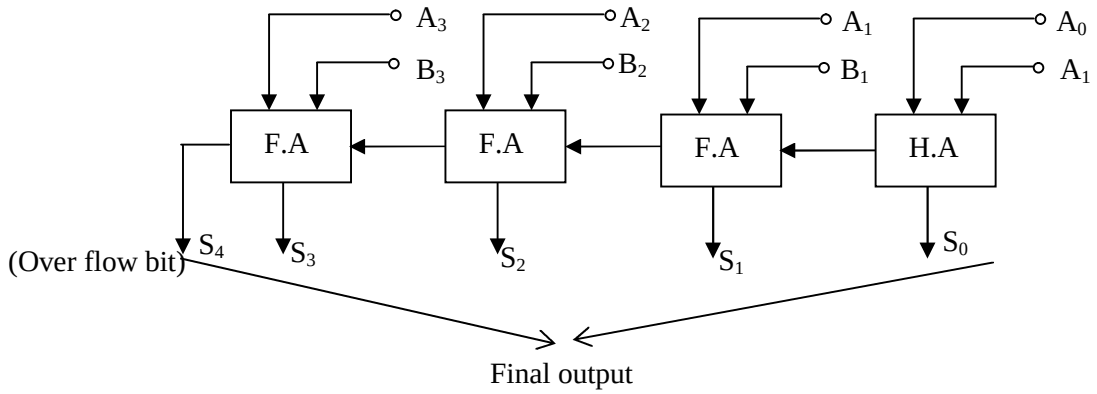


Fig 10.3

Note: In binary, the count starts from zero rather than from 1. The first bit is called S_0 rather than S_1. Similarly the nyj bit is S_{n-1}.

A parallel adder is an m-bi at a time. This produces the arithmetic sum of two n-bit binary numbers in parallel. This can be constructed by using one half adder and several full adders. The full adders are connected in cascade, with the output carry from one full adder connected to the input carry of the next full adder.

Fig 10.3 shows the logic diagram for 4-bit parallel adder circuit.

Suppose we want to add 4-bit binary numbers $A_3\ A_2\ A_1\ A_0$ and $B_3\ B_2\ B_1\ B_0$ then we get a sum $S_4\ S_3\ S_2\ S_1\ S_0$ where S_4 indicates overflow bit if the sum exceeds four bits. For adding the above two 4-bit numbers, we require three full adders and a half adder connected in parallel. The output (carry) of each adder is connected to next adder to get a parallel adder.

10.5 Decoder:

A decoder is a combinational circuit that converts binary information from 'n' input lines to a maximum of 2^n unique output lines. Decoders are available with several output configurations: active low voltage, high-sink current for direct driving of indicator lamps. Output voltage ratings range from +5V to over +100V. the decoder is used in conjunction with some code converters such as a BCD – to – 7 segment decoder, BCD – to – Decimal decoder. The decoder presented here is called n – to 2^n line decoders decoder. Its purpose is to generate the 2^n min terms of n input variables. These decoders form a combinational circuit with n input variables and 2^n output variables. For each binary input combination of 1s and 0s, there is one and only one output line that assumes the value of 1. Figure 10.4(a) shows a 2 by 4 decoder. It has four AND gates and two inverters.

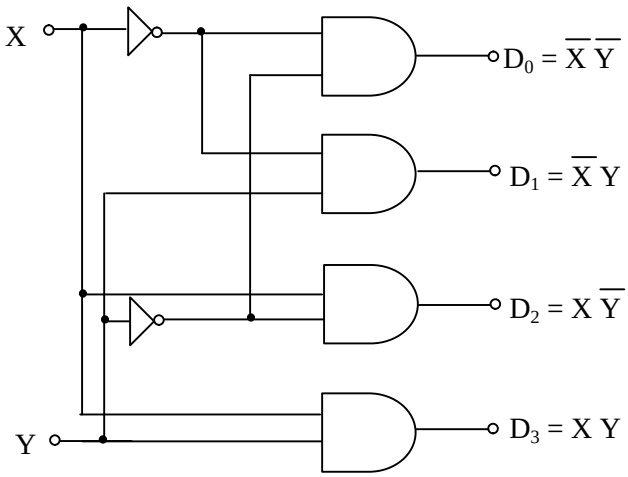


FIG 10.4(a) 2 – bit decoder

X	Y	D_0	D_1	D_2	D_3
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0
1	1	0	0	0	1

Fig 10.4(b) 2- bit decoder Truth table

Some times an enable input may be included with a decoder to control the circuit operation. In this, all outputs will be equal to 0, if the enable input is zero. When the enable input is 1, the circuit operates as a conventional decoder. Block diagram of 3 to 8 decoder with enable signal is shown in Fig 10.5.

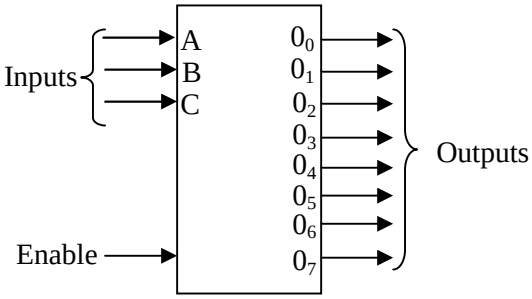


Fig 10.5 Block diagram of 3 to 8 bit decoder

10.6 Demultiplexer:

A demultiplexer is a circuit that receives information on a single line and transmits this information on one of 2^n possible output lines. The selection of a specific output line is controlled by the bit values of n selection lines. Figure 10.6 shows the block diagram of decoder and demultiplixer. The decoder with an enable input can function as a demultiplexer.

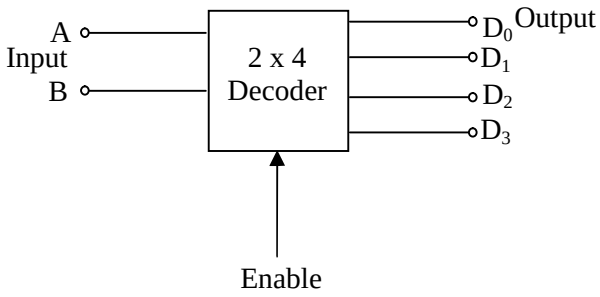


Fig 10.6(a)

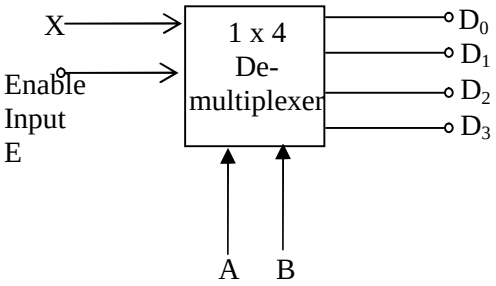


Fig 10.6(b)

The decoder of Fig 10.6(a) can function as a demultiplexer if the enable line is taken as a data input line and lines A and B are taken as the selection lines as shown in Fig 10.6(b). Out of 4 output lines, one gets link with input E depends upon the binary value of two selection lines A and B (i.e), if AB = 01, D_1 gets connected with the input E, so that input is available at D_1 output. While all other outputs are maintained at 1.

10.7 Encoder:

An encoder is a digital circuit, that produces a reverse operation from that of a decoder. An encoder has 2^n input lines and n output lines. The output lines generate the binary code for the 2^n input variables. One of

the encoders is shown in fig 10.7(a). It has eight inputs, one for each of the eight digits and three outputs that generate the corresponding binary outputs that generate the corresponding binary number.

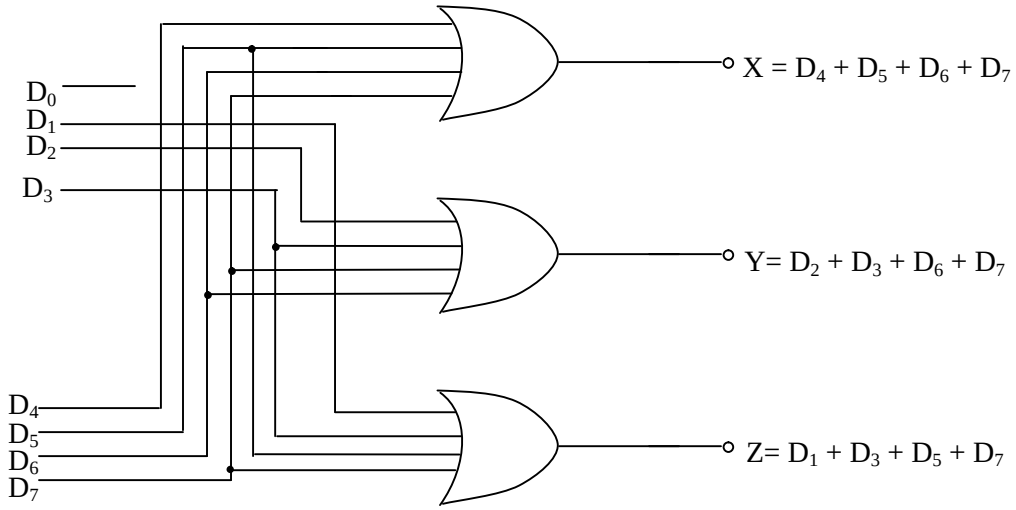


Fig. 10.7(a)

Fig 10.7(a) shows the octal to binary encoder. It is constructed with OR gates. It, has 8 input lines and could have 2^8 = 256 possible input combinations. Only eight of these combinations have meaning and others are don't care conditions. The output z is 1, if the input digits are odd. Output Y is 1 for octal digits 2, 3, 6 and 7. Output X is 1 for digits 4, 5, 6 or 7. It is 1 at any time. Do is not connected to any OR gate. It erises that only the highest priority input line is encoded.

INPUTS								OUTPUTS		
D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7	X	Y	Z
1	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0	0	1	1
0	0	0	0	1	0	0	0	1	0	0
0	0	0	0	0	1	0	0	1	0	1
0	0	0	0	0	0	1	0	1	1	0
0	0	0	0	0	0	0	1	1	1	1

Truth table 10.7(b)

10.8 Multiplexers / Data selector:

It is a combinational circuit which selects binary information from one of many input lines and directs it to single outputs line. The selection of a particular input line is controlled by a set of selection lines. For 'n' bits, there are 2^n input lines and n selection lines whose bit combinations determine which inputs is selected. This means "many into one". It is used when a complex logic circuit is to be shaved by a number of input signals. Fig 10.8 shows the logic diagram, block diagram and function table of a 4 to 1 line multiplexer.

A line to 1 line multiplexer has 4 inputs I_0 to I_3 and one output line. Each of the four input lines is applied to one input of a three input AND gate. The remaining two inputs of it is supplied by selection lines S_1 and S_0.

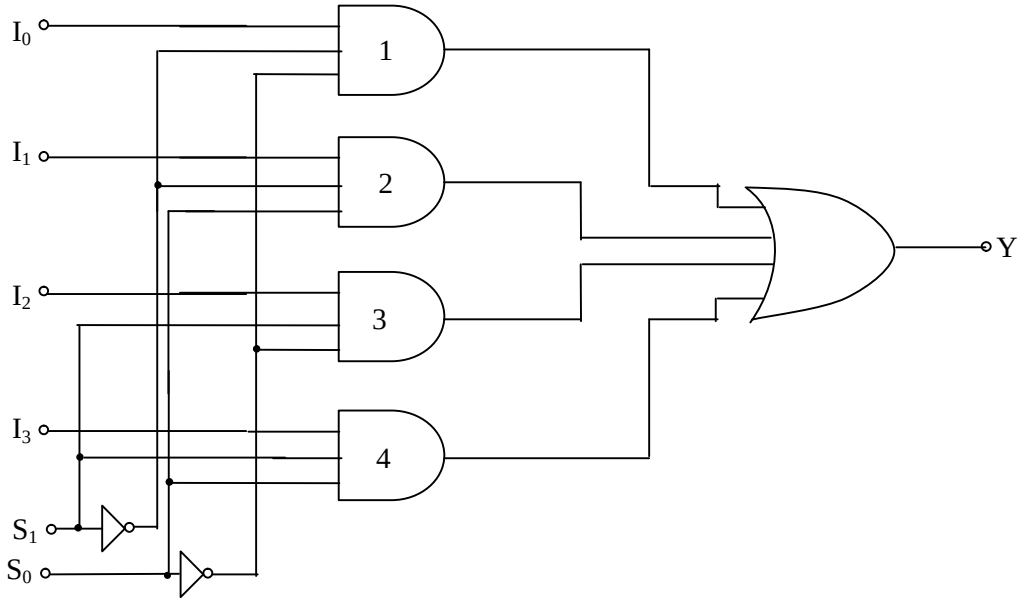
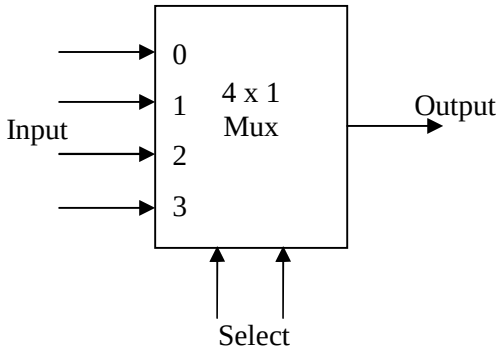


Fig 10.8(a): 4 to 1 Multiplexer

S_1	S_0	Y
0	0	I_0
0	1	I_1
1	0	I_2
1	1	I_3

Fig 10.8(b)



Fit 10.8(c)

Based on the combination of S_1 and S_0, the input associated with the selected AND gate finds out the path to reach the output through OR gate. For example: when $S_1S_0 = 01$, it selects the AND gate 2. So the input I_1 is transferred to the output i.e OR gate output is now equal to the value of I_1. thus providing a path from the selected input to the output. The remaining AND gates have at least one input equal to 0.

As in decoders, multiplexer IC have an enable input (or) storbe input to control the operation of the unit. It can also be used to expand two or more multiplexers to a digital multiplexer with a large number of inputs.

10.9 Summary:

Logic circuits can be either combinational or sequential. The five basic gates (OR, AND, NOT, NAND, NOR) and their combinations – half, full and parallel adders are known as combinational logic circuits. Here there is no feedback. These have no memory elements. Their output only depends upon their present input. Thus can operate as fast as the devices of which they are made. Half and full adders are bit adders, where as parallel adder is an adder of two n – bit word. The function performed by a multiplexer is to select 1 out of N input data sources and to transmit the selected data to a single information channel. A multiplexer performs the increase process of a demultiplexer. A decoder is a system which 1 on one (and only one) of 2^m output lines. In other words, a decoder recognizes a particular code, A decodes with an enable input can function as a demultiplexer. An encoder produces reverse operation from that of a decoder. The combinational logic circuits are mainly employed in ALU_S.

10.10 Key terminology:

Binary Adder: A logic circuit that can add two binary numbers.

Half Adder: A logic circuit that adds 2 bits.

Full Adder: **A** logic circuit than can add 3 bits.

Over flow: As the sum or difference that of lies outside the normal range.

ALU: ALU stands for arithmetic logic unit. It performs arithmetic one logic operations.

Encoding: The process of generating binary codes.

Decoding: The reverse process of encoding.

Demultiplexer: A combinational logic circuit, that accepts a single input and sends it to 1 out of N output lines which is selected by select lines.

10.11 Self – assessment questions:

1) Explain is a binary addition. Give examples.
2) What is a full adder? How it adds 3 bits?

3) Draw a full adder circuit, without using half adders.
4) Construct a 5-bit parallel adder.
5) Implement a full adder circuit with multiplexers.

10.12 Reference Books:

1. Integrated Electronics – Millman and Halkias
2. principles of digital Electronics – Malvino and Leach.

Name of the author:

Dr. B.V.S.Goud, M.Sc., Ph.D.,
Associate Professor and Head,
Dept. of Electronics and Instrumentation Technology
Acharya Nagarjuna University,
Nagarjuna Nagar – 522 510

Unit – III

LESSON – 11

Sequential Logic Circuits

(Flip – Flops, Registers and counters)

Objectives:

This lesson explains you the concept of

1) Flip – Flops
2) Different types of Flip Flops
3) Shift Registers
4) Synchronous and asynchronous counters

Structure of the lesson

INTRODUCTION:

The logic circuits whose present output depends on the past inputs are known as sequential logic circuits. These circuits store and remember information. The sequential circuits unlike combinational circuits are time dependent. Normally the current output of a sequential circuit depends on the previous state of the circuit and on the current input to the circuit. It is a connection of flip-flops and gates. What is a flip-flop? You will find the answer in this section. There can be two types of sequential circuits.

- Synchronous
- Asynchronous

Synchronous circuits use flip-flops and their status can change only at discrete instants. The Asynchronous sequential circuits may be regarded as combinational circuits with feedback path. Since the propagation delays of output to input are small they may tend to become unstable at times.

The synchronization in sequential circuits can be achieved using a clock pulse generators. A clock synchronizes the effect of input over output. It presents a signal of the following form:

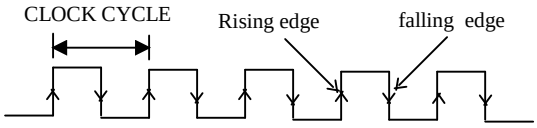


Fig 11.1 Clock signal of a Clock Pulse Generator

The signal produced by clock generator is in the form of clock pulse or clock signal. These clock pulses are distributed throughout the computer system for synchronization.

A clock can have two states, an enable or active state, otherwise a disable or inactive state. Both of these states can be related to zero or one levels of clock signals. Normally, the flip-flops change their state only at the active state of the clock pulse. In certain designs the active state of the clock is triggered, when transition from 0 to 1 or 1 to 0 is taking place in clock signal.

11.1 Flip-flop

A flip-flop is a binary cell, which can store a bit of information. It itself is a sequential circuit. A flip-flop maintains any one of the two stable states that can be treated as 1 or 0 depending on the presence or absence of output signal. The state of flip-flop can only change when clock pulse has arrived. Let us first see the basic flip – flop or a latch .

11.2 S-R Latch:-

The most fundamental circuit in the group of flip-flops is the S-R(set-reset) latch . The basic S-R latch has two NAND gates connected back – to – back. Fig 11.2 shows, such an S-R latch. The truth table of the S-R latch is given in Table 11.1. It describes the conditions of the flip-flops at times t_0 and t_1. The state at t = t_0 is called the present state (PS) and the state at t = t_1 is called the next state (NS).

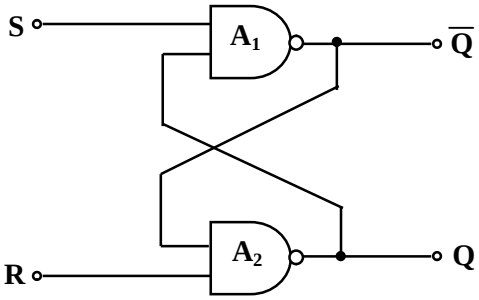


Fig 11.2 Circuit diagram of the S – R latch.

The S – R latch has two external inputs S and R, and two feed – back inputs Q and $\overline{Q}$. The present states of the inputs are S_0, R_0 and Q_0. After one clock pulse, the output will change to the next state Q_1. This changed value of the output will be the new present state of the input.

For testing the validity of the circuit, we have to test eight states corresponding to the three inputs. So, R_0 and Q_0. These states are tested by using fig 11.2 and table 11.1

Table 11.1 Truth Table of S – R latch

Present state (PS)			Next state (NS)
S_0	R_0	Q_0	Q_1
0	0	0	Not permitted
		1	
0	1	0	Reset to 0
		1	
1	0	0	Set to 1
		1	
1	1	0	0 } Q_0
		1	1 }

We shall test the outputs for the conditions: S = R= 0; S=0, R=1; S=R=1. In each of these cases, we test the state of the latch for the conditions of Q = 0, and Q = 1 respectively.

Case(i): We find that when S = R = 0 both Q and $\overline{Q}$ become 1. But this is not allowed because we want Q and $\overline{Q}$ to be complementary outputs. This state is designated as the forbidden state.

Case (ii): In S = 0, R = 1 case, when S = 0, output of NAND gate A_2, i.e., $\overline{Q}$ becomes 1. Now the inputs to gate A_1. i.e., R and $\overline{Q}$, both become 1, which make its output, Q = 0. This is called the reset (to 0) state.

Case (iii): In S = 1, R = 0 case, when R = 0, A_1 produces an output of 1 i.e., Q = 1. Now, both the inputs to A_2 are 1, which makes $\overline{Q}$ = 0. This state (S = 1) making Q equal to 1 is called the set (to 1) state.

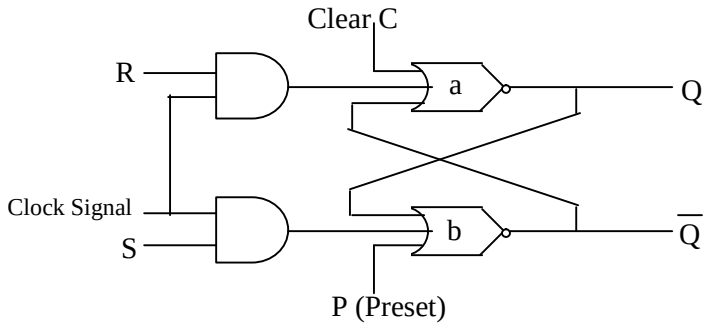
Case (iv): In S = 1, R = 1 case, both the inputs are equal to 1 . It will not produce any change in the output. Hence, for these cases, the outputs will remain as such in their previous states. This means that the present state and the next state are one and the same.

Using fig 11.2, we have described the behavior of a bi-stable multi-vibrator using NAND gates for all possible input states. Table 11.1 represents the truth table of the S-R latch. The behavior of the circuit shows that it will remain in one of the two states (i.e. 0 or 1) till an externally applied trigger pulse produces a transition in the output from 0 to 1, or 1 to 0. Thus, the circuit is able to store one bit of information in its memory. Therefore, this flip-flop is called a memory cell, we shall now discuss the clocked S-R Flip-Flop.

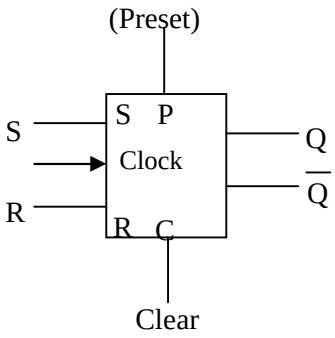
11.3 Clocked R-S Flip – Flop:

The main feature in R-S Flip-Flops is the addition of a clock pulse input. In this flip-flop change in the value of R or S will change the state of the flip-flop only if the clock pulse at that moment is one. It is shown in fig 11.3a.

The excitation or characteristic table basically represents the effect of S and R inputs on the state of the flip-flop, irrespective of the current state of flip-flop. The other two inputs P(present) and C(clear) are asynchronous inputs and can be used to set the flip-flop or clear the flip-flop respectively at the start of operation, independent of the clock pulse.



11.3(a) Logic Diagram



11.3(b) Symbolic representation

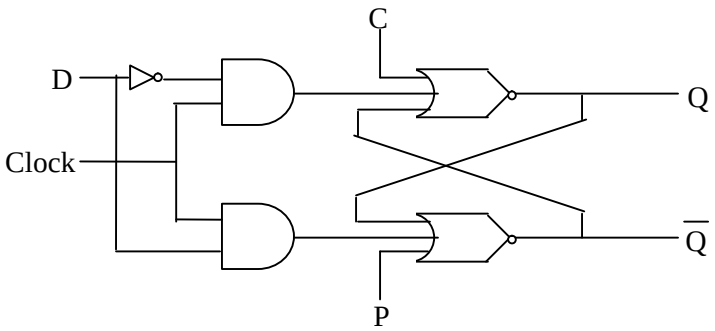
Input S	Input R	State on completion of clock cycle
0	0	No change in State
0	1	Clear the flip – flop (state 0)
1	0	Set the flip – flop (state 1)
1	1	Undesirable (Not allowed)

Fig 11.3(c) Characteristic table.

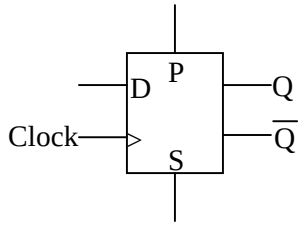
Fig 11.3: R – S flip – flop

11.4 D Flip – Flop:

D flip-flop is a special type flip-flop and it represents the currently applied input as the state of the flip-flop. It can store 1 bit of data information and sometimes refers to as Data flip-flop. The state of flip-flop changes with the applied input. It does not have a condition where the state does not change as the case in RS flip-flop, the state of R-S flip-flop does not change when S = 0 and R = 0. If we want that for a particular input state does not change, then either the clock is to be disabled during that period or a feedback of the output can be embedded with the D- flip-flop is also known as Delay Flip-Flop because it delays the 0 or 1 applied to its input by a single clock pulse.



11.4(a) Logic Diagram



11.4(b) Symbolic representation

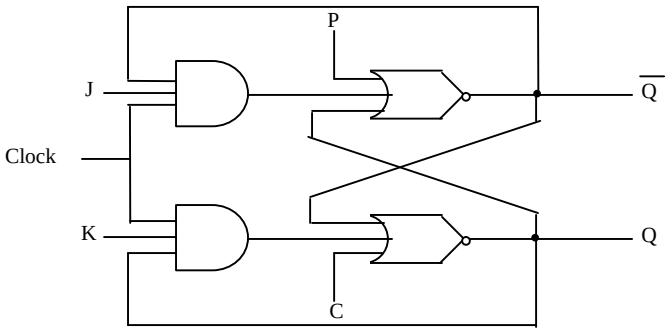
Input D	State on completion of clock cycle
0	0
1	1

Fig 11.4(c) Truth table of D-Flip-Flop.

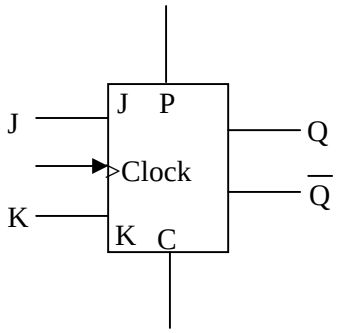
Fig 11.4: D flip – flop

11.5 J-K Flip – Flop:

It is not possible to achieve toggling (changing the state of the output wherever the input makes a transition from 1 to 0) with the simple flip-flops described above. By using NAND gates, one can construct a flip-flop that toggles. The symbol and truth table of one such circuit, called J-K flip-flop are shown in fig 11.5



11.5(a) Logic Diagram



11.5(b) Symbolic representation

Input J	Input K	State on completion of clock cycle
0	0	No change in State
0	1	Clear the flip –f lop (state 0)
1	0	Set the flip – flop (state 1)
1	1	Complement the state of flip - flop

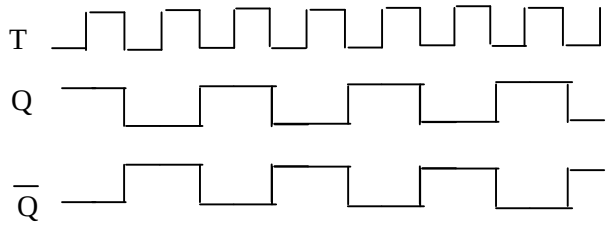
Fig 11.5(c) Truth table.

Fig 11.5: J – K flip – flop

A JK flip-flop is a refinement of the SR flip-flop in that the indeterminate condition of the SR type defines in the JK type. An examination of the truth table given in fig 11.5(a) reveals the following properties of the JK flip-flop.

a) It retains its present state if J = 0 and K = 0
b) If J = 0, K = 1, Q = 0, it resets to zero
c) If J = 1, K = 0, Q = 1, it sets
d) If J = 1, K = 1, it Toggles.

If the first clock pulse leaves Q = 1 and $\overline{Q}$ = 0 the second clock pulse make Q = 0 and $\overline{Q}$ = 0. The third clock pulse make Q = 1 and $\overline{Q}$ = 0 again and so on as shown

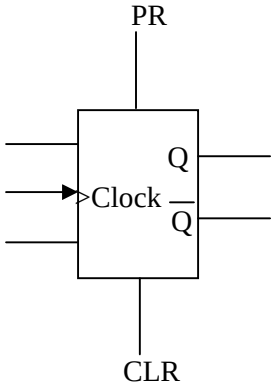


Relation between clock pulses and output JK flip-flop. When J = 1 and K = 1.

Please note Q = 1 once in every two clock pulses i.e., the output has half the frequency of the clock pulse. This property is utilized in binary counters.

Preset and Clear Inputs:

Usually, the J-K flip – flop works on the basis of the clock input. However, we can introduce two inputs called the PRESET and CLEAR in the last NAND level. As shown in fig 11.6(a) these inputs do not obey the clock pulses, and override them in action. Hence, they are called the asynchronous inputs. The working of these inputs is as given below.



11.6(a) Symbolic representation

PR	CLR	Q_0	Q_1
0	0	Forbidden	operations
0	1	0/1	1
1	0	0/1	0
1	1	Normal	operations

Fig 11.6(b) Truth table.

Fig 11.6: J – K flip – flop

Let PR = 0 and CLR = 1. Then, as and when PR = 1 and CLR = 0, Q will jump to 1 from whatever state it is in. This operation presets $\overline{Q}$ to 1. When PR = 1 and CLR = 0, the lower NAND produces 1 at its output, i.e., $\overline{Q}$ = 1. This makes Q = 0. Thus whatever be the state, Q will fall back to '0'. When both PR = CLR = 1, the NANDs are enabled and the J-K flip-flop functions in the normal way. However, the state PR = CLR = 0 is forbidden, since both Q and $\overline{Q}$ will be 1 in this state. The truth table for the PR and CLR terminals is as shown in Table 11.6a. fig 11.6b shows the symbol of J-K flip-flop.

RACE – AROUND CONDITIONS:

Consider the last two rows of the truth table of the J-K flip-flop, where both the J and K inputs are connected to logic-1 state. We find that when J = K = 1, Q_0 changes to $\overline{Q}_0$. That is, if Q_0 = 0, Q_1 = 1, and if Q_0 = 1, Q_1 = 0. These transitions gives rise to what is called the race-around problem. When J = K = 1 and CK = 1 the output will toggle from '0' to '1', and 1 to 0 continuously, until the clock pulse becomes zero. The result of this toggling action is that when the clock finally becomes zero, we well not know in what state Q would be. This is called race-around problem, as it is generated due to the racing of this signal around the feedback path, and around the circuit.

11.6 Master – Slave JK Flip – Flop (Solution to the Race-Around Problem):

The race around problem can be solved using two J-K flip-flops in the master-slave mode of operation. Here two S-R flip-flops are in the J-K mode of operation. The first flip-flop in the 'master' and the second one is the 'slave'. The master is converted into J-K mode of operation.

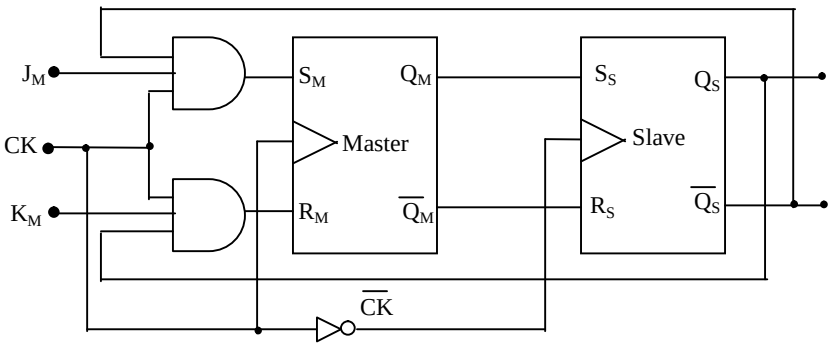


Fig 11.7 J – K master –slave flip – flop.

The slave inputs are connected to the outputs of the master. In this mode, the slave will also act as a J-K Flip – Flop. The Master is driven by clock pulse CK, while the slave is driven by $\overline{CK}$, which is obtained by inverting the CK pulses. The feedback connection can be seen to be from the Slave outputs to the master inputs.

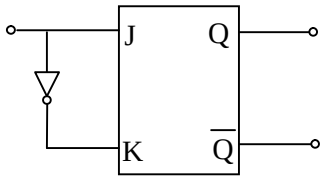
(i) When the clock is ON, the master gets enabled and the slave gets disabled. So, the inputs appearing at input terminals S_M and R_M of the master will appear at its outputs Q_M and $\overline{Q}_M$. respectively, which means that the external inputs now appear at the inputs S_s and R_s of the slave.

(ii) When the CK = 0, the master is disabled, and the slave is enabled because $\overline{CK}$ = 1. Thus the inputs at S_s and R_s will appear at Q_s and $\overline{Q}_s$, respectively, and hence at the inputs J_M and K_M of the master. However, as the master is now disabled, these inputs will not appear at S_s and R_s. Hence, no feedback problem occurs. Thus, the race-around problem is eliminated.

11.7 Applications of JK Flip – Flop:

a) JK Flip - Flop as D-Flip - Flop

The 'D' flip-flop has only one input, called the delay – or D – input. The truth table will have only four entries, as shown in fig 11.8b. when D is 0, Q will become'0'after the program delay t_p. similarly, if D = 1, output will become 1, irrespective of the previous value of the output. We thus find that the flip-flop will always produce an output equal to the input, after a time – delay equal to t_p. By cascading a number of D- Flip - Flops, the delay time can be increased.



11.8(a) D flip – flop.

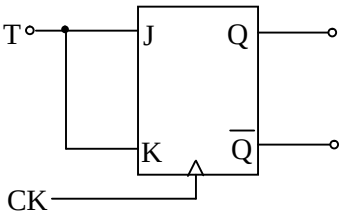
D_0	Q_0	Q_1
0	0	0
0	1	0
1	0	1
1	1	1

Fig 11.8(b) Truth Table of the D flip – flop.

Since the circuit output is produced after a time-delay, this is called the delay flip-flop. It is also called the D latch.

b) JK Flip-Flop as T-Flip- Flop :-

By shorting the J and K terminals, we can construct a T Flip-Flop. Usually, the T – input is tied to logical 1 . Fig 11.9a shows the circuit of the toggle flip-flop and fig 11.9b shows its truth table.



11.9(a) Toggle flip – flop.

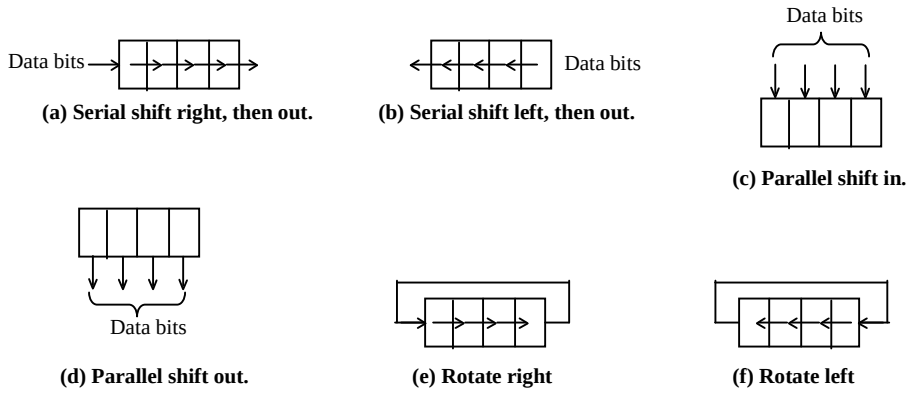
T_0	Q_0	Q_1
0	0	0
0	1	0
1	0	1
1	1	1

Fig 11.9(b) Truth Table of the T flip – flop.

Table shows that, when T = 0, $Q_1 = Q_0$. However, if T = 1, $Q_1 = \overline{Q}_0$. T flip – flops are useful in the construction of counters.

11.8 Shift Register:

Shift registers are very important in applications involving the storage and transfer of data in digital system. A register, in general, is used solely for storing and shifting data (1s and 0s) entered into it from all external source and possesses no characteristic internal sequence of states. The storage capability of a register is one of its two basic functional characteristics and makes it an important type of memory device.



(a) Serial shift right, then out. **(b) Serial shift left, then out.** **(c) Parallel shift in.**

(d) Parallel shift out. **(e) Rotate right** **(f) Rotate left**

Figure 11.10 Basic data movement in registers.

Registers are commonly used for the temporary storage of data within a digital system. The shift capability of a register permits the movement of data from stage to stage within the register or into or out of the register upon application of clock pulses.

Fig 11.10 shows symbolically the types of data movement in shift register operations. The block represents any arbitrary four-bit register, and the arrow indicates the direction and type of data movement.

11.9 Serial In – Serial out Shift Register:

This type of shift register accepts data serially, i.e. one bit at a time, and also outputs data serially. The logic diagram of a 4 – bit serial – in, serial – out, shift register is shown in fig 11.11. with four stages, i.e., four flip –flops, the register can store up to four bits of data.

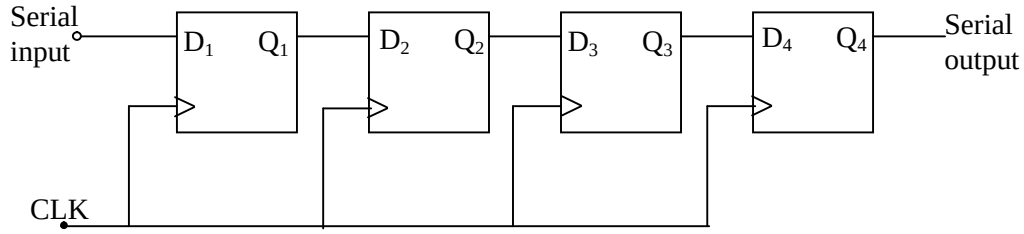


Figure 11.11 4 – bit serial – in, serial – out, shift register

Serial data is applied at the D input of the first Flip-Flop (FF). The Q output of the first FF is connected to the D input of the second FF, the Q output of the second FF is connected to the D input of the third FF and the Q output of the third FF is connected to the D input of the fourth FF. The data is outputted from the Q terminal of the last FF.

When serial data is transferred into a register, each new bit is clocked into the first FF at the positive – going edge of each clock pulse. The bit that was previously stored by the first FF is transferred to the second FF. The bit that was stored by the second FF is transferred to the third FF, and so on. The bit that was stored by the last FF is shifted out.

Fig 11.12 and 11.13a illustrate this process to store the data bits 0101 in the register. Initially all the FFs are reset, i.e., $Q_1 = 0$, $Q_2 = 0$, $Q_3 = 0$ and $Q_4 = 0$.

The right most bit '1' is applied at the D_1 input of FF_1. At the positive – going edge of the first clock pulse, this '1' is shifted into FF_1 and all other FFs store their respective bits at the D inputs. Therefore, $Q_1 = 1$, $Q_2 = 0$, $Q_3 = 0$ and $Q_4 = 0$, after the first clock pulse.

Then a '0' is applied at the D_1 input of FF_1. At the positive – going edge of the second clock pulse, this '0' is shifted to Q_1 of FF_1 and the D inputs of all other FFs are also shifted to their respective outputs. Therefore $Q_1 = 0$, $Q_2 = 1$, $Q_3 = 0$ and $Q_4 = 0$, after the second clock pulse.

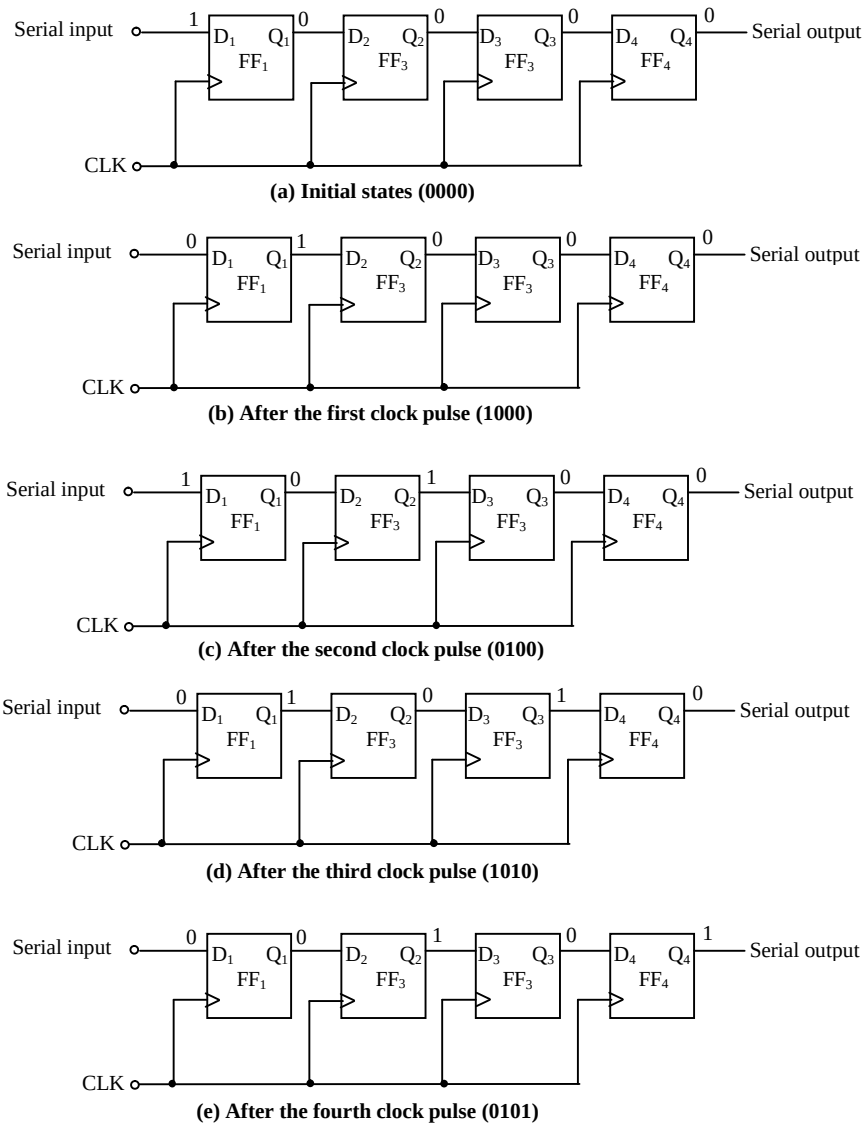


Fig11.12 Loading of the 4 – bit serial – in, serial – out, shift register

Then a '1' is applied at the D_1 input of FF_1. At the positive – going edge of the third clock pulse, this '1' is shifted into Q_1 of FF_1 and the D inputs of all other FFs are also shifted to their respective outputs. Therefore, $Q_1 = 1$, $Q_2 = 0$, $Q_3 = 1$ and $Q_4 = 0$, after the third clock pulse.

Then a '0' is applied at the D_1 input of FF_1. At the positive – going edge of the fourth clock pulse, this '0' is shifted into Q_1 of FF_1 and the D inputs of all other FFs are also shifted to their respective outputs. Therefore, $Q_1 = 0$, $Q_2 = 1$, $Q_3 = 0$ and $Q_4 = 1$, after the third clock pulse, this '0' is shifted to Q_1 of FF_1 and the D inputs of all other FF_S are also shifted to their respective outputs. Therefore, $Q_1 = 0$, $Q_2 = 1$, $Q_3 = 0$ and $Q_4 = 1$, after the fourth clock pulse.

After clock pulse	Serial input	Q_1	Q_2	Q_3	Q_4	
0	1	0	0	0	0	Initial states
1	0	1	0	0	0	
2	1	0	1	0	0	
3	0	1	0	1	0	
4	–	0	1	0	1	

Fig 11.13(a) Shifting in the data 0101 serially.

This completes the serial entry of 0101 into the 4-bit register fig 11.13 shows the timing diagram of the loading of serial input 0101 into the 4-bit serial-in, serial-out, shift register.

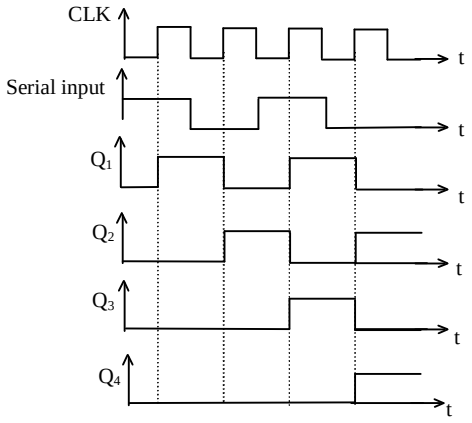


Fig 11.13(b) Timing diagram showing the loading of the serial input 0101 into the 4 – bit Serial-in, Serial-out, shift register.

The shifting out of the stored data 0101 serially from the register is illustrated in fig 11.14 . It requires four clock pulses to shift out the 4-bit stored data.

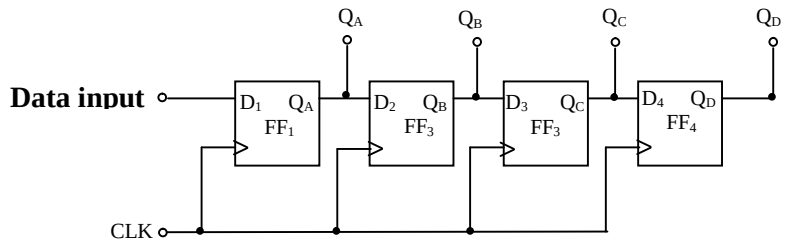
After clock pulse	Serial input	Q_1	Q_2	Q_3	Q_4
0	0	0	1	0	1
1	0	0	0	1	0
2	0	0	0	0	1
3	0	0	0	0	0
4	–	0	0	0	0

Fig 11.14 Shifting in the data 0101 serially.

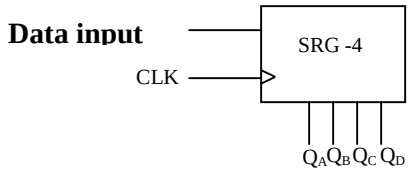
11.10 Serial In, Parallel – out Shift Register:

In this type of register, the data bits are entered into the register serially, but the data stored in the register is shifted out in parallel form. Fig 11.15 shows the logic diagram and the logic symbol of a 4-bit serial – in, parallel – out shift register.

Once the data bits are stored, each bit appears on its respective output line and all bits are available simultaneously, rather than on a bit-by-bit basis as with the serial output. The serial – in, parallel-out, shift register can be used as a serial-in, serial-out, shift register, if the output is taken from the Q terminal of the last FF.



(a) Logic diagram



(b) Logic symbol

Fig 11.15 A 4 – bit serial - in, parallel - out, shift register.

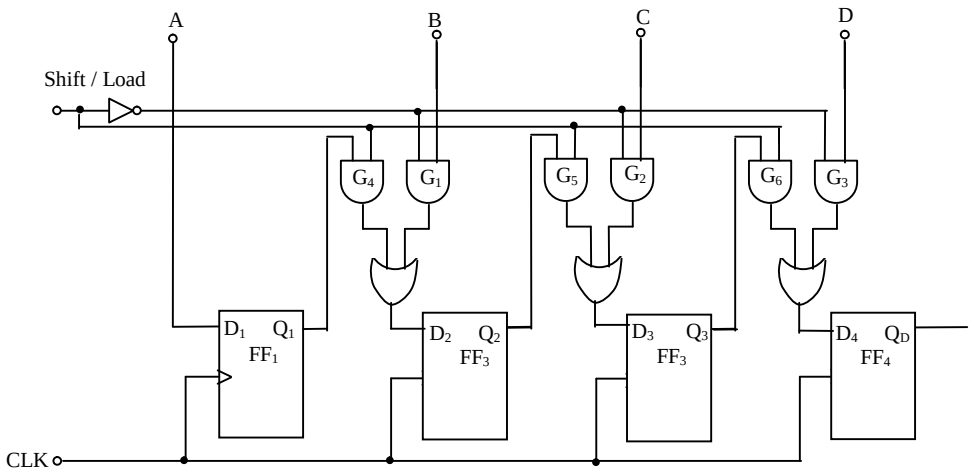
11.11 Parallel In, Serial – Out Shift Register:

In parallel – in, serial-out, shift register, the data bits are entered simultaneously into their respective stages on parallel lines, rather than on a bit-by-bit basis on one line as with serial data inputs, but the data bits are transferred out of the register serially, i.e., on a bit-by-bit basis over a single line.

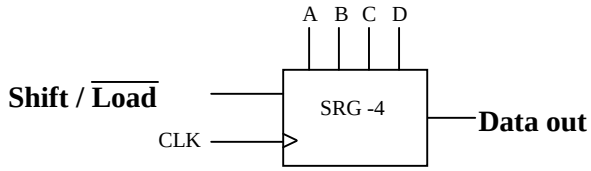
Fig 11.16 illustrates a 4-bit parallel-in, serial out, shift register using D flip-flops. There are four data lines A, B, C and D through which the data is entered into the register in parallel form. The signal shift / load allows (a) the data to be entered in parallel form into the register and (b) the data to be shifted out serially from terminal Q_4.

When Shift / Load line is high, gates G_1, G_2 and G_3 are disabled, but gates G_4, G_5 and G_6 are enabled allowing the data bits to shift – right from one stage to the next. When Shift / Load line is low, gates G_4, G_5 and G_6 are disabled, whereas gates G_1, G_2 and G_3 are enabled allowing the data input to appear at the D

inputs of the respective flip-flops. When a clock pulse is applied, these data bits are shifted to the Q output terminals of the flip-flops and therefore, data is inputted in one step. The OR gate allows either the normal shifting operation or the parallel data entry depending on which AND gates are enabled by the level on the Shift / Load input.



(a) Logic diagram



(b) Logic symbol

Fig 11.16 A 4-bit parallel, serial – out, shift register.

11.12 Parallel In, Parallel – out Shift Register:

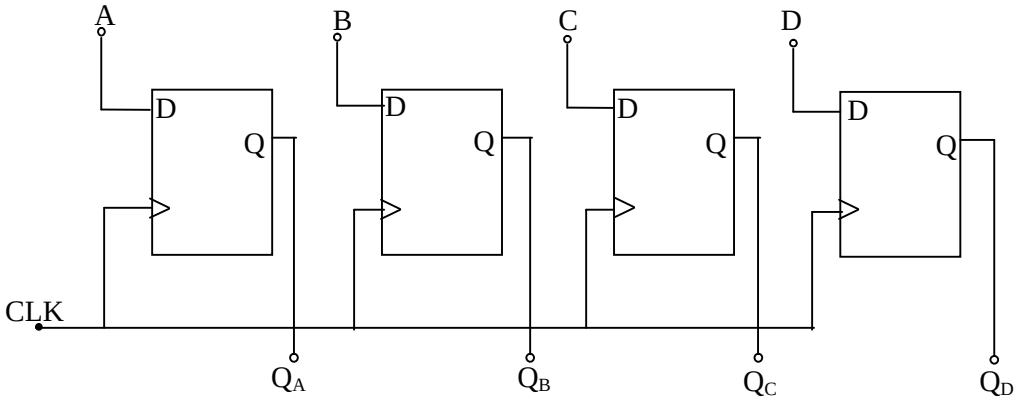


Figure 11.17 Logic diagram of a 4 – bit parallel – in, parallel – out, shift register

In a parallel – in, parallel – out, shift register, the data is entered into the register in parallel form, and also the data is taken out of the register in parallel form. Immediately following the simultaneous entry of all data bits, the bits appear on the parallel outputs.

Fig 11.17 shows a 4-bit parallel-in, parallel-out, shift register using D Flip-Flops. Data is applied to the D input terminals of the flip-flops. When a clock pulse is applied, at the positive going edge of that pulse, the D inputs are shifted in to the Q outputs of the Flip-Flops. The register now stores the data. The stored data is available instantaneously for shifting out in parallel form.

11.13 Bi-directional Shifter:

A bidirectional shift register is one in which the data bits can be shifted from left to right or from right to left.

Fig 11.18 shows the logic diagram of a 4-bit serial-in, serial-out, bi-directional shift register. Right/Left is the mode signal.

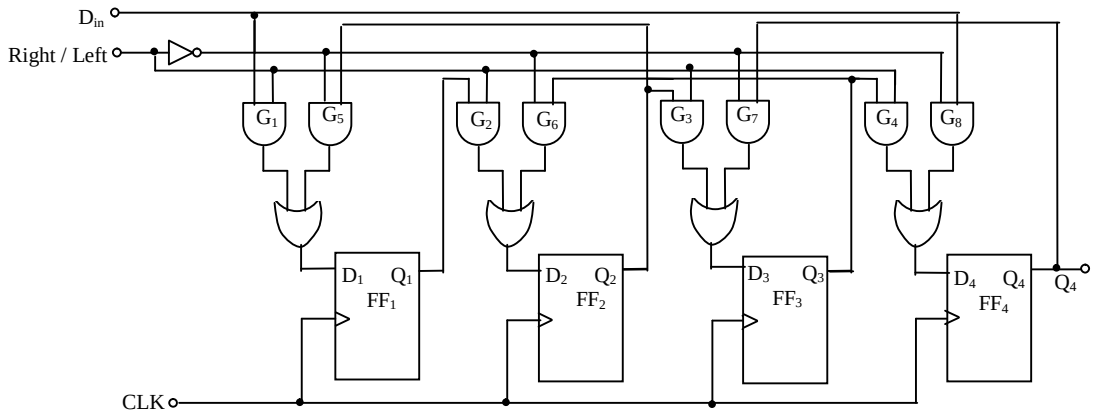


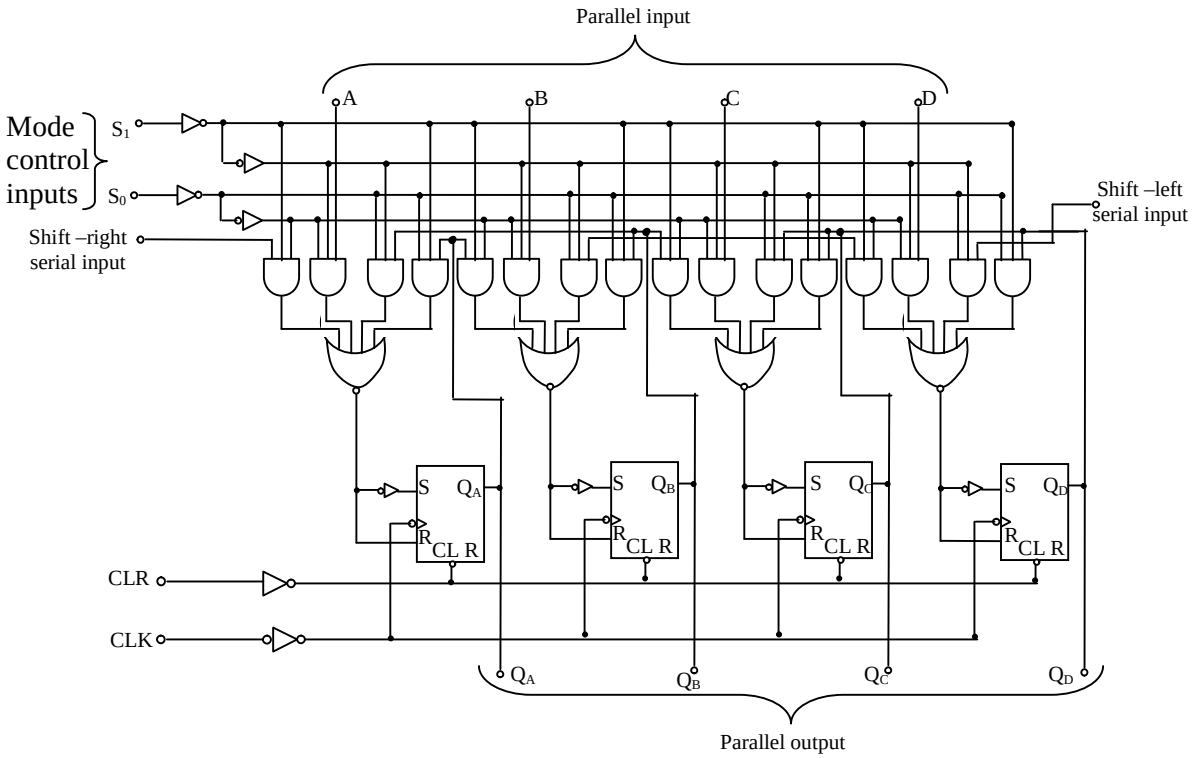
Fig 11.18 Logic diagram of a 4 – bit bi-directional shift register.

When Right / $\overline{\text{Left}}$ is a 1, the logic circuit works as a shift-right shift register. When Right / $\overline{\text{Left}}$ is a 0, it works as a shift – left register. The bidirectional operation is achieved by using the mode signal and two AND gates and one OR gate for each stage as shown in fig 11.18.

A HIGH on the Right / $\overline{\text{Left}}$ control input enables the AND gates G_1, G_2, G_3 and G_4 and disables the AND gates G_5, G_6, $G_7$and G_8 and the state of Q output of each FF is passed through the gate to the D input of the following FF. When a clock pulse occurs, the data bits are then effectively shifted one place to the right. A low on the Right / $\overline{\text{Left}}$ control input enables the AND gates G_5, G_6, $G_7$and G_8 and disables the AND gates G_1, G_2, G_3 and G_4 and the Q output of each Flip-Flop is passed to the D input of the preceding FF.

When a clock pulse occurs, the data bits are then effectively shifted one place to the left. Hence, the circuit works as a bidirectional shift register.

11.14 Universal Shift Register:



(a) Logic diagram

(b)Truth Table

Inputs S_1	Inputs S_0	Clock	Action
0	0	X	No change
0	1	m	Shift-right
1	0	m	Shift-left
1	1	m	Parallel load

Fig 11.19 The 74194 **4–bit universal shift register.**

A universal shift Register is a bidirectional register, whose input can be either in serial form or in parallel form and whose output also can be either in serial form or in parallel form.

Fig 11.19 shows the logic diagrams of the 74194 4-bit universal shift register. The output of each flip-flop is routed through AOI logic to the stage on its right and to the stage on its left. The mode control inputs, S_0 and S_1 are used to enable the left-to-right connections. When it is desired to shift-right, and the right-to-left connections. When it is desired to Shift-Left.

The truth table (fig 11.19b) shows that no shifting occurs when S_0 and S_1 are both LOW or both HIGH. When $S_0 = S_1 = 0$, there is no change in the contents of the register, and when $S_0 = S_1 = 1$, the parallel input data A, B, C and D are loaded into the register on the rising edge of the clock pulse. The combination $S_0 = S_1 = 0$ is said to inhibit the loading of serial or parallel data, since the register contents cannot change under that condition. The register has an asynchronous active – LOW clear input, which can be used to reset all the flip-flops irrespective of the clock and any serial or parallel inputs.

11.15 Counters:

A digital counter is a set of flip-flops (FFs) whose states change in response to pulses applied at the input to the counter. Thus, as its name implies, a counter is used to count pulses. A counter can also be used as a frequency divider to obtain waveforms with frequencies that are specific functions of the clock frequency. They are also used to perform the timing function as in digital watches, to create time delays, to produce non-sequential binary counts, to generate pulse trains, and to act as frequency counters, etc.

Counters may be asynchronous counters or synchronous counters. The term asynchronous refers to events that do not occur at the same time. Asynchronous counters are also called Ripple counters. The asynchronous counter has a disadvantage, in so far as the unwanted spikes are concerned. This limitation is overcome in parallel counters. Propagation delay is a major disadvantage in asynchronous counters because it limits the rate at which the counter can be clocked and creates decoding problem.

The term synchronous as applied to counter operation means that the counter is clocked such that each flip-flop in the counter is frigsered at the same time.

11.16 Asynchronous counters:

In a asynchronous counter, the flip-flop output transition serves as a source of triggering other flip-flops. In other words, the CP inputs of all flip-flops are triggered not by the incoming pulses but rather by the transition that that occurs in other flip-flops. In this section, we present some asynchronous counters and explain their operation.

Four – Bit Asynchronous Binary Counter:

Four – Bit asynchronous binary counter consists of a series connection of complementing flip-flops, with the output of each flip-flop connected to the CP input of the next higher-order flip-flop. The flip-flop lading the least significant bit receives the incoming count pulses. The diagram of a 4 – bit binary ripple counter is shown in fig 11.20. All J and K inputs are equal to ‘1’. The small circle in the CP input indicates that the

flip flop complements during a negative-going transition or when the output to which it is connected goes from 1 to 0.

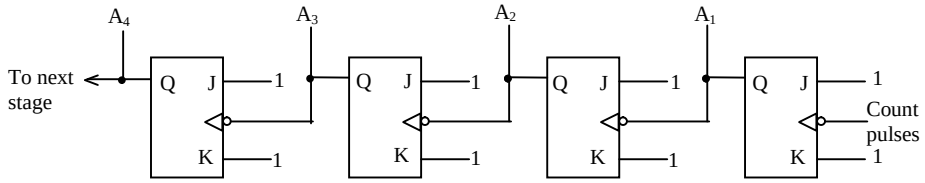


Fig 11.20 4 – bit binary ripple counter

Table 11.2 Count sequence for a binary ripple counter

Count sequence				Conditions for complementing flip-flops	
A_4	A_3	A_2	A_1		
0	0	0	0	Complement A_1	
0	0	0	1	Complement A_1	A_1 will go from 1 to 0 and Complement A_2
0	0	1	0	Complement A_1	
0	0	1	1	Complement A_1	A_1 will go from 1 to 0 and Complement A_2
					A_2 will go from 1 to 0 and Complement A_3
0	1	0	0	Complement A_1	
0	1	0	1	Complement A_1	A_1 will go from 1 to 0 and Complement A_2
0	1	1	0	Complement A_1	
0	1	1	1	Complement A_1	A_1 will go from 1 to 0 and Complement A_2
					A_2 will go from 1 to 0 and Complement A_3
1	0	0	0	and so on …	A_3 will go from 1 to 0 and Complement A_4

To understand the operation of the binary counter, refer to its count sequence given in Table 11.2. It is obvious that the lowest order bit A_1 must be complemented with each count pulse. Every time A_1 goes from 1 to 0, it complements A_2. Every time A_2 goes from 1 to 0, it complements A_3, and so on. For example, take the transition from count 0111 to 1000. The arrows in the table emphasize the transition in this case. A_1 is complemented with the count pulse. Since A_1 goes from 1 to 0, it triggers A_2 and complements it. As a result, A_2 goes from 1 to 0, which in turn complements A_3. A_3 now goes from 1 to 0, which complements A_4. The output transition of A_4, if connected to a next stage, will not trigger the next flip-flop, since it goes from 0 to 1. The flip-flops change are at a time in rapid success on, and the signal propagates through the counter in a ripple fashion. Hence asynchronous counters are sometimes called asynchronous counters.

The 7493A Four – Bit binary counter:-

The 7493A is presented as an example of a specific integrated circuit asynchronous counter. As the logic diagram in fig 11.21 shows, this device actually consists of a single flip-flop and a three –bit asynchronous

counter. This arrangement is for flexibility. It can be used as a divide – by-2 device using only the single flip-flop, or it can be used as a modulus – 8 counter using only the three-bit counter position. This device also provides gated reset inputs, RO(1) and RO(2). When both of these inputs are HIGH, the counter is RESET to the 0000 state by $\overline{CLR}$.

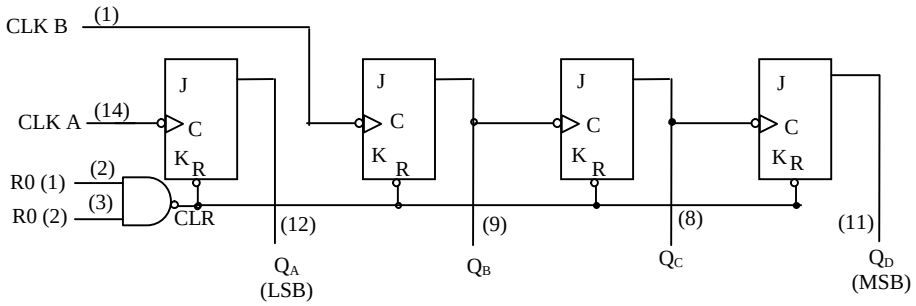
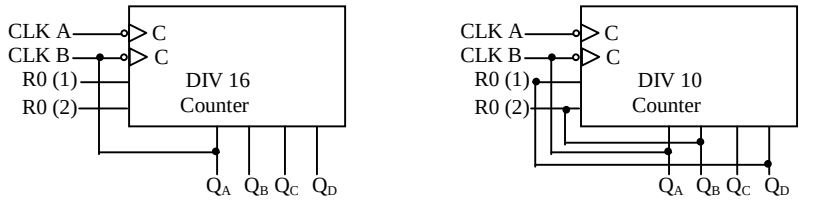


Fig 11.21 The 7493A four-bit binary counter logic diagram. (Pin numbers are in parentheses, and all J-K inputs are internally connected HIGH.

Additionally, the 7493A can be used as a four-bit modules-16 counter (counts 0 through 151) by connecting Q_A output to the CLK B input as shown in fig 11.22.



(a) 7493A connected as a modulus-16 counter

(a) 7493A connected as a decade counter

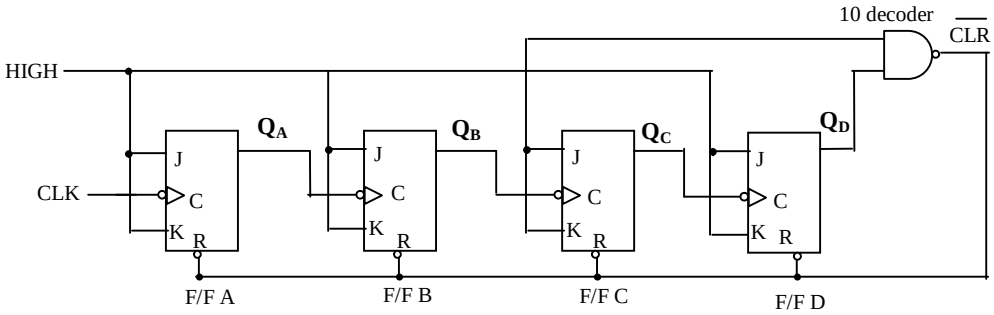
Fig 11.22 Two configurations of the 7493A asynchronous counter.

It can also be configured as a decade counter with asynchronous recycling by using the gated reset inputs for partial decoding of count 10_{10}, as shown in fig 11.22(b).

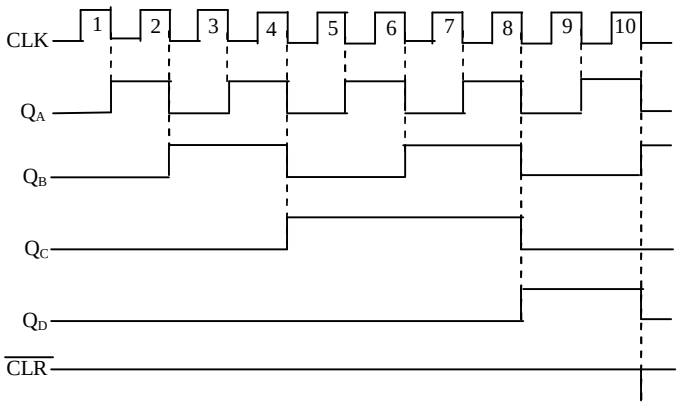
Asynchronous Decade Counters:

Counters with ten states in their sequence are called decade counter. A decade counter with a count sequence of 0(0000) through 9(1001) is a BCD decade counter because its ten-state sequence is the BCD code. This type of counter is very useful in display applications in which BCD is required for conversion to a decimal readout.

A decade counter requires four flip-flops. We will now take a four-bit asynchronous counter and modify its sequence in order to understand the principle of truncated counters. One method of achieving this recycling after the count of 9(1001) is to decode count 10_{10}(1010) with a NAND gate and connect the output of the NAND gate to the clear (CLR) inputs of the flip-flops, as shown in fig 11.23a.



(a)



(b)

Fig 11.23 An asynchronously clocked decade counter with asynchronous recycling.

Notice that only Q_B and Q_D are connected to the NAND gate inputs. This is an example of partial decoding, in which the two unique states (Q_B = 1and Q_D = 0) are sufficient to decode the count of 10_{10} because none of the other states (0 through 9) have both Q_B and Q_D HIGH at the same time. When the counter goes into count 10_{10}(1010), the decoding gate output goes LOW and asynchronously RESETS all of the flip-flops.

The resulting timing diagrams is shown in fig. Notice that there is a glitch on the Q_B wave form. The reason for this glitch is that Q_B must first go HIGH before the count of 10_{10} can be decoded. Not until several nano seconds after the counter goes to the count of 10_{10} does the output of the decoding gate go LOW. Thus, the counter is in the 1010 state for a short time before it is RESET back to 0000, thus producing the glitch on Q_B.

11.17 Synchronous Counters:

Synchronous counters are distinguished from ripple counters in that clock pulses are applied to the CP inputs of all flip-flops. The common pulse triggers all the filp-flops simultaneously, rather than one at a time in succession as in a ripple counter. The decision whether a flip-flop is to be complemented or not is determined from the values of the J and K inputs at the time of the pulse. If J = K = 0, the flip-flop remains unchanged. If J = K = 1, the flip-flop complements. In this section, we present some typical MSI synchronous counters and explain their operation.

Three – Bit Synchronous Binary counter:

Three bit synchronous binary counter is shown in fig 11.24(a) and its timing diagram in fig 11.24(b). An understanding of this counter can be achieved by a careful examination of its sequence of states as shown in Table 11.3.

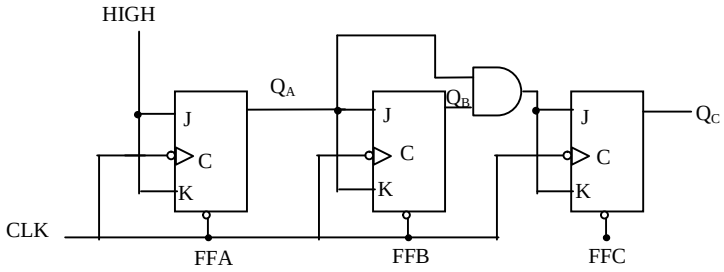


Fig 11.24(a) A three – bit synchronous binary counter.

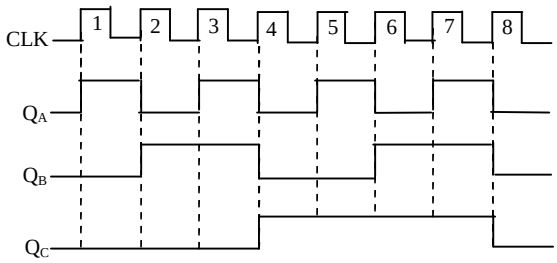


Fig 11.24(b) Timing diagram for the counter of figure.

First , let us look at Q_A. Notice that Q_A changes on each clock pulse as we progress from its original state to its final state and then back to its original state. To produce this operation, FFA must be held in the toggle mode by constant HIGH on its J and K inputs. Now let us see what Q_B does. Notice that it goes to the opposite state following each time Q_A is a 1. This occurs at CLK_2, CLK_4, CLK_6 and CLK_8. CLK_8 causes the counter to recycle. To produce this operation, Q_A is connected to the J and K inputs of FFB. When Q_A is a 1 and a clock pulse occurs. FFB is the toggle mode and will change state. The other times when Q_A is a 0, FFB is in the no-change mode and remains in its present state.

Table 11.3 State sequence for a three-stage binary counter.

Clock Pulse	Q_C	Q_B	Q_A
0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1

Next, let us see how FFC is made to change at the proper times according to the binary sequence. Notice that both times Q_C changes state, it is preceded by the unique condition of both Q_A and Q_B being HIGH. This condition is detected by the AND gate and applied to the J and K inputs of FFC. Whenever both Q_A and Q_B being HIGH, the output of the AND gate makes the J and K inputs of FFC HIGH, and FFC toggle on the following clock pulse. At all other times, the J and K inputs of FFC are held LOW by the AND gate output, and FFC does not change state.

11.18 Summary:

- A flip-flop is the basic memory element; it can store a '0' or a '1'.
- A flip flop is known more formally as a bistable mutivibrator. It has two stable states.
- Flip-Flops are used for data storage, counting, frequency division, parallel-to –serial and serial-to-parallel data conversion, etc.
- An unclocked flip-flop is called a latch, because the output of the flip-flop latches on to a 1 or a 0 immediately after the input is applied.
- The clocked D latch is called a transparent D latch, because its output follows the input when the clock is HIGH.
- The J-K Flip-Flop is the most versatile and most widely used of all the flip-flops.
- T flip-flops are not widely available as commercial items.
- The master-slave flip-flop is made up of two flip-flops- a master and slave. It was developed to make synchronous operation of flip-flops more predictable by overcoming the problem of race in clock flip –flops.
- A register is a set of FFs used to store binary data.
- A register in which shifting of data takes place is called a shift register.

- In a serial - in, serial - out, shift register, data is fed in serially, that is one bit at a time on a single line and data is also shifted out serially.
- In a serial-in, parallel-out, shift register, data is fed in serially but data is shifted out in parallel form, that is all bits at a time.
- In a parallel-in, serial-out, shift register, data is fed in, in parallel form but shifted out in serial form.
- In a parallel – in, parallel – out, shift register, data is both fed in and shifted out in parallel form.
- In a universal shift register, data can be shifted from left – to – right or right –to – left and also data can be shifted in or shifted out in serial form or in parallel form.
- Shift registers are used in digital system to provide time delays. They are also used for serial / parallel data conversion and in construction of ring counters.
- A digital counter is a set of FFS whose states change in response to pulses applied at the input to the counter.
- Counters may be asynchronous counters or synchronous counters.
- Asynchronous counters are also called ripple counters.
- In asynchronous counters, all the FF s do not change states simultaneously. They are serial counters.
- In synchronous counters, all the FF s change state simultaneously. They are parallel counters.
- In asynchronous counters if the clock frequency is very high owing to accumulation of propagation delay, skipping of states may occur.
- In synchronous counters, the propagation delays of individuals FF s do not add together.
- Synchronous counters have the advantages of high speed and less severe decoding problems, but the disadvantage of more circuitry than that of asynchronous counters.
- Generation of pulse trains using indirect logic has the advantage that any counter (ripple or synchronous) with the correct number of states can form the generator.

11.19 Key Terminology:

Flip – Flop: A memory device capable of storing a logic level.

D flip-flop: A type of bit table , multi vibrator in which the output follows the state of the D Input.

Register: A group of flip –flops capable of storing data.

Shift Register: A digital circuit capable of storing and shifting binary data.

Universal shift register: Shift – right, shift – left, shift register which can input and output data either serially or parallely.

Counter: A digital circuit capable of counting electronic events, such as pulses, by processing through a sequence of binary states.

Asynchronous counter: A type of counter in which the external clock signal is applied only to the first FF and the output of each FF serves as the clock input to the next FF in the chain.

Synchronous counter: A counter in which the circuit out puts can change state only on the transitions of a clock.

Binary counter: A counter in which the states of FF s represent the binary number equivalent to the number of pulses that have occurred at the input of the counter.

Ripple counter: A counter in which the external clock signal is applied to the first FF and then the clock input to every other FF is the output of the preceding FF.

11.20 Self – Assessment Questions:

1. Distinguish between combinational and sequential switching circuits
2. How do you build a latch using universal gates?
3. Explain the operation of clocked SR flip-flop.
4. List the different types of latches and flip-flops. Name the applications of each type.
5. How does a J-K flip –flop differ from an S-R flip-flop in its basic operation? What is its advantage over an S-R flip – flop?
6. What do you mean by toggling?
7. What is a serial-in, serial-out, shift register? Explain the operation of it.
8. What is a parallel-in, serial-out, shift register? Explain the operation of circuit.
9. What is a serial – in, parallel-out, shift register? Explain the operation of the circuit.
10. What is a parallel – in, Parallel – out, shift register? Explain the operation of the circuit.
11. What is a universal shift register? Explain the operation of the circuit.
12. What is the basic difference between a counter and a shift register?
13. What is the advantage of a synchronous counter over asynchronous counter? What is its disadvantage?
14. Explain the operation of 4-bit Asynchronous Binary counter.
15. Explain the operation of Decade counter.
16. What is synchronous counter? Explain the operation of 3-bit synchronous binary counters.
17. Explain the operation of synchronous decade counter.

11.21 References:

1 Digital Logic and Computer Design – M. Morris Mano, Prentice Hall Inc,.

2 Digital Fundamentals – 3rd Edition – Floyd, UBS Publishers.

3 Digital Electronics: An Introduction to Theory and Practice, 2nd edn., Prentice – Hall of India, New Delhi, 1987.

4 Switching and Finite Automatic Theory, 2nd edn., Tata -McGraw – Hill, New Delhi, 1978.

5 Tokheim, T., Digital Principles, Tata -McGraw – Hill, New Delhi, 1988.

Name of the Author:

S.V. RangaNayakulu, M.Sc., M.Phil., M.Ed.,

Dept. of Physics & Electronics

T.J.P.S. COLLEGE (P.G.COURSES)

GUNTUR-522 006

Unit III

LESSON – 12

Digital – to – Analog and Analog – to – Digital converters

Objectives of the lesson:

This lesson gives an idea of

1. Digital to Analog converters
2. Analog to Digital converters

After going through this lesson you will be in a position to explain

1. The working of weighted-Resistor type digital to Analog converter
2. The working of successive approximation Analog – to – Digital Converter
3. The sample and Hold Circuits.

Structure of the Lesson:

12.1 Introduction:-

The physical quantities such as voltage, current, temperature, pressure and time etc., are available in analog form. It is difficult to process, store or transmit the analog signal, even though, an analog signal represents a real physical parameter with accuracy. Therefore, for processing transmission and storage purposes, it is often convenient to express these variables in digital form. It gives better accuracy and reduces noise. The operation of any digital communication system is based upon analog signals to digital (A/D) and digital to analog (D/A) conversion.

Fig 12.1 highlights a typical application within which A/D and D/A conversion is used. This is used either in full or in part in applications such as digital audio recording and play back, computer, music and video synthesis, data acquisition, digital signal processing, microprocessor based instrumentation. Both A/D and D/A conversions are also known as data converters. We shall first discuss D/A followed by A/D, because A/D conversion usually makes use of a D/A conversion.

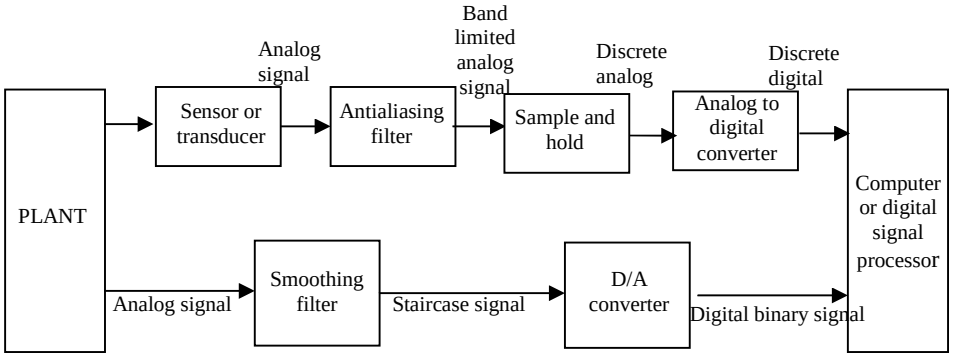


Fig 12.1 Applications of A/D and D/A converter.

12.2 Digital – to – Analog Converter [D/A] :

The Digital to Analog converter accepts data in digital form and converts it to a voltage or current which is proportional to the digital value.

The black diagram of a n-bit D/A converter is shown in fig 12.2. It has n-logic inputs and a single analog output, whose value varies discretely in response to the 2^n possible input bit patterns. The internal circuit can be all electronically switched ladder network.

Normally two types of networks are used:

i) Weighted resistor network ii) R – 2R ladder network.

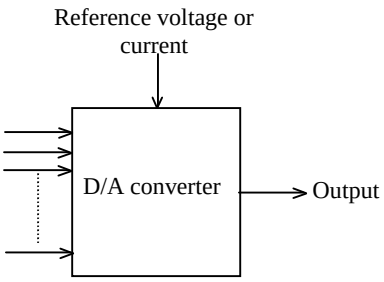


Fig 12.2 D/A Converter–Block diagram

12.2.1 The Weighted – Resistor Type D/A converter:

The diagram of the weighted – resistor D/A converter is shown in fig 12.3. The operational amplifier is used to produce a weighted sum of the digital inputs, where the weights are proportional to the weights of the bit positions of inputs. Since the op-amp is connected as an inverting amplifier, each input is amplified by a factor equal to the ratio of the feedback resistance divided by the input resistance to which it is connected. The most significant bit (MSB) i.e. D_3 is amplified by R_f / R, D_2 is amplified by R_f / 2R, D_1 is amplified by R_f / 4R, and D_0, the LSB is amplified by R_f / 8R.

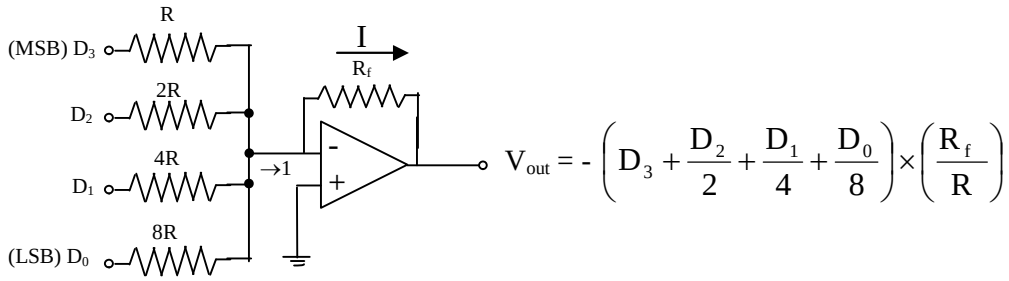


Fig 12.3 Weighted – Resistor Type D/A Converter

The inverting terminal of the op-amp in fig.12.3 acts a virtual ground. Since the op-amp adds and inverts.

$$V_{out} = - \left(D_3 + \frac{D_2}{2} + \frac{D_1}{4} + \frac{D_0}{8} \right) \times \left(\frac{R_f}{R} \right)$$

The main disadvantage of this type of D/A Converter is, that a different-valued precision resistor must be used for each bit position of the digital input.

12.2.2 Binary R – 2R Ladder D/A Converter:

This D/A Converter circuit avoids the above problem and requires only two values of resistors R and 2R (hence the name R-2R), although each bit position requires two resistors instead of the single resistor per bit

position of the binary weighted converter. The-n-bit R-2R ladder network with output operational amplifier is shown in fig 12.4(a).

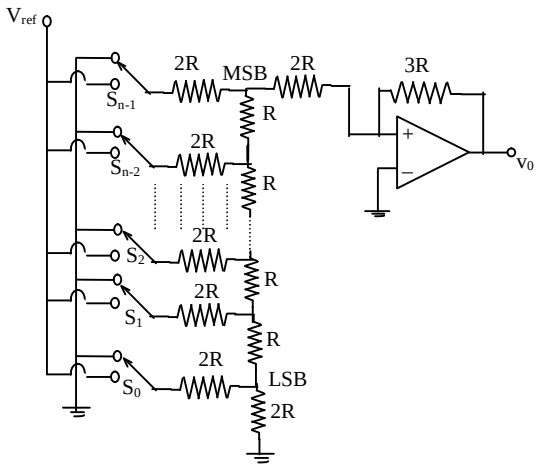
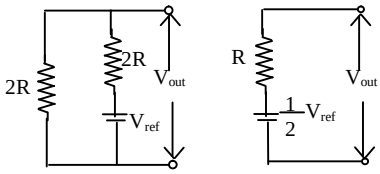


Fig 12.4(a) Binary R–2R Ladder D/A Converter.



(b) **(c)**

Fig 12.4(b) (c) Thevenin's equivalent circuits with only the LSB – switch.

The circuit may be analyzed by the use of Thevenin's theorem.

Assume that $S_0 = 1$, $S_1 = S_2 =$ ---- $S_{n-1} = 0$. The R- 2R network above the LSB position is then effectively the load for the circuit which can be redrawn as in Fig.12.4(b). The voltage across the open circuit terminals is thus V/2 and the value of series resistance is R (as 2R in parallel with 2R). The Thevenin's equivalent circuit is always the same as fig.12.4(c) for any bit position (i.e., $S_1 = 1$, $S_0 = S_2 = S_3 =$……..$S_{n+1} = 0$). To find the contribution of each bit position to the output voltage, let us redraw the circuit for the 3-bit D/A Converter and use the Thevenin's theorem, Fig.12.5 shows the effective Thevenin's circuits for each bit position.

We see that the weighted value of voltage of LSB, 1^{st},2^{nd} …MSB are respectively $V_{ref}/2^n$, $V_{ref}/2^{n-1}$, $V_{ref}/2^{n-2}$ … $V_{ref}/2$ volts. Therefore for n-bits D/A converter, the V_{out} at node X is given by

$$V_{out} = V_{ref}(S_0 2^{-n} + S_1 2^{-n-1} + S_2\, 2^{-(n-2)} + \ldots S_{n-2}\, 2^{-2} + S_{n-1}\, 2^{-1}) \text{ ---------- } (12.1)$$

In a common practice an additional resistor 2R is placed to the right of the MSB switch position so that the total load looking into the array is 3R. With a feedback resistor of value 3R, the gain of the operational amplifier is –1. Thus the output of Op-amp or of the n-bit R-2R ladder with Op-amp is given by

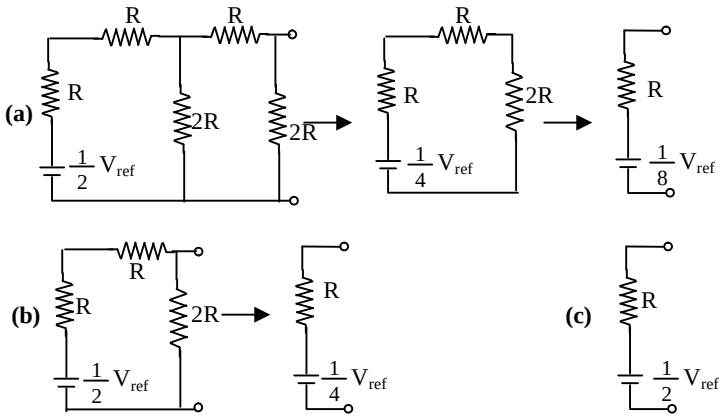


Fig 12.5Effective Thevenin's circuits for (a) MSB, (b) Next MSB and (c) LSB of the 3-bit R–2R D/A Converter.

$$V_0 = V_{ref}(S_0 2^{-n} + S_1 2^{-n-1} + S_2\, 2^{-(n-2)} + \ldots S_{n-2}\, 2^{-2} + S_{n-1}\, 2^{-1}) \quad \text{----------} \quad (12.2)$$

The negative sign corresponds to the inverting property of the Op-amp.

12.2.3 D/A converter Specifications:

(1). Resolution: The resolution is specified as the value of 'n' or the number of bits which the D/A converter can accept. The resolution can also be expressed as % Resolution = $100/(2^n - 1)\%$. Thus 3-bit D/A converter is of poor resolution. For a better resolution n must be large. The voltage conversion resolution is generally defined as $V_{ref}/2^n$.

(2). Linearity: An ideal D/A converter would have equal increments of the analog output for equal increments in the digital input. The deviation from the linear behavior is called non-linearity. It is expressed as

$$\% \text{ Non linearity} = 100D/M,$$

where D is the deviation of measured output from a straight line and M is the output range, usually the non-linearity is expressed as $\pm\frac{1}{2}$ LSB volts.

(3). Accuracy: The accuracy of a D/A converter is a measure of the difference between the actual analog output voltage and the ideal output voltage. There are many factors contributing to accuracy including the

linearity of the D/A converter, tolerances on reference voltages, amplifier gains, etc, and ability of the circuit to resist noise.

(4). Offset: For zero input there may be a finite output, which is called zero error. This is quite small and may be offset by adjustment (trimming) at the Op-amp stage.

(5). Gain error: The gain is the ratio of the D/A converter's full scale output value and the reference input value. The gain error is deviation of actual gain from the designed value. It can be minimized by using external preset resistors.

(6). Response or setting time: The setting time is defined as the time required for the output to reach and remain within, a specified band of voltage (or current) after an input change has occurred. It is caused by transients set up by voltage switching and the effects on the stray capacitance in the circuit. It is of the order of nano to micro seconds.

(7). Temperature Sensitivity: For a given fixed input value the output analog value varies with temperature, because of the sensitivity of the reference voltage, the resistors, Op-amp to temperature. It may be as high as ± 50 ppm/^{0}C.

12.3 Analog – to – Digital Converter:

Analog–to–digital conversion is slightly more complex than digital–to–analog conversion, and a number of different methods may be used. Four common conversion methods are described in this section. Of these, the successive approximation counter is the most widely used A/D converter, because it provides excellent performance for a wide range of applications at a reasonable cost.

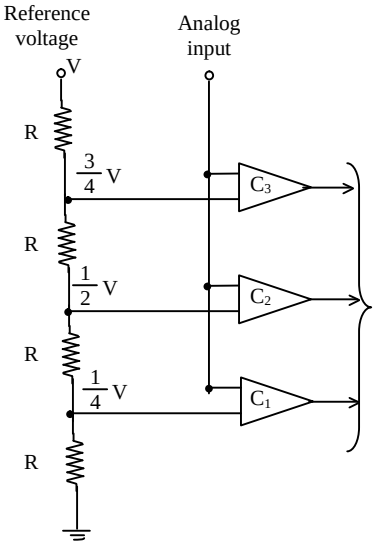
The comparator circuit forms the basis of all A/D converters. This circuit compares an unknown voltage with a reference voltage and indicates which of the two voltages is larger. A comparator is essentially a multistage high gain differential amplifier, where the state of the output is determined by the relative polarity of the two input signals. If, for instance, input signal A is greater than input signal B, the output voltage is maximum and the comparator is on. If input signal A is smaller than input signal B, the output voltage is minimum and the comparator is off. Since the amplifier has a very high gain, it either saturates or cuts off at relatively low differential input levels, so that it acts as a binary device.

12.3.1 Simultaneous A/D Converter:

Simultaneous A/D converter is shown in Fig 12.6, where three comparator circuit are used. Each of the three comparators has a reference input voltage, derived from a precision reference voltage source. A resistive divider consisting of four equal precision resistors is connected across the reference supply and

provides output voltages of $\frac{3}{4}$V, $\frac{1}{2}$V, and $\frac{1}{4}$V, where V is the reference output voltage. The other input terminal of each comparator is driven by the unknown analog voltage.

In this example the comparator is on (providing an output) if the analog voltage is larger than the reference voltage. If none of the comparators is on, the analog input must be smaller than $\frac{1}{4}$V. If comparator C_1 is on and both C_2 and C_3 are off, the analog voltage must be between $\frac{1}{4}$V and $\frac{1}{2}$V. Similarly, if C_1 and C_2 are both on and C_3 is off, the analog voltage must be between $\frac{1}{2}$V and $\frac{3}{4}$V; if all comparators are on, the analog voltage must be greater than $\frac{3}{4}$V. In total, four different output conditions may exist from no comparators on to all comparators on. The analog input voltage can therefore be resolved in four equal steps. These four output conditions can be coded to give two binary bits of information. This is shown in the table alongside the diagram of 12.6. Seven comparators would give three binary bits of information, fifteen comparators would give four bits, etc.



C_1	C_2	C_3	Input voltage
Off	off	off	0 to $\frac{1}{4}$V
on	off	off	$\frac{1}{4}$V to $\frac{1}{2}$V
on	on	off	$\frac{1}{2}$V to $\frac{3}{4}$V
on	on	on	> $\frac{3}{4}$V

Fig 12.6 Simultaneous analog-to-digital converter.

The advantage of the simultaneous system of A/D conversion is its simplicity and speed of operation, especially when low resolution is required. For a high – resolution system (a large number of bits), this method requires so many comparators that the system becomes bulky and very costly.

12.3.2 Successive – approximation A/D Converter:

This is one of the most popular A/D converter and has a resemblance with counter type A/D converter. Fig 12.7(a) gives the outline of a successive approximation A/D converter.

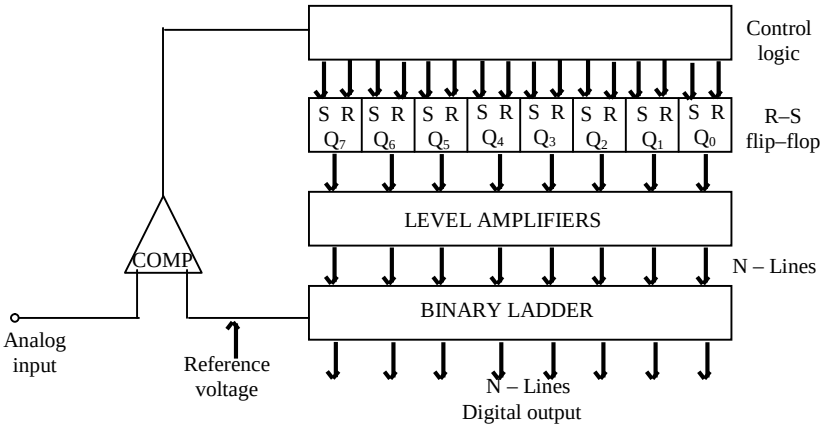


Fig 12.7(a) Block diagram of a successive approximation A/D converter

The converter circuit comprises of a control logic that is used to set or reset the flip-flops X_0, X_1…..X_7. The output of these flip flops are given to the level amplifiers and then to the binary ladder which provides the digital output as well as the analog reference voltage.

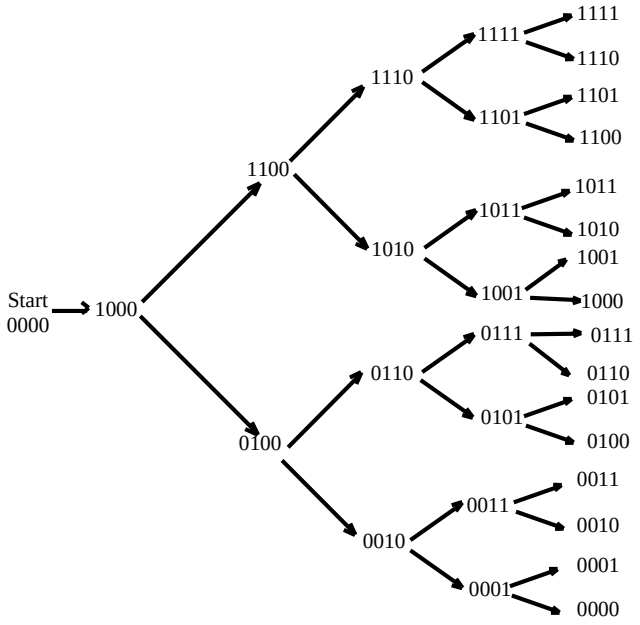


Fig 12.7(b) Operation diagram for a 4-bit successive approximation A/D Converter

Initially, the control logic resets all the flip-flops Q_0 to Q_7. Now its MSB is set to 1 so that digital signal reaching the binary ladder is 1000 0000. Analog voltage produced by binary ladder is now compared with the analog input signal. If the reference signal is smaller than the analog input, the comparator output gives a signal to the control logic which now switches the output of $Q_6$to 1 so that digital signal becomes 1100 0000 and its equivalent analog voltage is compared with the input signal. On the other hand if in the first comparison, comparator finds that reference voltage is more than the analog input, then a signal is given to the control logic to reduce the reference voltage accordingly. Q_7 is reset to '0' and Q_6 is set to 1. The digital signal now becomes 0100 0000 which is again compared with the analog signal. This process is repeated until analog input and reference voltage produced become equal. At this stage, a stable digital output is produced. The complete approximation procedure is illustrated for a four bit successive approximation A/D converter with the help of operation diagram fig 12.7(b).

An important feature of successive approximation A/D converter is that an 'n-bit' A/D converter requires only 'n' clock cycles for conversion. For a 8-bit A/D converter of this type, if a clock frequency of 1MHz is employed, the conversion time is only 8μs; which is quite low as compared to counter type or continuous type A/D converter.

Advantages:

1. It is more accurate than the staircase A/D converter.
2. It maintains a high resolution.
3. It is much faster.
4. Conversion time is much less.

Disadvantages:-

1. It requires a complex register called the successive-approximation register (SAR).
2. It is costly, as it contains more components.

12.33 Dual – slope A/D Converter:

The Dual – slope A/D converter is shown in fig 12.8

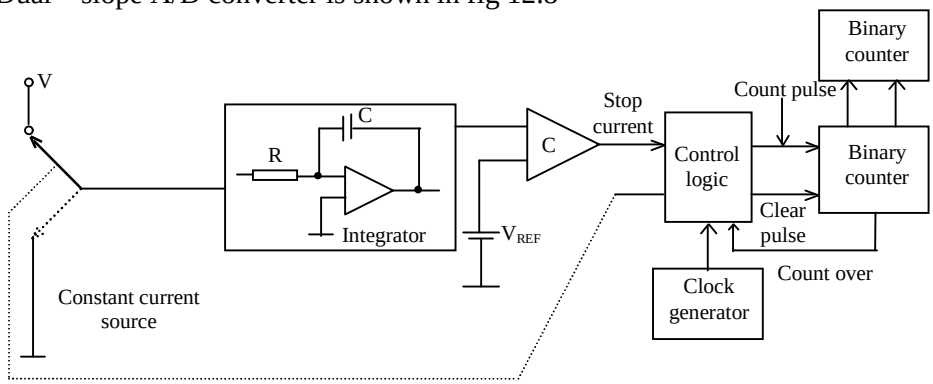


Fig 12.8 Circuit of dual–slope A/D converter.

The basic principle of operation of the circuit can be understood from fig 12.9. In Fig 12.9, we find that we charge a capacitor C for a fixed duration. This means that the integrator capacitor will charge to a voltage level depending on the value of the input analog voltage. Now, after the charging interval is over, the integrator capacitor is allowed to discharge at a constant (uniform)

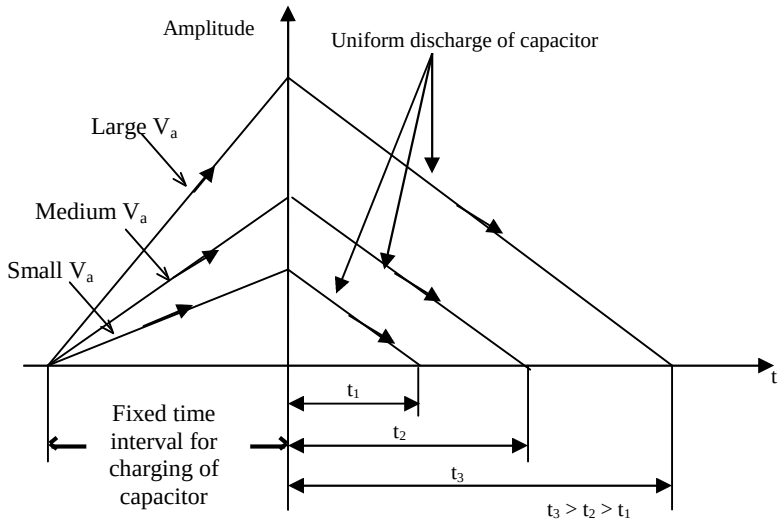


Fig 12.9 Charge–discharge intervals of integrator.

rate through a constant-current circuit. Depending on the level to which the capacitor had been initially charged to, the discharge time varies as shown, and hence the binary counter's counting time varies. This then reads the input voltage in terms of the count duration. The larger this duration, the more the count.

In the circuit shown in fig 12.8, there is a toggle switch that connects the input data and the constant-current generator to the integrator, alternately. The toggle switch is controlled by a control-logic circuit. The output of the integrator, alternately. This toggle switch is controlled by a control-logic circuit. The output of the integrator is applied to a comparator, which gives a stop command once its output exceeds the reference voltage V_{ref}. The control logic controls the clock, start count and clear count pulses of the binary counter, as well as the signal-pole, double-throw (SPDT) toggle switch.

At the start command, the integrator starts charging to the level of the input voltage for a fixed interval. Digital counter is not energized in this interval. At the end of the fixed-time interval the control switch throws the SPDT switch back to the constant-current generator, and the binary counter starts counting. When the integrator voltage drops to a value such that $V_{ref} > V_{int}$, the comparator stops the count. The reading on the digital counter represents the value of the input V_a.

Advantages:

1. Clock frequency drift is compensated for, as the same clock and integrator are used to perform conversions during the positive and negative intervals of the count period.

2. The above reason also increases accuracy.
3. Setting the clock rate and reference-input value can give desired scaling of the counter output.
4. The counter can be designed to be in binary, BCD, or any other desired display form.
5. Low conversion time.

Disadvantages:-

1. Complicated circuitry
2. Higher cost.

12.3.4 Counter Type A/D Converter: The principle is to adjust the D/A converter's input code until the D/A converter's output comes within ±(1/2) LSB to the analog input V_a which is to be converted to binary digital form. Thus in addition to the D/A converter, we need suitable logic circuitry to perform the code search and a comparator of adequate quality to announce when the D/A converter output has come within ±(1/2) LSB to V_a.

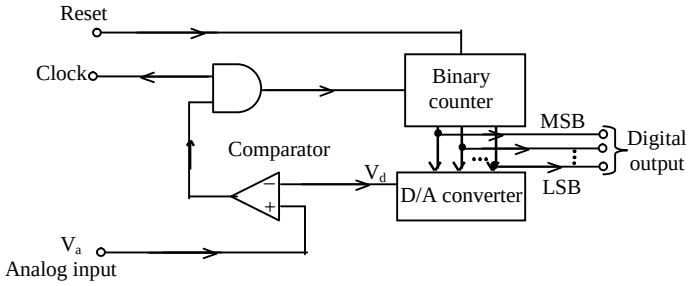


Fig 12.10(a) A counter type A/D converter.

A 3-bit counting A/D Converter based upon the above principle is shown in Fig.12.10 (a). The counter is reset to zero count by the reset pulse. Upon the release of RESET, the clock pulses are counted by the binary counter. These pulses go through the AND gate which is enabled by the voltage comparator high output. The number of pulses counted increase with time. The binary word representing this count is used as the input of a D/A converter whose output is a staircase of the type shown in Fig.12.10(b). The analog output V_d of D/A converter is compared to the analog input V_a by the comparator. If $V_a > V_d$, the output of the comparator becomes high and the AND gate is enabled to allow the transmission of the clock pulses to the counter. When $V_a < V_d$, the output of the comparator becomes low and the AND gate is disabled. This stops the counting at the time $V_a \leq V_d$ and the digital output of the counter represents the analog input voltage V_a. For a new value of analog input V_a, a second reset pulse is applied to clear the counter. Upon the reset, the counting begins again as shown in Fig 12.10(b). The counter frequency must be low enough to give sufficient time for the D/A converter to settle and for the comparator to respond. Low speed is the

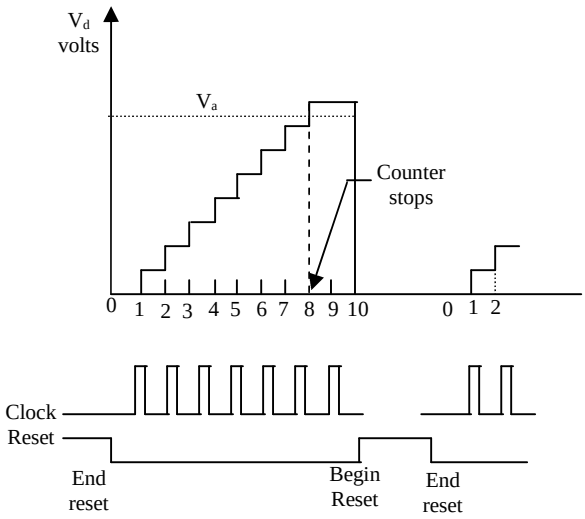


Fig 12.10(b) D/A output staircase waveform.

most serious drawback of the method. The conversion time can be as long as (2^n-1) clock periods depending upon the magnitude of input voltage V_a. For instance, a 12-bit system with 1 MHz clock frequency, the counter will take $(2^{12}-1)\mu s = 4.095$ ms to convert a full scale input.

If the analog input voltage varies with time, the input signal is sampled, using a sample and hold circuit before it is applied to the comparator. If the maximum value of the analog voltage is represented by n-pulses and if the clock period is T seconds, the minimum interval between samples is nT seconds.

12.35 A/D Converter Specifications:

(i) Analog input voltage:- This is the maximum allowable input voltage range.

(ii) Resolution:- It is the voltage charge at the input needed for 1 bit charge at the output. It is the resolution of D/A converter which is in A/D Converter, which is also known as the "quantization error". Thus all 8-bit A/D Converter has better resolution thus that of 4-bit A/D Converter.

(iii) Accuracy:- The accuracy of all A/D Converter is the maximum deviation of the digital output from the ideal linear curve. It depends upon the accuracy of its circuit elements, such as the D/A converter, comparator and reference supplies. This is in addition to the quantization error and is in general temperature dependent.

(iv) Conversion time:- It is defined as the time between two successive conversions at the maximum possible rate. It depends upon the analog voltage, clock frequency and number of bits.

12.4 Sample & Hold Circuit:-

The sample-and-hold circuit is basically an operational amplifier that charges a capacitor during the sample mode and retains the charge of the capacitor during the hold mode. The sample-and-hold circuit can be represented by the simple switch and capacitor of fig.12.11

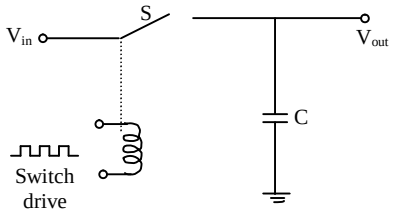


Fig 12.11(a) Circuit

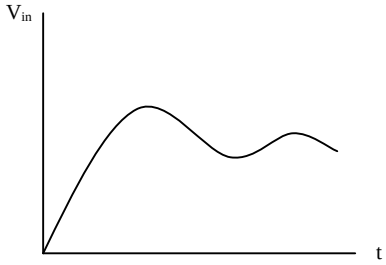


Fig 12.11(b) Input waveform

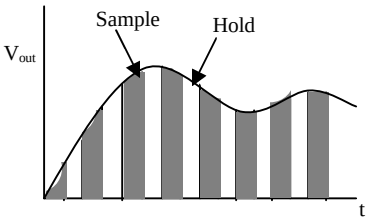


Fig 12.11(c) Operation of the sample-and-hold circuit.

When the switch is first closed, the capacitor charges to the value of the input voltage and then follows the input (assuming a low driving source impedance). When the switch is opened, the capacitor holds the voltage that it had at the time the switch was opened (assuming a high-impedance load).

The acquisition time of the sample-and-hold is the time required for the capacitor to charge up to the value of the input signal after the switch is first shorted. The aperture time is the time required for the switch to change state and the uncertainty in the time that this change of state occurs. The holding time is the length of time the circuit can hold the charge without dropping more than a specified percentage of its initial value.

It is possible to build a sample-and-hold circuit exactly as shown in Fig.12.11. However the sample-and-hold circuit is built, it always acts as the simple switch and capacitor shown.

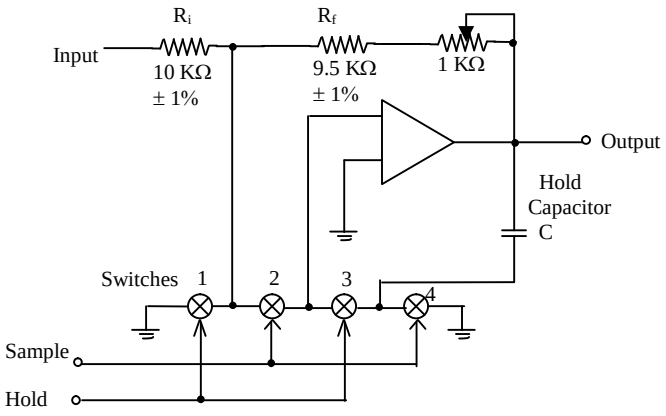


Fig 12.12 Sample–and–hold circuit

An actual sample-and-hold circuit is shown in the schematic diagram of Fig.12.12. The sample pulse operates switches 1 and 3; the hold pulse operates switches 2 and 4. The sample-and-hold control pulses are complementary. In the sample mode, the hold capacitor is charged up by the operational amplifier. In the hold mode the capacitor is switched into the feedback loop, while the input resistor R, and the feedback resistor R_f are switched to ground. Since the input to the amplifier remains within a few μV to ground (except during switching), the input impedance is 10kΩ in both the sample and the hold modes.

12.5 Summary:-

- Digital to analog conversion is the process of converting a value represented in a digital code to a voltage or current proportional to the digital value.
- The output of a D/A converter can be either a voltage or current.
- D/A converter essentially requires resistors, electronic switches and op-amp.
- A weighted resistor D/A converter requires a wide range of resistor values for better resolution where as, an R-2R ladder D/A converter requires only two values of resistors.
- The R-2R ladder type D/A converter is the most popular D/A converter.
- Analog to digital conversion is the process of converting an analog input voltage to a number of equivalent digital output levels.

- An A/D Converter produces a digital output proportional to the value of the input analog signal.
- A/D converts are either direct type or indirect type. Most direct type A/D Converter's require a D/A converter.
- The important direct A/D Converter techniques are counting type, parallel comparator and successive approximation technique.
- The important indirect A/D Converter techniques are Dual slope A/D Converter and charging balancing A/D Converter and dual slope A/D Converter.
- Successive approximation type A/D Converter is the most versatile. It completes n-bit conversion in just n-clock periods. Most monolithic A/D Converter's are successive approximation type.
- Dual slope converters are suitable for precise measurement of slowly varying signals. All digital voltmeters use dual-slope A/D Converter. The disadvantage is of long conversion time.
- Monolithic dual slope A/D Converter's are available in microprocessor compatible and display – oriented versions.
- The successive – approximation type A/D Converter is the most widely used type of A/D Converter.
- The important converter characteristics are: Resolution, linearity, Accuracy, monotonocity, settling time, stability etc.

12.6 Key Terminology:-

A/D Converter: The circuit, which converts all

D/A Converter: The circuit, which converts a digital input into an analog output.

Linearity error: The maximum deviation in step size from the ideal step size in D/A converter.

Percentage Resolution: The ratio of the step size to the full-scale value of a D/A converter. It can also be defined as the reciprocal of the maximum number of steps of a D/A converter.

Resolution: In a D/A converter, the smallest that can occur in the output for a change in the digital input. It is also called the step size. In all A/D Converter, the smallest amount by which the analog input must change to produce a change in the digital output.

Monotonicity: A property whereby the output of a D/A converter either increase or stays at the same value, but never decrease as the input is increased.

12.7 Self-assessment Questions:-

1. With the help of neat diagrams explain the working of R-2R ladder type digital to analog converter.
2. What is the advantage of the R-2R ladder D/A converter over the weighted resistor type D/A converter ?.
3. Explain the operation of Dual-slope A/D Converter.
4. Explain the operation of successive – approximation type A/D Converter.
5. Explain the important specifications of D/A and A/D converters.
6. Give two advantages and one disadvantage of the dual – slope A/D Converter.
7. Explain the operation of counter – type A/D Converter.

12.8 References :-

1. Op-amps and linear Integrated circuit Technology, Gayakwad, R.A., Prentice-Hall India Ltd., New Delhi.
2. Digital Circuit and Logic design, Lee, S.C., Prentice – Hall of India Ltd., New Delhi.
3. Digital Integrated Electronics, Taub & Schilling, McGraw Hill, New York.
4. Linear Integrated circuits, Roy Choudhury, Shail Jain. New Age International (p) Limited, Publishers. New Delhi.

Name of the Author:

S.V. RangaNayakulu, M.Sc., M.Phil., M.Ed.,
Dept. of Physics & Electronics
T.J.P.S. COLLEGE (P.G.COURSES)
GUNTUR-522 006

UNIT IV

LESSON 13

INTEL 8085 MICROPROCESSOR

Objectives:

1. To understand what a microprocessor is and what are its applications
2. To understand what is a microcomputer system.
3. To understand the function of various parts of 8085 microprocessor and its internal architecture.
4. To understand the propose of various pins and the bus-timings of 8085
5. To acquire knowledge about various groups of instructions and addressing modes.

Structure:

13.1 Introduction:

The latest development in the field of computer technology is the microprocessor, which is the central processing unit of a microcomputer. The microprocessors are the outcome of the trend towards smaller computers, which started in the middle of 1960s. The microprocessors appeared in 1970s and had made a remarkable progress in recent years. It has numerous applications such as industrial control, military applications, consumer and commercial equipment etc.

The Central Processing Unit (CPU) of a digital computer built on a single semiconductor chip using LSI or VLSI technology is called a microprocessor. A digital computer whose CPU is a microprocessor, is called a Microcomputer.

The first microprocessor INTEL 4004, introduced in 1971 by INTEL corporation, was an instant success. Several applications were developed using this programmable device. In quick succession, several processors with enhanced address bus size, improved instruction set were introduced into the market. INTEL 8080, 8085A, Zilog 80 (Z80), 6500, 6800 are examples of popular 8-bit processors.

Unlike several other manufactures, the INTEL corporation supplied detailed instruction set, technical details and developed a kit for learning about 8085 microprocessor. The 16-bit processors of INTEL namely 8086 and 8088 are similar members with similar internal architecture. IBM adopted INTEL processors as CPUs in its brand of personal computers. Further, all over the world INTEL 8085 is being used as model to acquaint about processors and various interfacing aspects. So, it is still in circulation and we will learn about this processor in this lesson.

13.2 Introduction to microcomputer:

Computer being a general purpose device, can be used in different varieties of applications – word processing, computation or controlling industrial process are few of them. But the microcomputers, as the name implies, are small computers having limited capacity, are used for specific applications.

Figure (13.1) shows a block diagram of a simple micro computer consists of a CPU, memory, input and output devices.

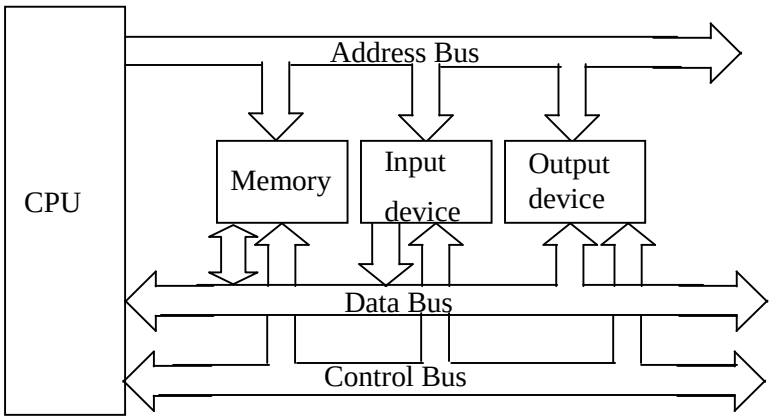


Fig 13.1 Block diagram of a simple microcomputer consists of a CPU

The function of a microcomputer is receiving data and information and processing it. It means performing arithmetic are logical computations on data, store the results or data or information in memory and display the results of the computation.

The means used to receive data are known as input devices. Some of the input device are: keyboards, switches, mouse etc. A device which performs computations is known as arithmetic logic unit (ALU). In

microcomputers, this task , together with control of all devices is performed by a single chip called microprocessors. Some of the well-known microprocessors are: 8085, 8086, Z80, 650Z etc. Storage of data instructions is accomplished using memories. Cassettes, floppy disks, CD ROMs, semiconductor memories are some examples for memories. Presenting of results is done by output devices. Commonly used output devices are monitor (VDU), printers, Seven segment LED displays, LCDs. Figure (13.1) gives a block diagram of a typical microcomputer.

Various peripherals-I/o devices, memories etc – are connected to the microprocessor by means of three types of buses: Address bus, data bus and control bus. A bus is a group of conducting wires.

Over the address bus, the microprocessor transmits the address of the device, with which it desires to communicate (access). The number of devices or memory locations that a CPU can address is determined by the number of address lines. For example, a CPU with 16 address lines can address 2^{16} or 65,536 or 64K(2^{10} = 1024 = 1K) locations. This bus is unidirectional. It means information can flow in one direction only. The CPU sends the address information on these lines.

Over the data bus, CPU can read in or writes out the data to various memory or I/O devices. Therefore, this bus is bi-directional. Normally data bus widths of sizes 8, 16, 32 etc. are used.

Contol bus may consist of some input lines, some output lines and some bi-directional lines depending upon the processor. CPU sends out some control signals like memory read, I/O write etc that are necessary for memory and I/O devices to work.

13.3 Architecture of 8085 microprocessor :

The internal logic design of the microprocessor is called its architecture, determines how and what various operations are performed by the microprocessor. As shown in the block diagram (figure 13.2), it consists of arithmetic and logic unit ALU, register section, a timing and control unit and other sections like address / date bus buffers, interrupt control and serial I / O control.

13.3.1 Register Section:

Inside the 8085, there are several internal registers, which are used for executing a program.

a) Accumulator: It is also designated as register ‘A’ or Acc. It is an 8-bit register used for executing various arithmetic and logical operations. In executing many of the instructions, one of the operands must be in Acc. The accumulated result will also be in Acc, as the name ‘Accumulator’ implies. For example, during the addition of two 8-bit integers, one of the operands must be in Acc and the result also stored in Acc.

In the I/O mapped I/O mode of data transfer communication with external devices takes place via Accumulator. To send data to output devices in I/O mapped I/O mode the 8-bit data to be sent must be kept in Acc before give OUT instruction. Similarly IN instruction transfers data from input port addressed to Acc.

b) General purpose registers: The 8085 has six general purpose 8-bit registers labeled as B, C, D, E, H and L. They can be used individually to store 8-bit information or can be used in pairs to hold 16-bit information or data or address. When used in pairs, only contains combinations viz., B – C, D – E and H – L are permitted. These registers are programmable, means that a programmer can use them to load or transfer data from the registers by using instructions.

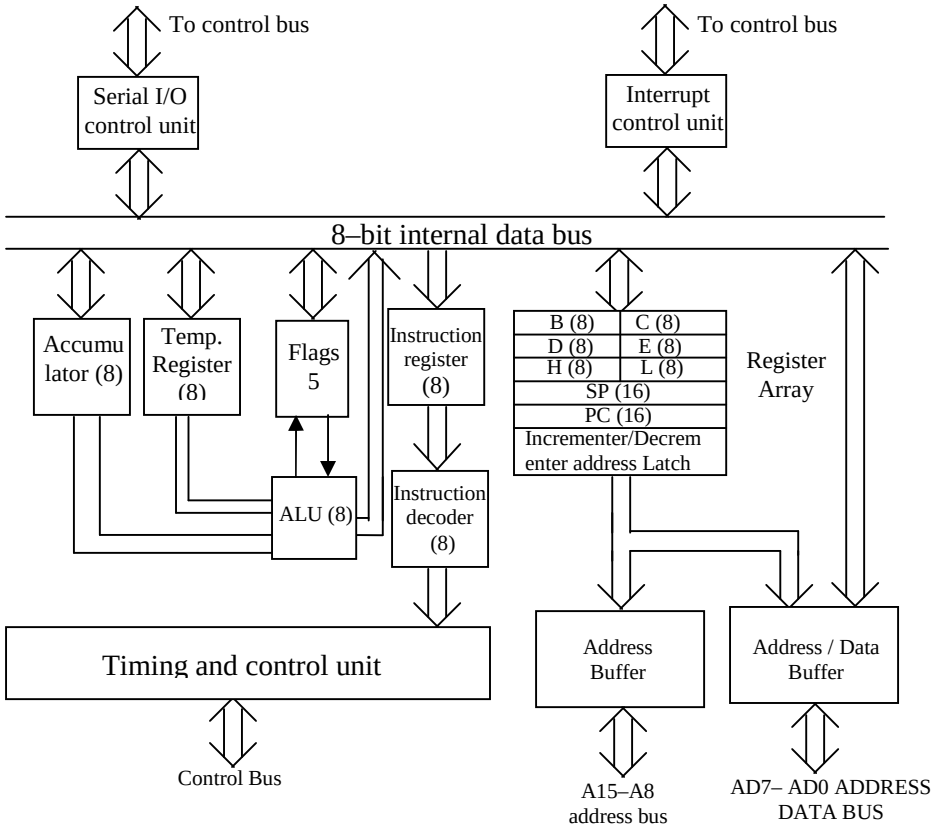


Fig 13.2 Block diagram of 8085 Microprocessor

c) Program Counter (PC): PC is a 16-bit register used to point the memory address from which the next instruction is to be fetched. Some times it is also called memory pointer. The contents of the PC is automatically incremented by the microprocessor during the execution of the instruction, so that at the end of execution of the present instruction it points to the address of the next instruction in the memory.

The signal of RESET pin of timing and control unit presets the PC to 0000H memory location which is the address of the very first instruction to be executed. The instruction JUMP and subroutine call, also change the contents of PC.

d) Stack Pointer (SP): Stack is a portion of R/W memory set aside by the programmer. The stack is used to save the contents of a register during the execution of a program. Stack pointer holds the address of the top element of the stack. Whenever something is added to the stack, the stack pointer is decremented and whenever something is received from the stack, the stack pointer is incremented. Hence, the SP always points to the top of the stack.

e) Instruction register and instruction decoder: After fetching an instruction code from memory, the microprocessor stores it in the instruction register, which is then decoded by instruction decoder and then microprocessor performs the required operation.

f) Status flag register: The flag register or status register consists of five status flip-flops called flags. Each of these flip-flops holds 1-bit information that indicates certain condition that occurs during an arithmetic or logic operation. The five status flags available in 8085 as shown below are Sign (S), Zero (Z), Auxiliary carry (AC), Parity (P), and carry (Cy).

D_7	D_6	D_5	D_4	D_3	D_2	D_1	D_0
S	Z	X	AC	X	P	X	CY

Fig 13.3: Bit position of the flags in flag register. (X stands for unused)

Sign flag S – After the execution of a signed arithmetic or logical operation in ALU, if D_7 bit of the result in accumulator is 1, the sign flag is 1 indicating that the result is negative.

Zero flag Z – Zero flag is set to 1, if the ALU operation results in 0, and the flag resets to zero, if the result is not 0.

Auxiliary Carry AC – In an arithmetic operation, when a carry is generated by digit D_3 and passed on to D_4, the AC flag is set.

Parity flag P – If an operation produces even number of 1 in its result, the flag is set to 1. For odd number of 1s in the result, the flag is reset to 0.

Carry flag C – If an ALU operation produces a carry or borrow in its result, the carry flag set to 1, otherwise it is reset.

g) Temporary register: This register is used to hold information from the memory or from the register array of the ALU. The other input to the ALU comes from the accumulator.

h) Address Buffer and Address/Data buffer: Through these buffers microprocessor transfers address as well as data required for memory and input/output devices.

13.3.2 The Arithmetic and Logic Unit (ALU):

The ALU performs the computing functions like addition, subtraction, logical AND, OR and EX-OR, increment, decrement, complement, compare etc. It includes the accumulator, the temporary register, the arithmetic and logic circuits and flag register. The result is stored in accumulator and flags one set or reset according to the result of the operation.

13.3.3 Timing and Control Unit:

Main function of the microprocessor in the microcomputer system is to perform computations and controlling peripherals like memory and I/o devices. The timing and control unit provides status, control and timing signals, which are necessary for the control of these peripherals.

13.4 Pin Configuration of 8085 microprocessor:

The INTEL 8085 is an 8-bit microprocessor fabricated in a 40 Pin DIP package. It operates with single +5VDC supply and with a clock speed of 3 MHz. All the signals corresponding to 40 pins can be divided into six functional groups: 1) Address bus, 2) data bus, 3) Control and status signals, 4) power supply and clock signals, 5) Interrupts and peripheral initiated signals, and 6) serial I/O ports. The pin diagram and signal diagram are shown in Figs15.4 a and 15.4b.

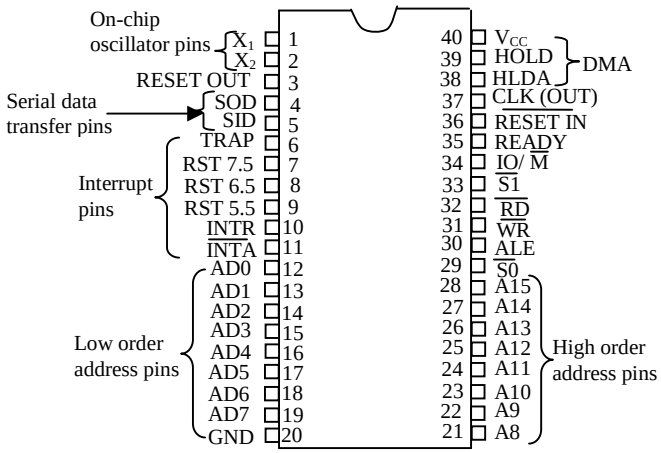


Fig 13 . 4(a) 8085 IC pin diagram

13.4.1 Address Bus:

The 8085 has 16-bit address bus capable of addressing 2^{16} (= 2^6 x 2^{10} = 64 x 1k = 64k) memory locations. The high order 8 address lines $A_8 – A_{15}$ carries high order address of memory through this bus.

13.4.2 Multiplexed Address/data bus:

These 8 lines are time multiplexed low order 8-bit address as well as 8-bit data bus. In the execution of an instruction, during the first clock cycle microprocessor sends out low order 8-bit address and during the remaining clock cycle microprocessor sends out 8-bit data on these pins in multiplexed mode. However, the low – order address bus is separated from data bus using a latch.

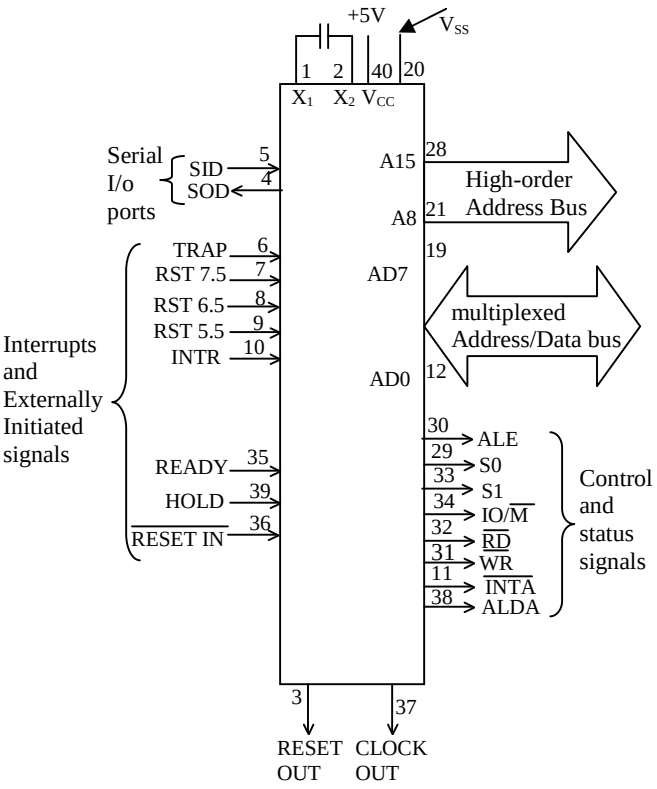


Fig 13.4b 8085 signal diagram

13.4.3 Control and status signals:

This group of signals includes two control signals, $\overline{RD}$ and $\overline{WR}$, three status signals, IO/$\overline{M}$, S_1 and S_0, and one special signal ALE. The functions of these signals are as follows:

ALE – Address Latch Enable: A positive going pulse is generated every time when microprocessor sends out address over multiplexed bus. It is used to latch the low-order address from the multiplexed address/data bus on to an octal latch, to generate separate set of eight address lines A_0 to A_7.

$\overline{RD}$ - This active low control signal indicates that the selected I/O or memory device is to be read and data is available on the data bus.

$\overline{WR}$ - This active low write control signal indicates that the data on the data bus, is to be written into a selected memory or I/O device.

IO/$\overline{M}$ - This status signal is used to differentiate between I/O and memory operations. A high on this line indicates an I/O operation whereas a low indicates a memory operation. This signal is combined with $\overline{RD}$ and $\overline{WR}$ to generate I/o and memory control signals.

S_1 and S_0 - These status signals can be used to identify various operations of microprocessor as shown below.

Status signals			Operations
Io/$\overline{M}$	S_1	S_0	
0	1	1	Opcode fetch
0	1	0	Memory Read
0	0	1	Memory write
1	1	0	I/O Read
1	0	1	I/O write
1	1	1	Interrupt Knowledge

13.4.4 Power Supply and clock frequency:

The power supply signals: V_{cc} is +5 volts DC supply and V_{ss} - ground reference X_1 & X_2 – An external crystal connected to these two pins generates a clock frequency by the internal oscillator for the operation of microprocessor. The oscillator frequency is divided by two internally; therefore, to operate a system at 3MHz, the crystal should have the frequency of 6MHz.

CLK OUT – This signal can be used as the system clock for other devices.

13.4.5 Interrupts and Externally initiated operations:

Interrupts : When an interrupt is recognized, the next instruction is executed from a fixed location in the memory. The 8085 has five interrupt signals that can be used to interrupt a main program execution. These are TRAP, RST 7.5, RST 6.5 RST5.5 and INTR. The TRAP is a non-maskable interrupt having highest priority among the other interrupts.

INTR – INTR is an interrupt request signal. An interrupt is used by an I/O device to transfer data to the microprocessor without wasting its time.

$\overline{INTA}$ - $\overline{INTA}$ is an interrupt acknowledge signal sent by the microprocessor after INTR is received.

$\overline{RESET\ IN}$ - It resets the program counter to zero, causes the microprocessor to execute the first instruction at the 0000H location.

READY – Input signal . It is used by slower devices to request the microprocessor to wait until they are ready for data transfer.. Microprocessor enters into WAIT states as long as this pin is active.

13.4.6 Serial I/O Ports:

The 8085 has two signals to implement the serial date transmission.

SID – Serial data input line. A serial data bit on this line is loaded into the most significant bit (D7) of accumulator, when RIM instruction is executed.

SOD – Serial data output line. The most significant bit (D7) of accumulator is output on this line when SIM instruction is executed.

13.5 8085 BUS configuration:

The program instructions are stored in memory. To execute an instruction the microprocessor first sends out the address of the memory location on address bus. The address is decoded and the memory location ,where the instruction is stored is accessed. The instruction operation code(op code) is fetched through data bus, and placed in instruction decoder. Here the number of bytes further needed is identified and the number of machine cycles needed is determined . Timing and control section issues necessary signals to proceed further with the instruction and the execution of the instruction is completed. These steps are known as "Fetch, Decode and Execute". All these operations are performed in synchronization with system clock. Therefore these operations are performed at a given time period, depending on the clock frequency. During these operations the 8085 signal activity be visualized by studying its timing diagram. The timing diagrams will give us a clear understanding of how the microprocessor works.

Timing diagram:

Figure 13.5 is the timing diagram of MOV B, A (op code 47H) stored in a memory location say, 3050H. Figure shows the timing activities of various signals like address bus, data bus, $\overline{RD}$, $\overline{WR}$, ALE, IO/$\overline{M}$ etc.

The first one in the figure is the clock waveform. One cycle of this clock is called a state labeled as T.

A basic microprocessor's operation such as reading a byte from memory or writing into a port is called a machine cycle. The time a microprocessor requires to fetch and execute an entire instruction is known as an instruction cycle. An instruction cycle may consist of one or more machine cycles.

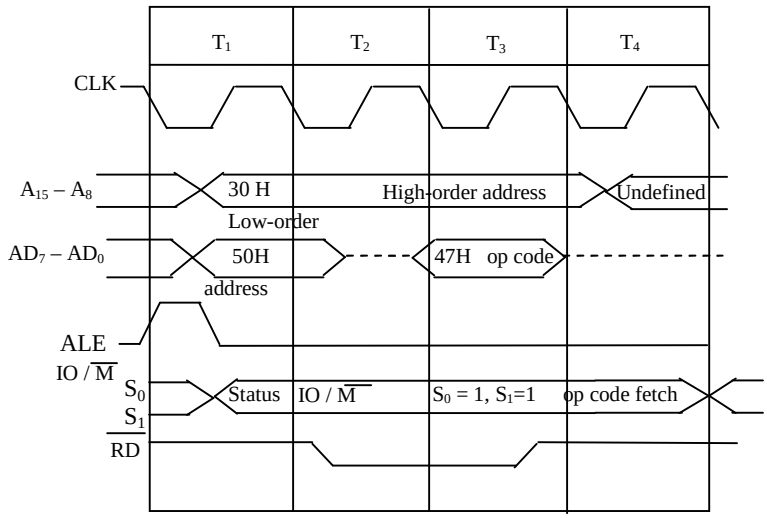


Fig. 13.5 MOV B,A timing diagram

The sequence of steps involved in the execution of an instruction MOV B, A having the operation code 47H stored in memory location 3050H are as follows (See figure 13.5):

Step 1 – The microprocessor places the contents of program counter 3050H on the address bus, such as 30H on the high order address bus and 50H on the low – order AD_0-AD_1 bus.

Step 2 – ALE goes high indicating $AD_0 – AD_7$ bus contains low order address. This signal is used to store lower byte of address in an octal latch for further use as AD bus gets converted to data bus from second clock cycle.

Step 3 – IO/$\overline{M}$ goes low to indicate the operation relates to memory.

Step 4 – Both S_1 and S_0 goes high to tell the operation is op code fetch. All these operations from step 1 to step 5 is performed in T_1 state.

Step 5 – During T_2, $\overline{RD}$ goes low since it is memory read operation, the program counter is incremented by one i.e.3051H, ALE goes low. The content of memory location 3050H, which is 47H, the op code of MOV B, A is placed on the data bus.

Step 6 – In T_3 state, 47H is read by microprocessor and transferred instruction decoder.

Step 7 – During T_4, microprocessor decodes the instruction and executes the specified operation i.e. transferring of accumulator contents into register B.

Above figure shows a complete machine cycle (in this case it is also equal to one instruction cycle) consisting of four states. If the clock frequency (f) is 2 MHz, the time for one clock period is

$T = \frac{1}{f} = \frac{1}{2 \times 10^6} = \frac{10^{-6}}{2} = 0.5\mu s$

$\therefore$ Total execution time = T- states **x** clock period = 4 x 0.5 x 10^{-6} = 2 μs.

13.6 Addressing modes:

A microprocessor system operates under the control by a series of instructions that make up a program. An instruction is a command to the microprocessor to perform a specific operation on certain data.

Let us consider the following instructions

a) MOV C, B – This instruction moves the contents of register B into register C.

b) LDA 2050H – Loads the contents of the memory location 2050H into accumulator.

Each instruction has two parts: first part indicates the task to be performed and it is called Mneumonic (code in English characters that stands for binary op code) The second part is called operand.

In the first example MOV specifies the tasks of instruction (to move the content of one register to another)and is therefore called op code. In the remaining part of the instruction two data registers (source , destination) called operands are specified

In the second instruction, LDA is the op code(in this instruction Acc is one of the operands). A 16 bit number is specified for operand . This number stands for a memory location and represents the second operand. According to INTEL notation if 2050 location is intended the instruction has to be entered in the order as op code, lower byte of address, upper byte of address i.e 3A 50 20 . Here 3A is the hexadecimal equivalent for the mneumonic of LDA op code.

Above examples indicates that each instruction specifies an operation to be performed on certain data. There are certain techniques by which the address of the data to be operated upon may be specified. The various ways of specifying the operand are called the addressing modes.

There are five addressing modes

i) Direct addressing mode ii) Register addressing mode iii) Register indirect addressing mode

iv) Immediate addressing mode v) Implicit addressing mode

13.6.1 i) Direct addressing modes: In this mode, the address of the operand is specified directly within the second and third byte of the instruction itself.

Examples:

a) STA 1850H: Store the contents of accumulator into memory location 1850H.

b) LHLD 2015H: Load the contents of memory location 2015H into register L and the contents 2016H into register H.

13.6.2 ii) Register addressing mode: In this addressing mode, the source or destination or both operands are located in the microprocessor registers

Examples:

a) MOV A, B – The contents of register B is copied into register A
b) ADD B – The contents of register B is added to the contents of accumulation and the result stored in accumulator.

13.6.3 iii) Register indirect addressing mode : In this addressing mode, the address of the operand is specified by a register pair.

Examples:

a) MOV A, M – The contents of the memory location whose address is specified in H-L register pair is copied into accumulator.
b) LDAX B – The contents of the memory location whose address is specified by B-C register pair is copied into register A.

13.6.4 iv) Immediate addressing mode: In this addressing mode, the operand is specified in the instruction itself.

Examples:

a) MVI B, 25H – The 8-bit data 25H is moved in to register B
b) LXI H, OF5AH – The 16-bit data OF5AH is moved into H-L register pair such that 5AH is moved into register L and OFH is moved into register H.

13.6.5 v) Implicit addressing mode: Some instructions do not require the address of the operand. The operands are implicit in the instruction itself. Most of these instructions operate on the contents of the accumulator.

Examples:

a) CMA – Compliment the contents of accumulator.
b) STC – Set carry flag.

13.7 Instruction set:

It is clear from the previous studies that the 8085 has a collection of 8-bit and 16-bit registers and 8-bit memory locations that can be used for programming purpose. The collection of these register and memory

locations used by the programmer is called programmer's model. The programmer's model of 8085 consists of

i)8-bit Accumulator ii) Six 8-bit registers: B, C, D, E, H & L or three 16-bit registers: B-C, D-E & H-L .iii) 16-bit program counter PC iv)16-bit stack pointer SP v) 8-bit flags register vi) 64KBmemory locations divided into ROM, RAM and stack areas and vii) 64 I/O ports.

An instruction is a command to the microprocessor to program a specific task. The entire group of instructions are called its instruction set. The 8085 instruction set has 74 op-codes that result in 246 instructions. Depending on their function all these instructions can be classified into the following categories:

i) Data transfer group ii) Arithmetic group iii) Logical group iv) Branch group

v) Stack, I/o and machine – control group.

13.7.1 i) Data transfer group: These instructions copy data from a source into a destination without modifying the contents of the source and without affecting the flags.

8-bit data transfer as

1) MOV destination, source – Copies the contents of source into destination where the source may be
   - an immediate value (e.g.# 35H)
   - an 8-bit register (A, B, C, D, E, H & L)
   - the contents of a memory location whose address is specified in H-L reg. Pair

   The destination may be
   - an 8-bit register (reg.)
   - the contents of a memory location whose address is specified in H-L reg. pair. But both source and destination should not be memory contents and the destination should not be an immediate value.

Examples:

Between registers :

MOV B, C – data of register C is copied into register B.

Specific data byte into a register or memory

MVI B, 2FH – data 2FH copied into register B

MVI M, 2FH – data 2FH into a memory location whose address is specified by H-L pair (Before this instruction H-L pair should be loaded with a memory address)

Data transfer between memory location and a register

MOV M, B – copy the contents of reg. B into a memory location address by H-L pair.

MOV B, M – copy the contents of memory location address by H-L pair into reg. B.

LDA 16-bit address: Load accumulator directly the contents of memory location whose address is specified in the second and third byte of instruction.

e.g.: LDA 2050 H – Load the contents of memory location 2050H into the acc.

STA 16-bit address: Store accumulator direct.

The contents accumulator is stored into a memory location whose address is specified with second and third byte of instruction.

STA 30A0H – Store accumulator content directly into memory location whose address is 30A0H.

LDAX rp: Load acc . indirect rp = B – C / D – C.

The contents of memory location whose address in the register pair rp is loaded into the acc.

STAX rp: Store acc. Indirect rp = B – C / D – E.

The contents of acc. is stored in memory, whose address is in the register pair rp.

e.g.: STAX B – Store the contents of acc. Into the memory location addressed by B – C.

16-bit data transfer:

1) LXI rp, data 16: Load register pair immediate.

rp = B – C / D – C | H – C | sp.

The 16-bit immediate data is loaded into the specified register pair.

e.g. LXI H, 2450H – Loads 2450H into HL pair such that 50H is in L and 24H in H.

e.g: LDAX B – This instruction will load the contents of the memory location, whose address is in B – C reg. Pair into acc.

2) LHLD 16-bit addr. : Load H – L pair direct.

The contents of memory location, whose address is specified in second and third byte of instruction is loaded into register L. The contents of the next memory location is loaded into register H.

e.g: LHLD 3050H – This instruction loads the contents of 3050H into register L and the contents of 3051H into register H.

3) SHLD 16-bit addr: Store H – L pair direct.

The contents of register L is stored into the memory location whose address is specified in the instruction, and the contents of register H is stored in the next memory location.

e.g: SHLD 3500H – This instruction will store the contents of register L in the memory location 3500H and the contents of register H is the location 3501H.

4) XCHG: Exchange the contents of H – L with D – E.

The contents of register pair H – L are exchanged with that of the register pair D – E.

e.g. if DE=2050;HL=307F ; after the execution of XCHG instruction DE=307F; HL=2050

13.7.2 ii) Arithmetic group:

These instructions perform arithmetic operations such as addition, subtraction, increment and decrement.

i) 8-bit addition and subtraction: The 8085 can execute two types of addition / subtraction instructions – addition or subtraction with out carry and addition or subtraction with carry. Both addition instructions can be used to add the contents of accumulator to the contents of one of the general purpose registers or the contents of memory locations (address in H – L pair) or all immediate data byte.

a) These instructions assume that the acc. is one of the operands.
b) Place the result in the acc.
c) Do not affect the contents of another operand and
d) Modify all the flags according to the data condition in acc.

e.g.1: ADD B – Adds the contents of register B with the contents of the accumulator and the result is stored in the acc. If a=05H and B-0 7H ,execution of ADD B modifies the content of A as 0CH. Reg B content is not altered. As 8085 handles data in binary in the display we see the hexadecimal equivalent of result OCH .Don't expect the decimal answer of 12.

e.g.2: ADD M – The contents of the memory location addressed by the pair is added to the contents of acc. one result is stored in the acc.

e.g.3: ADI 35H – The 8-bit immediate data 35H is added to the contents of the accumulator and the result is stored in acc.

e.g.4: ADC D – The contents of register D and carry flag are added to the contents of acc. and result is stored in the acc.

e.g.5: ADC M – The contents of memory location addressed by H – L and carry are added to the contents of the acc and the result is stored in the acc.

e.g.6: ACI 35H – 35H with carry are added to the contents of acc.

e.g.7: SUB B – The contents of register B is subtracted from the contents of the acc. and the result is stored in the acc.

e.g.8: SBB M – The contents of the memory location addressed by H – L pair and carry are subtracted from the contents of acc. And the result is stored in the acc.

ii) 16-bit addition:

This instruction adds 16-bit data of H – L register to the 16-bit data of another 16-bit register contents. Result is stored in HL pair. Except carry no other flags are effected.

DAD rp: Add the content of register pair specified to the content of H-L pair and store the result in H-L pair.

rp = B-C / D-E / H-L / sp.

The contents of register pair r_p are added to the contents of H – L pair and the result is stored in H – L pair.

e.g.: DAD D – The contents of register pair D – E are added to the contents of HL pair and the result is stored in H – L pair. If the result is more than 16-bits carry flag is set. No other flags are affected.

iii) 8-bit increment/decrement:

These instructions increment or decrement the contents of 8-bit registers or the contents of memory location whose address is specified in H – L register pair.

Carry flag is not affected.

e.g.: INR D – The contents of register D is incremented by one.

INR M – The contents of memory location addressed by H – L pair is incremented by one.

DCR C – The contents of register C is decremented by one.

DCR M – The contents of the memory location addressed by H – L pair is decremented by one.

iv) 16-bit increment/decrement:

1N x rp: Increment register pair contents

rp = B – C / D – E / H – L / sp.

e.g.: 1NX H – The contents of the H – L pair is incremented by one. No flags are affected.

DCX rp: decrement register pair contents.

rp = B – C / D – E / H – L / sp.

e.g.: DCX B – The contents of the B – C pair is decremented by one. No flags are affected.

However it becomes necessary in some programs to check whether the execution of DCX rp resulted in zero or not . After each decrement by checking the logical OR result of the registers involved one can ascertain it See how this is done in program no. .

v) Decimal adjusting data:

DAA: Decimal adjusting accumulator.

DAA instruction is used in the program after an addition instruction. The DAA instruction operates on the result in the accumulator and the final result in decimal form. It uses carry and auxiliary carry for decimal adjustment. To use the DAA instruction the two numbers added should be in BCD form.

13.7.3 iii) Logical Instruction:

These instructions perform logical operations such as AND, OR, EX-OR, compare, rotate and complement of data in register or memory.

1) Logical OR, AND, NOT and EX-OR operations.

a) These instructions implicitly assume that the accumulator is one of the operands
b) Place the result in accumulator.
c) Do not affect the contents of the operand reg. (except NOT operation)
d) Some of the flags are affected according to the result.

e.g.1: ORA D – The contents of the register D is logically OR ed with the contents of the accumulator and the result is stored in the accumulator.

e.g.2: ANI 23H – The data 23H is logically AND ed with the contents of the accumulator and the result is stored in the accumulator.

e.g.3: XRA M – The contents of the memory location addressed by H – L pair is EX – OR ed with the contents of the accumulator and the result is stored in the acc.

e.g.4: CMA – The contents of the accumulator is complemented and the result stored in accumulator.

2) Comparison:

The contents of the accumulator is compared with the contents of the other register or the contents of the memory location addressed by H – L pair or an 8-bit data.

CMP R – Compare register with accumulator

CMP M – Compare memory contents with accumulator.

CPI 8-bit data – Compare immediate data with accumulator

Here comparison is performed through subtraction. However, neither contents are modified, but the result of the subtraction is only reflected in the form of conditional flags.

i) If the content of A is lesser than compared data , carry flag set 1.
ii) If the content of A is greater than compared data, carry flag is zero.
iii) If both are equal, zero flag set to 1.

3) Rotate Instructions:

a) The rotate instructions are register specific, operates only on the accumulator
b) In these instructions, the bit that is shifted out is used as the new bit shifted in

c) There are four rotate instructions to rotate a bit pattern in the accumulator to the left or to the right with or without carry.

RLC – rotate accumulator left

RRC – rotate accumulator right

RAL – rotate all (including carry) left

RAR – rotate all (including carry) right.

13.7.4 iv) Branching Instructions:

The flow of a program proceeds sequentially, from instruction to instruction, unless a control transfer command is executed. The branch instructions allow microprocessor to go to a different memory location, either unconditionally or conditionally (under certain test conditions) and μρ continues executing instructions from that new location.

The branch instructions are classified in three categories:

- Unconditional/conditional jump instructions.
- Unconditional / conditional call and return instructions
- Restart instructions.

Jump Instructions:-

a) Unconditional jump

JMP addr – The program counter is loaded with 16-bit address and the program execution jumps to that address unconditionally.

b) Conditional jump

conditional jump instructions first check for the condition and if the condition is satisfied then only execution jumps to new location. Flag(I), sign flag(S), and purity flag (P).

e.g.1: JC addr – Jump on Carry

JNZ addr – Jump on No zero

JPE addr – Jump on even parity.

Call and Return instructions:-

The call instruction is used in the main program to call a subroutine, and the return instruction is used at the end of the subroutine to return to the main program. When a subroutine is called, the contents of the program counter, which is the address of the instruction following the call instruction, is stored in the stack and the program execution is transferred to the subroutine address. When the return instruction is executed at the end of the subroutine, the memory address stored in the stack is retrieved, and the sequence of execution is resumed in the main program.

a) Unconditional call

CALL addr – unconditional call

b) Conditional call:

CC addr – call if the carry flag set

CNC addr – call of the carry flag not set

CNF addr – call on No zero

a) Unconditional return

RET – unconditional return

b) Conditional return

RC – Return on carry

RNC – Return on No carry

RP – Return on positive

Restart Instructions:

Restart is a one byte CALL instruction. Because these instructions are having specific call addresses for specific restart instructions as specified bellow. The program jumps to the instruction starting at restart location.

Restart Instruction	Location
RST 0	0000H
RST 1	0008H
RST 2	0010H
RST 3	0018H
RST 4	0020H
RST 5	0028H
RST 6	0030H
RST 7	0038H

One interesting use of RST instruction is to create break point. Finding mistakes(bugs) in a big program can be made simple by inserting break points at desired places. Program terminates when RST instruction is executed. If there are no bugs up to that point, the break point can be shifted to some other place . The process of removing bugs is called debugging. Instead of HALT instruction one may use Break point for the safe transfer of control to monitor program.

13.7.5 v) Stack, I/o and machine – control group:

a) Stack instructions

PUSH rp – PUSH register pair onto stack

rp = B – C / D – C / H – C / psw.

The content of the specified register pair is pushed onto stack after the stack pointer is decremented by two.

e.g.: PUSH B – The contents of register pair B – C is pushed on to stack.

POP rp – POP off stack to register pair.

rp = B – C / D – C | H – C | psw.

The top two elements of the stack are POPed into the specified register pair. The stack pointer is incremented by two.

e.g.: POP D – The contents of stack top POPed into register pair D – E.

In some programs the available internal registers may not be sufficient to write the program. However some of the registers may not be needed immediately. In such situations the contents of immediately not used register contents are pushed on to the stack. These registers are used for the present need and once the purpose is served, they are reloaded with their earlier values with POP instruction. If a series of push instructions like PUSH B, PUSH D, PUSH H are used, to restore their original contents the POP instructions must be in the order POP H, POP D, POP B as stack operations follow first in last out policy.

b) I/O Instructions

IN port-addr – Input to accumulator from input port

The data available on the port is moved into the accumulator. Port address is 8-bit.

OUT port-addr – Out put from accumulator to output port. The contents of the accumulator is moved to the port whose 8-bit address is specified in the instruction

External devices with suitable interfacing can be connected to input and output ports. The power of microprocessor lies in its ability communicating with external devices. Industrial digital control systems, Robots and other smart devices work under microprocessor control. All together

256 I/O devices can be connected . The devices which use IN and OUT instruction for data transfer are said to be operating in I/o mapped I/o mode. If more than 256 devices are to be connected , Memory mapped I/O mode can be used. In this mode each device is assigned with a 16 bit address.

c) Machine control instructions

These instructions are used to control the microprocessor operation.

e.g.: HLT – Halt

The execution of the HLT instruction stops the microprocessor.

8085 Instruction set

Data transfer group

MOV r1,r2 Cycles:1 States:4 Flags: none
MOV r, M Cycles:2 States:7 Flags: none
MOV M, r Cycles:2 States:7 Flags: none

MOV	B,B 40	MOV	C,B 48	MOV	D,B 50	MOV	E,B 58
	B,C 41		C,C 49		D,C 51		E,C 59
	B,D 42		C,D 4A		D,D 52		E,D 5A
	B,E 43		C,E 4B		D,E 53		E,E 5B
	B,H 44		C,H 4C		D,H 54		E,H 5C
	B,L 45		C,L 4D		D,L 55		E,L 5D
	B,M 46		C,M 4E		D,M 56		E,M 5E
	B,A 47		C,A 4F		D,A 57		E,A 5F

MOV	H,B 60	MOV	L,B 68	MOV	M,B 70	MOV	A,B 78
	H,C 61		L,C 69		M,C 71		A,C 79
	H,D 62		L,D 6A		M,D 72		A,D 7A
	H,E 63		L,E 6B		M,E 73		A,E 7B
	H,H 64		L,H 6C		M,H 74		A,H 7C
	H,L 65		L,L 6D		M,L 75		A,L 7D
	H,M 66		L,M 6E		M,M 76		A,M 7E
	H,A 67		L,A 6F		M,A 77		A,A 7F

MVI r, data Cycles:2 States:7 Flags: none
MVI M, data Cycles:3 States:10 Flags: none
LXI rp, data 16 Cycles:3 States:10 Flags: none

Move immediate	Load immediate
MVI A,data 3E	LXI B,data 01
B,data 06	D,data 11
C,data 0E	H,data 21
D,data 16	SP,data 31
E,data 1E	
H,data 26	
L,data 2E	
M,data 36	

Load / Store	Opcode	Cycles	States	Flags
LDAX B	0A	2	7	none
LDAX D	1A	2	7	none
LHLD addr	2A	5	16	none
LDA addr	3A	4	13	none
STAX B	02	2	7	none
STAX D	12	2	7	none
SHLD addr	22	5	16	none
STA addr	32	4	13	none
XCHG	EB	1	4	none

Arithmetic group

	Cycles	States	Flags
ADD r, ADC r, SUB r, SBB r	1	4	All
ADD M, ADC M, SUB M, SBB M	2	7	All
ADI data, ACI data, SUI data, SBI data	2	7	All
INR r, DCR r,	1	4	Z, S, P, AC
INR M, DCR M	3	10	Z, S, P, AC
INX rp, DCX rp	1	6	none
DAD rp,	3	10	CY
DAA	1	4	All

ADD		ADC		SUB		SBB	
B	80	B	88	B	90	B	98
C	81	C	89	C	91	C	99
D	82	D	8A	D	92	D	9A
E	83	E	8B	E	93	E	9B
H	84	H	8C	H	94	H	9C
L	85	L	8D	L	95	L	9D
M	86	M	8E	M	96	M	9E
A	87	A	8F	A	97	A	9F

INR		DCR		INX		ADI		
A	3C	A	3D	B	03	data		C6
B	04	B	05	D	13	ACI	data	CE
C	0C	C	0D	H	23	SUI	data	D6
D	14	D	15	SP	33	SBI	data	DE
E	1C	E	1D					
H	24	H	25	DCX		DAA		27
L	2C	L	2D	B	0B	DAD	B	09
M	34	M	35	D	1B	DAD	D	19
				H	2B	DAD	H	29
				SP	3B			

Logical group

	Cycles	States	Flags
ANA r, ORA r, XRA r, CMP r	1	4	All
ANA M, ORA M,XRA M, CMP M	2	7	All
ANI data, ORI data, XRI data, CPI data	2	7	All
RLC, RRC, RAL, RAR	1	4	CY
CMA	1	4	none
CMC, STC	1	4	CY

ANA		XRA		ORA		CMP	
B	A0	B	A8	B	B0	B	B8
C	A1	C	A9	C	B1	C	B9
D	A2	D	AA	D	B2	D	BA
E	A3	E	AB	E	B3	E	BB
H	A4	H	AC	H	B4	H	BC
L	A5	L	AD	L	B5	L	BD
M	A6	M	AE	M	B6	M	BE
A	A7	A	AF	A	B7	A	BF

ANI	data	E6	RLC	07	DAA	27	
XRI	data	EE	RRC	0F	CAN	2F	
ORI	data	F6	RAL	17	STC	37	
CPI	data	FE	RAR	1F	CMC	3F	

Branch CONTROL Instructions

			Cycles	States	Flags
JMP	addr	C3	3	10	none
JNZ	addr	C2	2/3	7/10	none
JZ	addr	CA	2/3	7/10	none
JNC	addr	D2	2/3	7/10	none
JC	addr	DA	2/3	7/10	none
JPO	addr	E2	2/3	7/10	none
JPE	addr	EA	2/3	7/10	none
JP	addr	F2	2/3	7/10	none
JM	addr	FA	2/3	7/10	none
PCHL		E9	1	6	none

			Cycles	States	Flags
CALL	addr	CD	5	18	none
CNZ	addr	C4	2/5	9/18	none
CZ	addr	CC	2/5	9/18	none
CNC	addr	D4	2/5	9/18	none
CC	addr	DC	2/5	9/18	none
CPO	addr	E4	2/5	9/18	none
CPE	addr	EC	2/5	9/18	none
CP	addr	F4	2/5	9/18	none
CM	addr	FC	2/5	9/18	none

RETURN		Cycles	States	Flags
RET	C9	3	10	none
RNZ	C0	1/3	6/12	none
RZ	C8	1/3	6/12	none
RNC	D0	1/3	6/12	none
RC	D8	1/3	6/12	none
RPO	E0	1/3	6/12	none
RPE	E8	1/3	6/12	none
RP	F0	1/3	6/12	none
RM	F8	1/3	6/12	none

RESTART		Cycles	States	Flags
RST	0	C7	3	12	none
RST	1	CF	3	12	none
RST	2	D7	3	12	none
RST	3	DF	3	12	none
RST	4	E7	3	12	none
RST	5	EF	3	12	none
RST	6	F7	3	12	none
RST	7	F8	3	12	none

STACK operation		Cycles	States	Flags
PUSH	B	C5	3	12	none
PUSH	D	D5	3	12	none
PUSH	H	E5	3	12	none
PUSH	PSW	F5	3	12	none
POP	B	C1	3	12	none
POP	D	D1	3	12	none
POP	H	E1	3	12	none
POP	PSW	F1	3	12	none
XTRL		E3	5	16	none
SPHL		F9	1	6	none

I / O and Machine control

INPUT / OUTPUT			Cycles	States	Flags
OUT	data	D3	3	10	none
IN	data	DB	3	10	none

CONTROL

DI	F3	1	4	none
EI	FB	1	4	none
NOP	00	1	4	none
HLT	76	1	5	none
SIM	30	1	4	none
RIM	20	1	4	none

13.7 Summery:

The central processing unit of a digital computer built into a single LSI or VLSI chip is called a microprocessor. It is the latest development in the field of computer technology. A VLSI chip contains more than 10,000 transistors. Intel 8085 is an 8-bit NMOS microprocessor. This means the transistors used in the manufacturing of 8085 IC are N-channel Metal Oxide Semiconductor Field Effect Transistors. The 8085 IC has 40 pins and operates on a single +5V DC power supply. Its clock speed is 3MHz. Its memory capacity is 64k Bytes. In the IC, there are 4 functional units:

Timing and control unit, to send central signals and synchronize the μp operations with the clock; ALU, to perform arithmetic-logic operations on the data given, Interrupt control unit, to interrupt the execution of program; and serial I/O central unit, to content, the serial form of data into 8-bit, parallel data and vice versa. To reflect the data conditions, 5 Flags are used.

In order to use the microprocessor for a particular application, it has to be connected to various devices. To communicate with the outside world or internal parts, a Microprocessor uses 3 buses namely: address bus, data bus and control bus. The address bus is unidirectional, data bus is bi-directional. Some lines in control bus are input lines and others are output lines. The internal architecture of 8085 Microprocessor determines how and what operations can be performed with data. The operations are: To transfer 8-bit data, to perform arithmetic and logic operations, to text, for a condition and to sequence the execution of instructions.

The 8085 Microprocessor instruction set, has 74 operation codes that result in 246 instructions. The set includes data transfer, arithmetic, logic, branching and machine control instructions. All these instructions may be either one byte, two byte or three byte long. In the instructions, the source and destinations are operands. There are various ways or formats of specifying the operands, which are called the addressing modes. The 8085 microprocessor has 5 types of addressing modes namely., direct, Register, Immediate, Register indirect and Implicit.

13.8 Key terminology:

LSI – Large scale Integration: The process of designing more than a thousand gates on a single semiconductor chip. Similarly VLSI stands for very large scale Integration.

Unidirectional: If the bits flow in one direction, i.e. from the microprocessor to the devices, then the operation is called unidirectional.

Bi-directional: If the bits follow in both directions, then the operation is called Bi-directional. 8085 data bus is Bi-directional.

Interrupt: To break the execution of program temporarily.

Flag: It is a Flip-flop which keeps track of changing conditions.

Flip-flop: It is bi-stable multi-vibrator exhibiting two states 0 and 1.

Program-center: It is a 16-bit register that deals with the operation of sequencing and execution of instructions.

Buffer: It, is a logic circuit which amplifier the current or power.

DMA: Direct memory access. i.e. accessing the memory directly, by bypassing the microprocessor , with I/O devices.

13.9 Self-assessment questions:

1) Define op code, operand and specify the operand in the instruction ADD B.
2) Classify the instructions as per the length of instructions and give examples.
3) Explain the function of ALE and IO/M signals. of 8085 microprocessor.
4) What are the functions of serial I/O control unit and timing and control unit?.
5) Why data bus is bi-directional in 8085 microprocessor?. Explain.
6) Explain few branching instructions with an example.

13.10 Reference books:

1) Microprocessor Architecture, Programming and Applications – Gaonkar (Wiley eastrun)
2) Microprocessor Theory and Applications – M.Rafiquzzaman (P.H.I.)
3) Digital Computer Electronics – Albert Paul Malvino (TMH)
4) Fundamentals of microprocessors and Microcomputers – B.Ram (Dhanpat Rai & Sons)

Name of the Author

Dr. B.V.S.Goud, M.Sc., Ph.D.,
Associate Professor and Head,
Dept. of Electronic and Instrumentation Technology,
Acharya Nagarjuna University,
Nagarjuna Nagar – 522 510

UNIT - IV

Lesson – 14

8085 Assembly Language Programming and Developments in Microprocessors

Objectives:

1) To understand what is an assembly language.
2) To know, what is an assembly language (AL) program.
3) To know how to write an AL program and how to execute it.
4) To learn about flow charting
5) To know the latest developments.

Structure:

14.1 Introduction:

Microprocessors are programmable devices like computers. In fact they are the CPUs of present day personal computers. For computers, high level languages are preferred. High level languages, how much versatile they are, do not allow the users to middle with hardware. So to deal with hardware and to create code to produce special effects needed by the individuals it becomes necessary to program the processor in its machine code. Machine codes are binary codes written in terms of 1s and 0s only. As it is very difficult to understand binary codes, English like assembly language were developed. Assembly language usually consists of mnemonics and hexadecimal numbers. Assemblers which convert assembly language programs written in terms of mnemonics and hexadecimal code to machine language also were developed. Assembly language programs can

also be written manually on paper and entered in the program area of memory of a microprocessor system. One disadvantage with assembly language is that it is not portable. It means programs written in the code of one processor cannot be used for another processor. If two 8085 systems differ in their memory map, then also the code cannot be transported and run directly.

INTEL developed microprocessors which have 8086 mode of working. Assembly language programs written with 8086 can be run on all upgraded versions of 8086, above 8086 to Pentium. Writing ALP for 8086 is difficult for the biginners.To understand programming techniques for writing ALPs, 8085 is convenient. In this lesson details about assembly language programming of 8085 are given.

Program is a set of instructions written in a specific sequence to accomplish a task.

A set of programs written for a particular computer is known as software for that computer. A program written using `0's and '1's is called a machine language program. If the program is written using mnemonics, then it is called Assembly Language (AL) program. Here the term mnemonic represents the shorthand form of English words for the operation performed by the instruction. For example ADD, SUB, MOV are the mnemonics for the addition, subtractions, movement of data respectively. The AL program written using mnemonics is then converted into machine language either by manual translation with the help of binary for each instruction given in the manufacturer's data books or by using a special program called assembler. The assembler is a program, that translates an AL program written mnemonics into a machine language program. The assembler performs various functions such as error checking and memory allocation.

14.2 Writing and executing an AL program:

In writing an Assembly language program to solve a particular problem, first step is to develop an algorithm, next draw a flow chat and finally the actual AL program to solve the problem.

The algorithm is a precise statement of the procedure received for solving the problem. It may be expressed using the English language. A flow chart is a pictorial representation of steps necessary to write the program. Various symbols are used in flow chart representation. They are: circle with an arrow, rectangle, diamond, oval, double-sided rectangle. Some of the common symbols used in flow charts are shown in Fig 14.10. From flow charts one can visualize the program flow and can identify the flaws in the algorithm. Based on flowchart, algorithm is revised

and again the flow chart is redrawn for consistency. After satisfying oneself that both flow chart and algorithms are in harmony, one starts writing assembly language programs which are directly hand coded in to machine language or converted in to machine code using an Assembler.

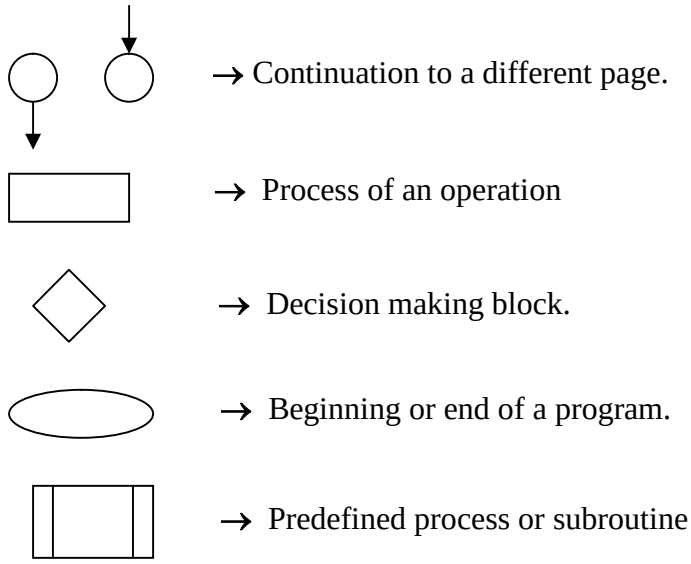


Fig 14.1 Flow chart symbols.

A standard AL program statement consists of four fields namely: Label, op code, operand and comment.

A label field represents the address, the op code field contains the mnemonics of the instruction, the operand field contains the data and comment field contains comments related to that operation.

EXECUTING A PROGRAM;- To manually write and execute an AL program on a Micro processor trainer kit with a hexadecimal (hex) keyboard for input, we follow the steps given below. An AL program written for specific task can be executed and tested for its validity in the following steps.

(i) Write an AL program using algorithm, flow chart and instructions.

(ii) Write the hexadecimal (hex) code for each instruction by searching through the instruction set.

(iii) Enter the program in the user memory (RAM) in a sequential order by using keyboard as the input device.

(iv) Execute the program by pressing the execute key. The answer will be displayed on the seven segment display.

14.3 Illustrative programs:

Example 1: To give a better idea of how the data transfer group of instructions are used in programs, we will now write a simple program to input a value from the key board, store this value in memory location and display it. Here we assume that the key board is connected through input port 1 whose address is 05 H, display is connected through output port of address 02H and memory location address is 1050H.

Solution :

Algorithm for the above program:

1. Input a value from key board connected to the port at address 05H.
2. Store the value in memory location 1050H.
3. Output the value to a display connected to the port at address 02H.

Assembly language program:

Table 14.1

Label	Mnemonics		Comment
	Op code	Operand	
	IN	05H	; Input from port 05H
	STA	1050H	; Store in memory
	OUT	02H	; Send it to display
	HLT		; End of the program

Example 2:-

To use the arithmetic group and branching group of instructions and to have better concept about counters and loops, we will write a program to find the sum of a series of numbers stored in the memory locations, starting from 0850H to 0856H. Store the 16-bit sum in 0860H and 0861H locations.

In this example, we use the concept of looping. Looping is a technique used to instruct the processor to execute certain instructions repeatedly. The number of repetitions are specified in a counter.

Flow chart:

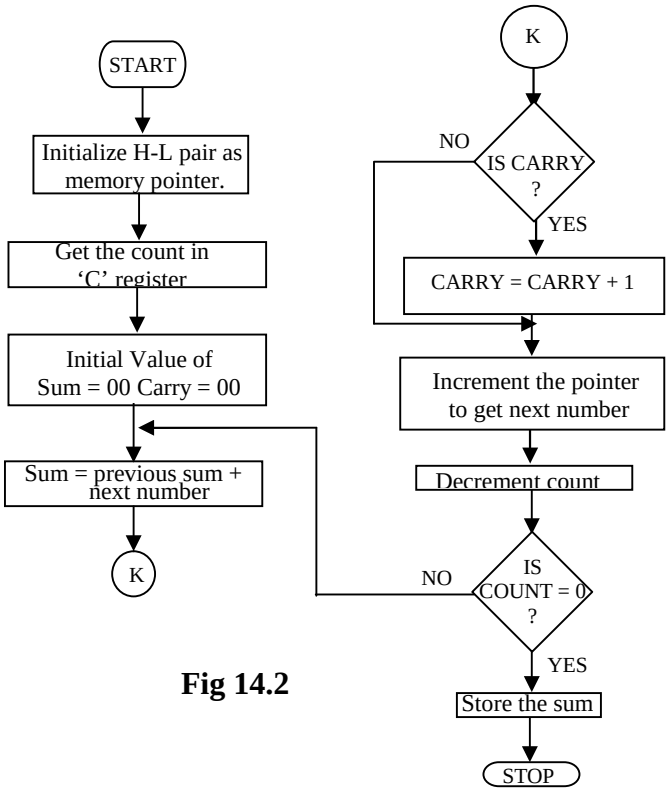


Fig 14.2

Algorithm:

1) Take initial sum as zero
2) Add the first number to the previous sum (initial sum)
3) Reduce the number of items by one.
4) Add the next number of the given series to the previous sum
5) Reduce the number of items by one
6) Repeat the steps (4) and (5) till all the numbers are added
7) Final sum is the result
8) Stop the process.

Program:-

Label	Mnemonics		Comments
	Op code	Operand	
	LXI	H, 0850	; Initialize the memory
	MVI	C, 07	; Get the count
	MVI	A, 00	; Clear A register
	MVI	B, 00	; Clear B register
Repeat:	ADD	M	; Add the number to previous sum
	JNC	Next	; Jump to 'next' if there is no carry
	INR	B	; Register carry
Next:	INX	H	; Increment the memory pointer
	DCR	C	; Decrements the counter
	JNZ	Repeat	; Jump to 'repeat' if counter is not zero.
	STA	0861	; Store the sum
	MOV	A, B	; Save the carry
	STA	0860	; Store the carry
	HLT		; End of the program

Table 14.2

The above program can be executed on a microprocessor kit, after converting the program into hex code and assigning one address to each byte of the program as in the previous example.

Example 3:- 8085 micro processor does not have a multiply instruction. So, when we want to multiply two numbers, we have to write a program for this. There are various algorithms for the multiplication of two numbers. Here we adopt a simple repeated addition algorithm .The method is to initialize a register with zero value and call it RESULT and go on adding one number to the RESULT second number times. For example if 5 is to be multiplied by 6 it can be done adding 6 to itself 5 times or adding 6 to itself 5 of times.

Algorithm:

1) Take the multiplicand (I number)
2) Take the initial sum as zero

3) Add the multiplicand to the previous sum
4) Decrement the multiplier
5) Repeat the steps (3) and (4) till the multiplier becomes zero
6) The final sum gives the products of two numbers
7) Stop the process

Flow chart:

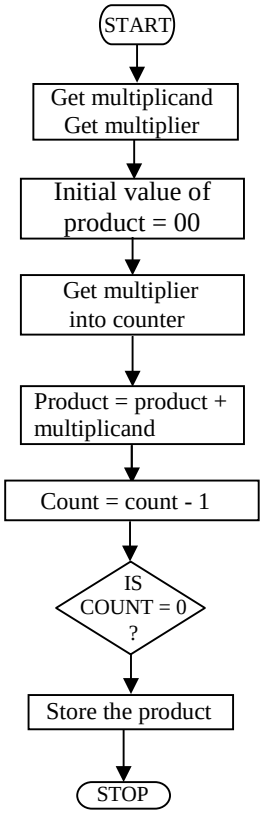


Fig 14.3

Let the numbers (multiplicand, multiplier) be stored at memory locations with the address 0850, 0851; let the result (product) be of 16-bit be stored at locations with the address 0852, 0853 (for higher and lower order byte of the product). Product is found by using repetitive addition.

Program:

Label	Mnemonics		Comments
	Op code	**Operand**	
	LXI	H, 0850	; Initialize the memory pointer
	MOV	B, M	; Get the multiplicand
	INX	H	; Increment the pointer
	MOV	C, M	; Get the multiplier
	SUB	A	; Make the initial product as 00
Next:	MOV	D, A	; Clear reg 'D' for carry
	ADD	B	; Add the multiplicand
	JNC	Loop	; Check for carry
Loop:	INR	D	; Register the carry
	DCR	C	; Decrement the multiplier
	JNZ	Next	; If (C) ≠ 0, go to 'next'
	INX	H	; Increment the pointer
	MOV	M, D	; Save the carry
	INX	H	; Increment the pointer
	MOV	M, A	; Save the sum
	HLT		; End of the program

Table 14.3

EXAMPLE 4:- To have the better concept of using compare instruction in finding the smaller or bigger or equality of two given numbers we write a program to find the largest number among the given set of n- numbers.

Let the list starts from a memory location with the address 0850H and let the largest number be stored at a memory location 0870H

Algorithm:

1) Take the first number from the given series
2) Compare this number, with the next number of given series
3) If the first number is larger than the next number then retain the first number
4) If the first number is smaller than the next number then retain the next number
5) Decrement the number of comparisons
6) Repeat the process of comparison till the largest number in the given list is found.
7) Stop the process.

Flow chart:

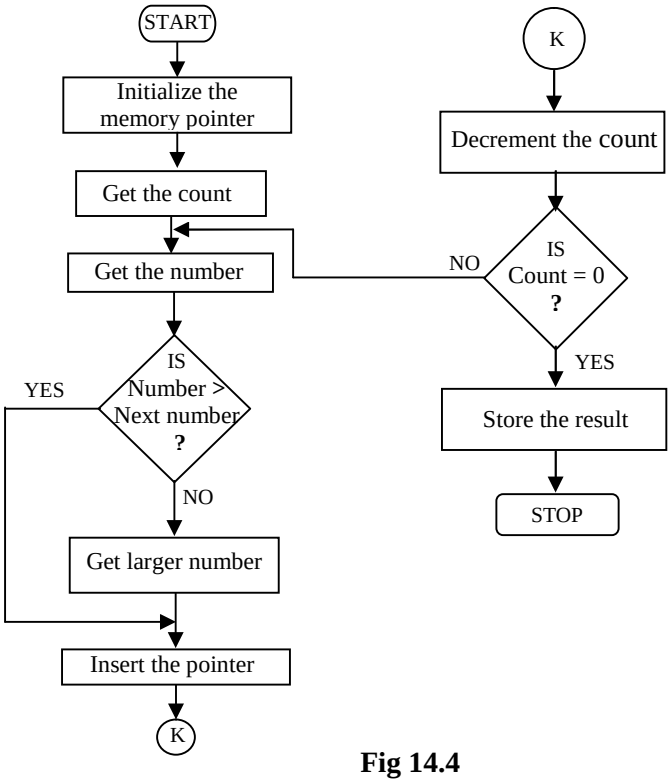


Fig 14.4

PROGRAM:

Label	Mnemonics		Comments
	Op code	**Operand**	
	MVI	C, n – 1	; Get the count
	LXI	H, 0850	; Initialize the memory pointer
	MOV	A, M	; Get the number
Next:	INX	H	; Increment the pointer
	CMP	M	; Compare with memory for largeness
	JNC	Loop	; If larger, jump to 'loop'
	MOV	A, M	; If smaller change with memory
Loop:	DCR	C	; Reduce the counter
	JNZ	Next	; If comparisons are not over, go to 'next'
	STA	0870	; Store the largest
	HLT		; End of the program

Table 14.4

Here, after converting the AL program into its hex codes and assigning with suitable addresses, the actual series of data items is to be to loaded into the memory, whose starting address is specified in the second instruction of above program. The method of loading the data items into the memory is same as that of loading the program. As an example, suppose the number of items , n is 8, and let the data items are

0A, 29, 02, 0B, 0E, 22, 39, 49

After executing the program using the above data, the largest number i.e. OE can be viewed at the specified memory location 0870.

Address	Data
0850	0A
0851	29
0852	02
0853	0B
0854	0E
0855	22
0856	39
0857	49

Result

Address	Data
0870	0E

Fig 14.4b

Note that the same program can be used to find the smallest number simply using JC instruction instead of JNC in the above program.

Example 5:. In most of the microprocessor systems, the key board input is in BCD format(in BCD format a decimal digit, 0 to 9, is represented by its 4-bit binary equivalent). But the processing of data inside of the microprocessor is performed in binary format. Hence there is a need to convert BCD number into its equivalent binary value. In doing so we appreciate the usage of logical AND and ROTATE instructions.

Algorithm:

1) Take the given 8-bit packed BCD number,
2) Separate the number into two 4-bit un packed BCD digits BCD, and BCD_2
3) Convert each digit into its binary value according to its position.
4) Add both binary numbers to obtain the binary equivalent of the given BCD number
5) Stop the process.

Flow chart:

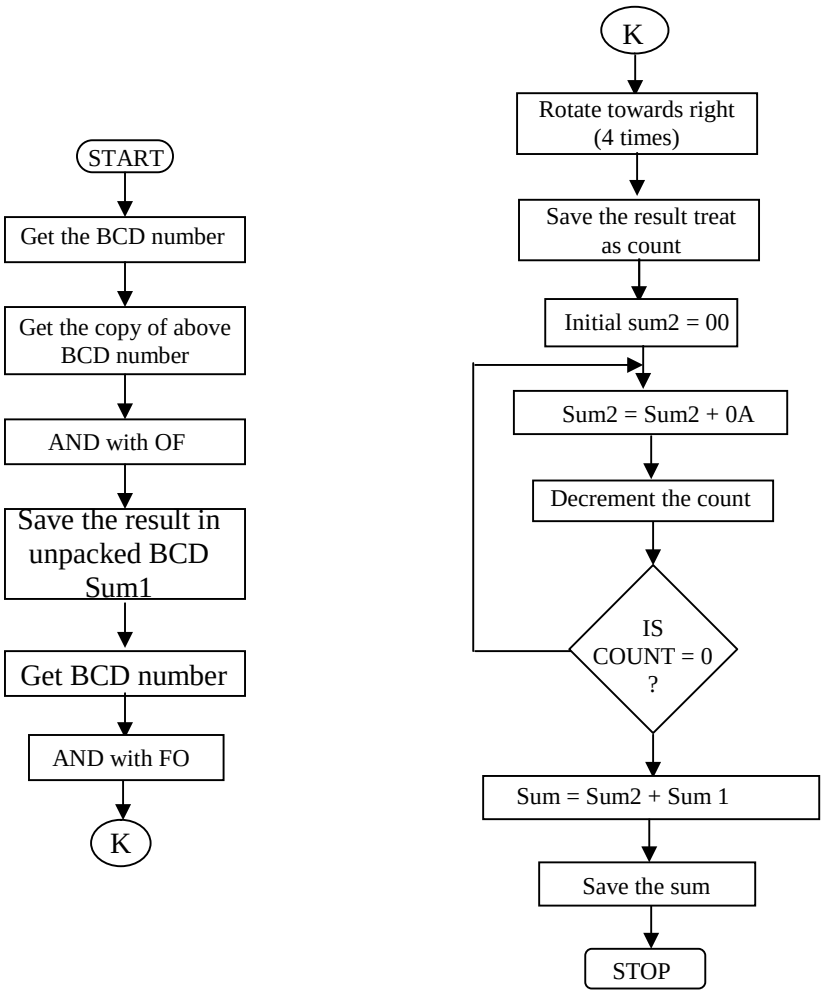


Fig 14.5

In this example we assume that the packed BCD number stored in memory location 1050H and we store the result in 1060H.

The conversion of a BCD number into its equivalent binary number employs the principle of positional weighting in a given number. For example in a number 96_{10}, the digit 9 represents 90 based on its second position from the right. So while converting 96_{BCD} into its binary equivalent requires the multiplication of 9 by 10 and adding the first digit of 6.

Label	Mnemonics		Comments
	Op code	Operand	
	LDA	1005H	; Get packed BCD number
	ANI	0FH	; Mask most significant four bits
	MOV	C, A	; Save unpacked BCD, in regC
	LDA	1050H	; Get BCD again
	ANI	FOH	; Mask least significant four bits
	RRC		
	RRC		
	RRC		
	RRC		
	MOV	D, A	; Save BCD_2 in D
	XRA	A	
	MVI	E, 0AH	
AGAIN:	ADE	E	
	DCR	D	
	JNZ		
	ADD	C	
	STA	1060H	
	HLT		

Table 14.5

Example 6: Accurate time delays between two events can be generated using assembly language programming. For example clock generators, digital clocks, flashing of lights etc require these time delays. The process of generating time delays using software instructions is more flexible than generation of time delays using hardware. To illustrate this , we write a program to generate time delay by using counters.

ALGORITHM:

1) Load a register with required delay.
2) Decrement the register.
3) Repeat it till register becomes zero.

4) Get out of the loop.

Flow chart:

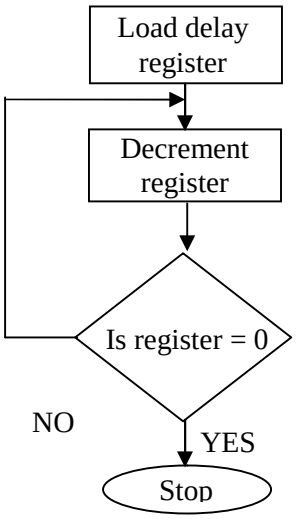


Fig 14.6

Label	Mnemonics		Comment
	Op code	Operand	
	MVI	C, N	; Load RegC with number N
BACK:	DCR	C	; Decrement C
	JNZ	BACK	;If C ≠ 0 Jump to BACK to decrement
	HLT		

Table 14.6

In the above program, the number N in register C determines the duration of the time delay. For the calculation of N, we should know the number of states required for the execution of each instruction and the number of times the instruction is executed. The number of states required can be found from the data supplied by the microprocessor manufacturer(see the instruction set given) and the number of times the instruction is executed is determined by the program as shown below.

Instruction	States	Total number of states
MVI C,N	7	7 x 1
DCRC	4	4 x N
JNZ	10/7	10(N-1)+7 x 1
HLT	5	5 x 1

$\therefore$ Total states = 7x1 + 4N + 10(N-1) + 7x1 + 5x1

= 7+4N+10N-10+7+5

= 14N + 4

Time for one state if the processor is operative with 2MHz clock = .5μs

Therefore Time delay = $.5*10^{-6}$ (N-1)

Using the above relation, the N value can be calculated for the required time delay.

Example 7:

To have the concept of subroutine, we write an assembly language program to switch ON and OFF an LED connected to an out port of the address 05H.

To visualize the ON and OFF states of LED we include a delay program in between these ON and OFF states.

Algorithm

1) Switch ON the LED
2) Insert delay
3) Switch OFF the LED
4) Insert delay
5) Loop continuously

Flowchart

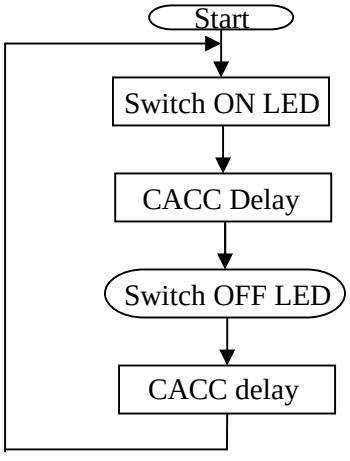


Fig 14.7

Program

```
START: MVI A, FFH
       OUT  O5H
       CALL DELAY
       MVI A, 00H
       OUT 05H
       CALL DELAY
       JMP START
```

Table 14.7a

```
Delay
  MVI B, FFH
Back: DCR B
      JNZ  Back
      RET
```

Table 14.7b

14.4 Developments in Microprocessors and Applications:

With the success of INTEL 8 bit Micro processors as programmable devices and CPUs of personal computers, several manufacturers produced a variety of Microprocessors. IBM , the Computer Giant was rather forced to enter in to PC market. To suit the computational needs INTEL produced the 8086, a 16 bit microprocessor in 1972 . It's data bus size is 16-bits. It can handle 16 bit data . It can directly access one mega byte of memory, because its address bus is 20-lines width (2^{20} = 1MB). It uses 16-bit data bus. The 8086 processor is available in a 40-pin DIP. A single +5V power supply is required. The clock input signal is generated by the 8284 clock generator / driver chip. Instruction execution times vary between 2 clock cycles and 30 clock cycles. The 8086 family consists of two types of 16-bit microprocessors, the 8086 and 8088. The main difference is how the processors communicate with the out side world. The 8088 has an 8-bit, data path to memory and I/O, while the 8086 has a 16-bit external data path. The 8088 was used in designing the IBM PC – XT personal computers.

Next, Intel introduced the high performance 80186 and 80188 processors, which are enhanced versions of the 8086 and 8088 respectively. These were followed by 80286 and 80386 and 80486 processors.

The 80286 is fabricated in a 68-pin package. The 80286 has added memory protection and management capabilities to the basic 8086 architecture. An external 82284 clock generator chip is required to generate the clock. The 80286 was used as the CPU of the IBM PC/AT personal computer. An enhanced version of the 80286 is the 80386 microprocessor. It is a 32-bit processor. The 80386 is used as the microprocessor in the IBM 386PC. The 80486 with its advanced features and support from Microsoft software survived long in the market. It has on-chip floating-point circuitry. It contained a circuitry of 2,75,000 transistors. It could directly address 4 giga bytes (2^{32}

bytes) of physical memory. With these processor based systems several new applications like desktop publishing, computer based digital communications were realized. INTEL introduced a new generation PC named Pentium. Even in this generation several versions are available and already up to Pentium III,excepting CELRON a locost version of Pentium II, all have become obsolete. Pentium IV is the current processor with a clock speed of 2.5GHz Even the Celron sytem was upgraded in clock speed to 2.8 GHz INTEL has introduced a new generation processor named Ethenium. The trend of developing new and powerful processors will go on.

Even other manufactures too introduced micro-processors. To quote some of them, Zilog introduced Z80, an 8-bit processor . Z8000 series 16-bit processors. Motorola introduced MC 6800 an 8-bit NMOS processor MC 68000 series 16-bit processors. Advanced Microdevices Inc is also producing processors which can compete with INTEL processors. INTEL processors have become so popular that several manufactures are fabricating clones to INTEL processors and IBM PC compatibles.

Applications:

Microprocessors find applications in the area of science, for the measurement and control of various physical parameters like temperature, pressure, PH value, humidity etc; in industries for process control; in day-to-day life for traffic control, for automatic lighting systems; in computers as their CPUs for data collection, analysis; in navigation, in aviation, in remote sensing, in satellite communications, in home appliances, in kid's toys, video games, in Robotics, in sun tracking systems, in data loggers for recording the data. Besides as CPUs, to suit dedicated applications Micro controllers, single chip computers and bit-slice systems were also developed.

14.5 Summary :

A microprocessor can be used for a specific application, only, when it is programmed properly. The microprocessor can be programmed using assembly language. The assembly language uses mnemonics to write the program. The AL program consists of statements. Each statement is to be written with four fields namely: Label, op code, operand and comment. Labels are to be followed with a colon, and comments are to be preceded with a semi colon. Before writing the AL program, the idea of solving the problem is to be first expressed in the form of algorithm. After this, the idea is to be interpreted in the form of flow chart. Finally, the AL program is to be written using the

mnemonics. In this chapter programming techniques-such as looping, counting and indexing were illustrated using memory related and date transfer instruction.

14.6 Key terminology:

Mnemonic: Symbolic code for each instruction.

Flow chart: The pictorial format, representing the thinking process and steps necessary to write the program.

Byte: A group of 8-bits

Assembly language: A medium of communication that, with a microprocessor in which programs are written in English-link words for representing binary instructions.

Assembler: A program that translates an AL program from mnemonics to the binary machine code.

Instruction cycle: Time required to complete the execution of an instruction

14.7 Self-assessment questions:

1) Differentiate between machine language and assembly language.
2) What is an assembler? What are its uses?
3) Name the four fields of standard AL program statement.
4) Write an AL program to find the largest between two given numbers.
5) Write an AL program to divide two 8-bit hexadecimal numbers using repetitive subtraction.
6) Write an AL program to count the number of 1 s in a given 8-bit number.

14.8 Reference books:

1) Microprocessor, Architecture, Programming and Applications – Gaonkar.
(Wiley Eastern)
2) Fundamentals of Microprocessors and Microcomputers – B. Ram.
(Dhanpat Rai & Sons Publications)

Name of the Author

Dr. B.V.S.Goud, M.Sc., Ph.D.,

Associate Professor and Head

Dept. of Electronic and Instrumentation Technology,

Acharya Nagarjuna University,

Nagarjuna Nagar – 522 510

UNIT IV

LESSON-15

INTEL 8086 MICROPROCESSOR

Objectives:-

This lesson explains you the

1. Description of 8086 Architecture
2. Various types of addressing modes
3. Instruction set of 8086 Microprocessor.

Structure of the lesson:

Section-I

SECTION-II

Section-I

8086 MICROPROCESSOR

15.1 Introduction:

The Intel 8086, a 16 – bit microprocessor, contains approximately 29,000 transistors and is fabricated using the HMOS technology. The expanded memory capability made multiprogramming feasible and several multiprogramming features have been incorporated into the 8086's design. The 8086 processor also includes a number of features which enhance its multiprocessing capabilities, thus allowing it to be used with other processing elements such as the 8087 numeric data processor.

15.1.1 8086 Pin Description:

A pin assignment diagram for the 8086 is given in Fig.15.1. The 8086 has 20 address pins, 16 of which are also used as data pins. The use of pins for both addresses and data means that both an address and datum cannot be sent to the system bus at the same time. There are 16 control lines for providing hand - shaking signals during bus transfer and for permitting at least some external control of the CPU. The 8086 requires only one supply voltage, +5V d.c. and one clock phase whose frequency can be up to 5MHz.

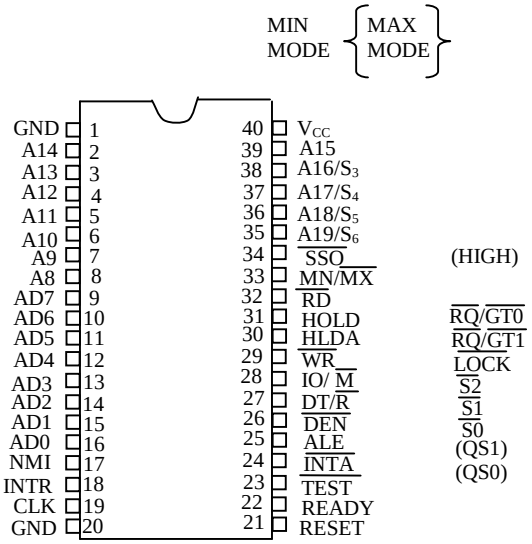


Fig. 15.1 8086 Pin – out diagram

The description of the pins of 8086 is as follows:

AD0-AD15. (Bi-directional) Address/Data lines. These are low-order address bus. They are multiplexed with data.

When AD lines are used to transmit memory address, the symbol A is used instead of AD, for example A0-A15. When data are transmitted over AD lines, the symbol D is used in place of AD, for example D0-D7, D8-D15 or D0-D15.

A16-A19(Output). High-order address lines. These are multiplexed with status signals.

A16/S3, A17/S4. A16 and A17 are multiplexed with segment identifier signals S3 and S4

A18/S5 A18 is multiplexed with interrupt status S5

A19/S6 A19 is multiplexed with status signal S6.

$\overline{\text{BHE}}$ **/S7 (Output)** Bus High Enable / Status. During T1 it is low. It is used to enable data onto the most significant half of data bus, D8-D15. 8-bit device connected to upper half of the data bus use $\overline{\text{BHE}}$ signal. It is multiplexed with status signal S7. S7 signal is available during T3 and T4.

$\overline{\text{RD}}$ **(Read).** This signal is used for read operation. It is an output signal. It is active when LOW.

READY (Input). The addressed I/O or memory sends acknowledgment through this pin. When HIGH, it indicates that the peripheral is ready to transfer data.

RESET (Input). System reset. The signal is active HIGH.

CLK(Input). Clock. 5,8 or 10MHz.

INTR (Interrupt request**)**

NMI(Input). Non - maskable interrupt request.

$\overline{\text{TEST}}$ **(Input).** Wait for test control. When it is low the microprocessor continues execution otherwise waits.

VCC. Power supply, +5V d.c.

GND. Ground.

15.1.2 Operating Modes of 8086:

There are two modes of operation for Intel 8086, namely the minimum mode and the maximum mode. When only one 8086 CPU is to be used in a microcomputer system the 8086 is used in the minimum mode of operation. In this mode, the CPU issues the control signals required by memory and I / O devices. In a multiprocessor system, it operates in the maximum mode. In case of maximum mode of operation, control signals are issued by Intel 8288 bus controller, which is used with 8086 for this very purpose. The level of the pin MN / $\overline{\text{MX}}$ decides the operating mode of 8086. When MN / $\overline{\text{MX}}$ is high, the CPU operates in the minimum mode. When it is low, the CPU operates in the maximum mode. Pins 24 to 31, issue two different sets of signals. One set of signals is issued, when the CPU operates in the minimum mode. The other set of signals is issued, when the CPU operates in the maximum mode. Thus the pins from 24 to 31 have alternate functions.

15.1.3 Pin Description for Minimum mode:

For the minimum mode of operation the pin MN / $\overline{\text{MX}}$ is connected to 5V d.c. supply, i.e. MN/ $\overline{\text{MX}}$ = V_{CC}. The description of the pins from 24 to 31 for the minimum mode is as follows:

$\overline{\text{INTA}}$ **(Output).** Pin No.24. Interrupt acknowledge. On receiving interrupt signal, the processor issues an interrupt acknowledge signal. It is active LOW.

ALE(Output). Pin No.25. Address latch enable. It goes HIGH during T1. The microprocessor sends this signal to latch the address into the Intel 8282 / 8283 latch.

$\overline{\text{DEN}}$ **(Output).** Pin No.26. Data enable. When Intel 8286 / 8287 octal bus transceiver is used, this signal acts as an output enable signal. It is active LOW.

DT/R (Output). Pin No.27. Data Transmit/Receive. When Intel 8286/8287 octal bus transceiver is used, this signal controls the direction of data flow through the transceiver. When it is HIGH, data are sent out. When it is LOW, data are received.

M/$\overline{IO}$ (Output). Pin No.28. Memory or I/O access. When it is HIGH, the CPU wants to access memory. When it is LOW, the CPU wants to access I/O device.

$\overline{WR}$ **(Output).** Pin No.29. Write. When it is LOW, the CPU performs memory or I/O write operation.

HLDA (Output). Pin No.30. HOLD acknowledge. It is issued by the processor when it receives HOLD signal. It is active HIGH signal. When HOLD request is removed, HLDA goes LOW.

HOLD (Output). Pin No.31. HOLD When another device in microcomputer system wants to use the address and data bus, it sends a HOLD request to CPU through this pin. It is an active HIGH signal.

15.1.4 Pin Description For Maximum Mode

For the maximum mode of operation, the pin MN/$\overline{MX}$ is made LOW. It is grounded. The description of the pins from 24 to 31 is as follows:

QS1, QS0 (Output). Pin Nos.24,25. Instruction Queue Status logic are given below:

QS1	QS0	Mode of operation
0	0	No operation
0	1	1st byte of op code from queue
1	0	Empty the queue
1	1	Subsequent byte from queue

S0,S1,S2 (Output). Pin Nos. 26,27,28. Status signals. These signals are connected to the bus controller Intel 8288. The bus controller generates memory and I/O access control signals. Table 15.1 shows the logic for status signals.

Table 15.1 Logic for Status Signals

S2	S1	S0	Mode of operation
0	0	0	Interrupt acknowledge
0	0	1	Read data from I/O port
0	1	0	Write data into I/O port
0	1	1	Halt

15.1.5 CPU Architecture:-

Fig 15.2 shows the internal architecture of the 8086. the 8086 CPU is divided into two independent functional units.

1. Bus Interface unit
2. The execution unit

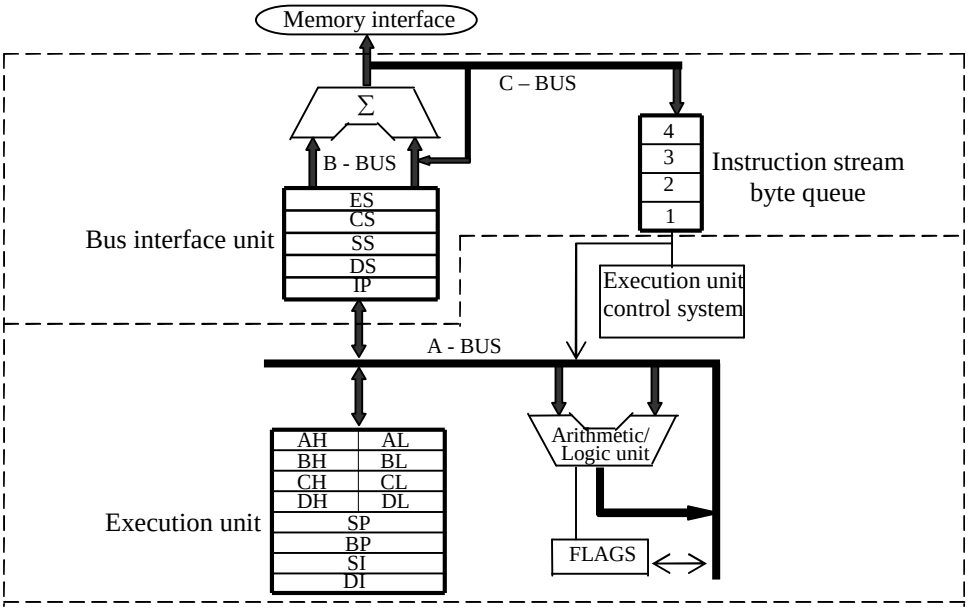


Fig 15.2 CPU Architecture

15.1.6 The Bus Interface Unit:-

The BIU sends out addresses, fetches instructions from the memory, reads data ports and memory and writes data to ports and memory. The following section describes the various functional parts of the BIU.

(a) The Queue:-

The main work of the BIU is to fetch the instruction / data bytes from the memory and pass them on to the execution Unit (EU) for further decoding and executing. These data bytes are fetched in advance and stored in a first-in-first-out like data structure, called Queue, while the buses are not being used by the EU. When the EU is ready for the execution of the next instruction, instead of fetching the byte from the memory it simply reads the byte from the Instruction Queue. As the time required to access register is less than the time required to access the memory, it increases the over all processing speed of the microprocessor. This feature is called Pipelining.

(b) The Segment Registers:

The address of the memory bytes that need to be accessed is generated with the help of segment registers. The BIU contain four sixteen bit registers, viz., the code segment, the data segment, the stack segment and the extra segment. These segment registers are used to hold the upper 16 bits of the starting address of the logical group of memory, called the segment, that 8086 is working with at a particular time.

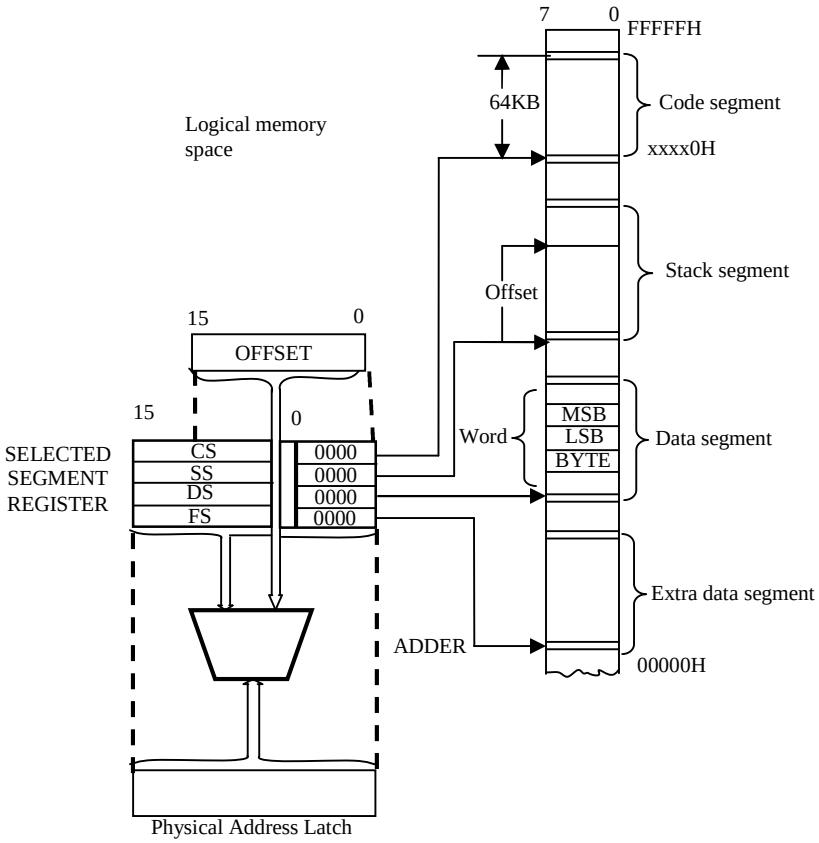


Figure 15.3 The 8086 / 88 memory segmentation

In the 8086 microprocessor, the total memory is divided into four segments. Each segment is of 64K bytes in size. Each of these segments can be used for a specific function. The code segment is used for storing the instructions, the stack segment is used as a stack, and data and extra segments are used for storing the data bytes.

The advantages of using segment registers are that they

i) allows the memory capacity to be one megabyte even though the address associated with the individual instructions are only 16 bits wide.

ii) allows the instruction, data or stack portion of a program to be more than 64K bytes long by using more than one code, data or stack segment.

iii) facilitate the use of separate memory area for a program, its data, and the stack.

iv) permit a program and / or its data to be put into different areas of memory each time the program is executed.

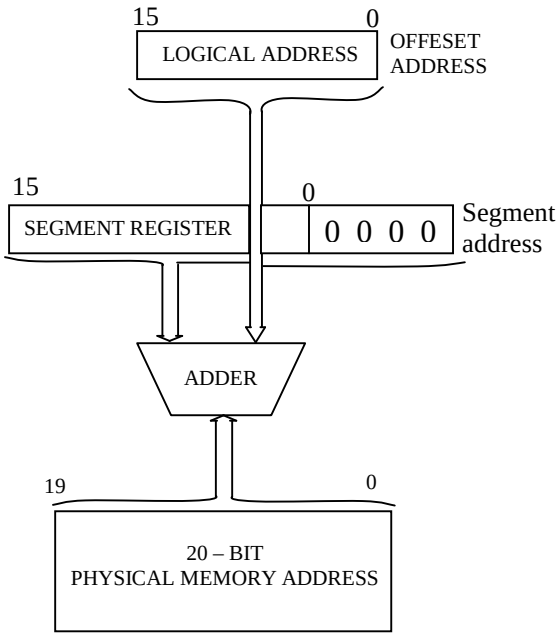


Fig. 15.4 Generating the 20 – bit physical address

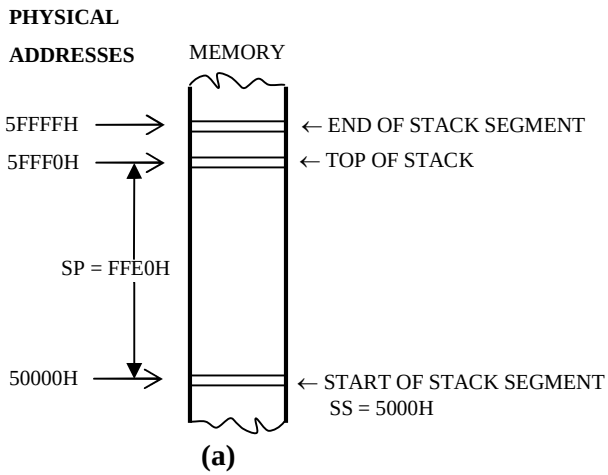
The starting address of the segments which are currently active, or in use, at any time, is stored in the respective segment register. As the segment register is of only 16 bytes, while the address width is 20 bytes, therefore to get the complete starting address of the segment, the segment register is multiplied by 16. The physical address is calculated as follows:

Example 1:- (Refer to fig 15.5)

The value of the stack pointer (SP) in = FFE0H

The value of the stack segment register (SS) = 5000H

Physical address of the top of the stack in = 5000H x 16 + FEE0H = 5FEE0H



(a)

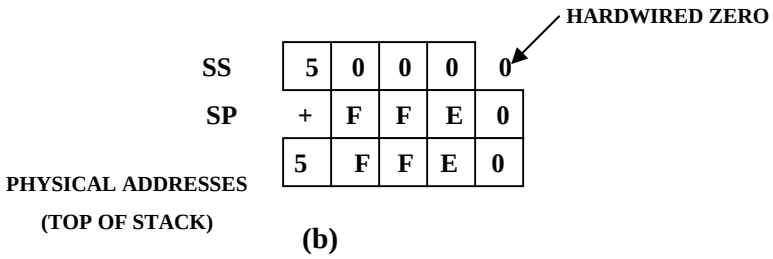


Fig 15.5Addition of SS and SP to produce the physical address of the top of the stack. (a) Diagram. (b) Computation.

Example 2:- (Refer to fig 15.6)

The offset of the data byte = 437AH

The value of the data segment register (DS) = 2000H

Physical address of the data byte = 2000H x 16 + 437AH = 2437AH

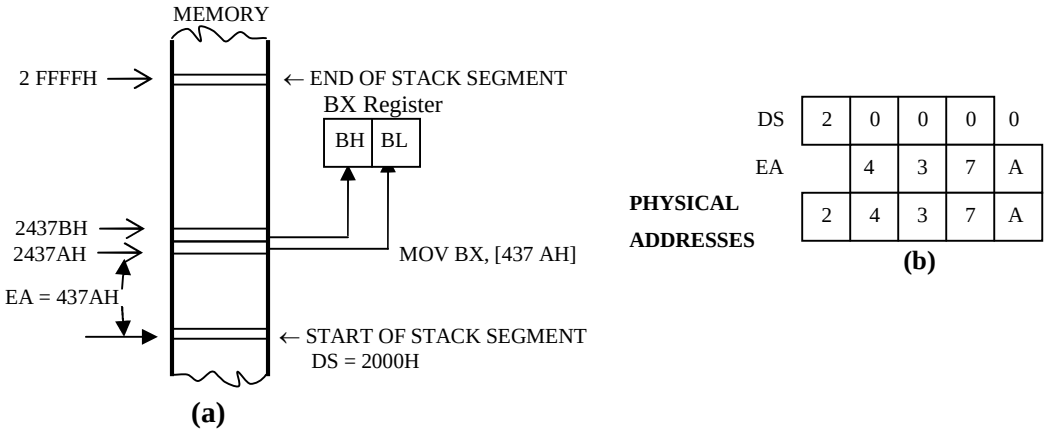


Fig 15.6 Addition of data segment register and effective address to produce the physical address of the data byte. (a) Diagram. (b) Computation.

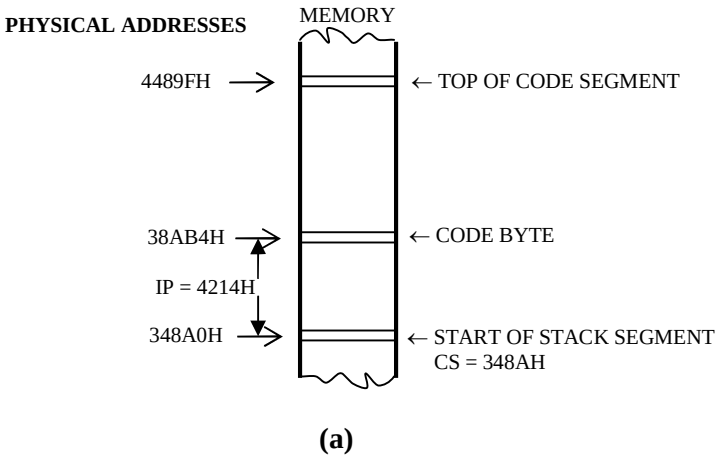
Example 3:- (Refer to fig 15.7)

The value of the Instruction Pointer, holding address of the instruction currently being executed = 4214H

The value of the code segment register (CS) = 348AH

Physical address of the instruction currently being executed = 348AH x 16 + 4214H

= 38AB4



(a)

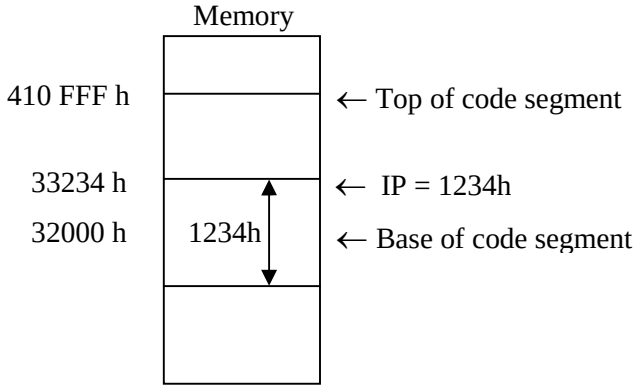
CS	**3**	**4**	**8**	**A**	**0**	**← HARDWIRED ZERO**
IP	**+**	**4**	**2**	**1**	**4**	
PHYSICAL ADDRESSES	**3**	**8**	**A**	**B**	**4**	

(b)

Fig 15.7 Addition of IP to CS to produce the physical address of the code byte. (a) Diagram. (b) Computation.

(c) Instruction Pointer :-

The entire memory of 8086 is divided into logical partitions, called segments. One of the segments, called the code segment, is used to store the segment address of the instructions. The pointer to the instruction that is currently being executed, is actually the offset of the instruction within the code segment. This pointer is called the instruction pointer. Actual address of the instruction is calculated as shown in the Fig 15.8.



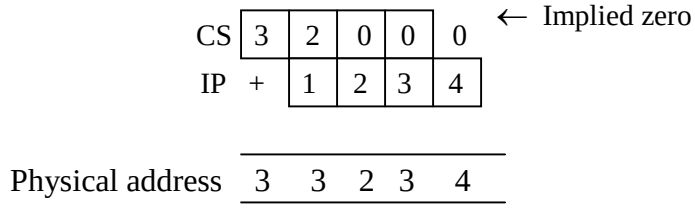


Fig 15.8 Physical address calculation using Instruction pointer and code segment register.

15.1.7 Execution Unit:-

Execution Unit tells the BIU, which memory location to access, and what to do with that. This involves the decoding and execution of the instructions. Execution Unit consists of the following sections:

(i) Control Circuitry, Instruction Decoder and ALU:-

The instruction are fetched from the Instruction Queue, and stored in the decoder, where the instruction is translated into sequence of actions which the EU carries out. The arithmetic and logical operations are performed in the ALU. All the actions are controlled by the control circuitry, which generates appropriate actions at fixed intervals of time.

(ii) Registers:-

Execution Unit in addition to the control circuitry, decoder and the ALU, also contain several general purpose and special purpose registers. The registers are discussed in details in the following sections.

15.1.8 CPU Registers:-

The registers of CPU can be divided into five groups according to their functionality. These groups are:

(a) General purpose Registers
(b) Segment Registers
(c) Pointer and Index Registers
(d) Special Registers
(e) Flags Registers

15.1.9 General purpose Registers:-

8086 has four 16-bit general purpose registers, viz, AH, AL, BH, BL, CH, CL, DH, and DL. These registers can be used individually for temporary storage of 8-bit data. The AL register is also called the accumulator. It has some specific features. Certain pairs of these general-purpose registers can be used together to store 16-bit data words. The acceptable register pairs are AH and

AL, BH and BL, CH and CL, and DH and DL. The AH-AL pair is referred to as the AX register, the BH-BL pair is referred to as the BX register, the CH-CL pair is referred to as the CX register, and the DH-DL pair is referred to as the DX register.

AX register is also called the accumulator. Instructions like, divide, rotate, shift etc., assume one of the operands to be present in the accumulator. BX register is mainly used as a base register. This means by default it is assumed to contain the offset of memory region within data segment. CX register in some loop like instructions is assumed to work as a counter, which contains the number of times the loop is to be executed.

DX register in I/O instructions is assumed to contain the port address.

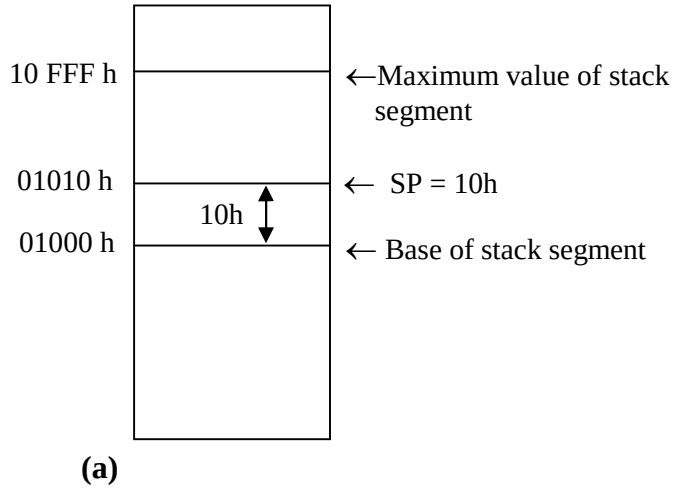
15.1.10 Segment Registers:-

We have already discussed the use of the segment registers, and how they are used to calculate the physical address of the data associated with them. Segment registers cannot be used as byte registers, for obvious reasons.

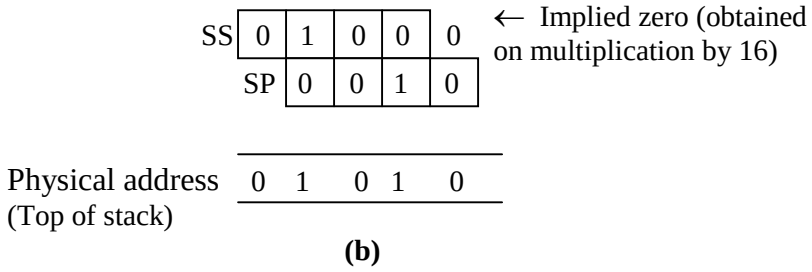
15.1.11 Pointer and Index Registers:-

8086 contain three 16-bits pointer and index registers. They are Base Pointer (BP), Source Index (SI) and Destination Index (DI). These three registers can also be used as general purpose registers. Their main purpose, however, is to contain indexes, in stack segment, data segment and extra segment respectively.

15.1.12 Special Registers:-



(a)



(b)

Fig 15.9 Addition of SS and SP to produce the physical address of the top of the stack. (a) Diagram (b) Computation.

The 8086 allows one to set aside on entire 64-Kbyte segment as a stack. The upper 16-bits of the starting address for this segment are kept in the stack segment register. The stack pointer (SP) register in the execution unit holds the 16-bit offset from the start of the segment to the memory location where a word was most recently stored on the stack. The memory location where a word was most recently stored is called the top of stack. The physical address for a stack read or a stack write is produced by adding the contents of the stack pointer register to the segment box address represented by the upper 16-bits of the box address in SS.

15.1.13 Flag Register:-

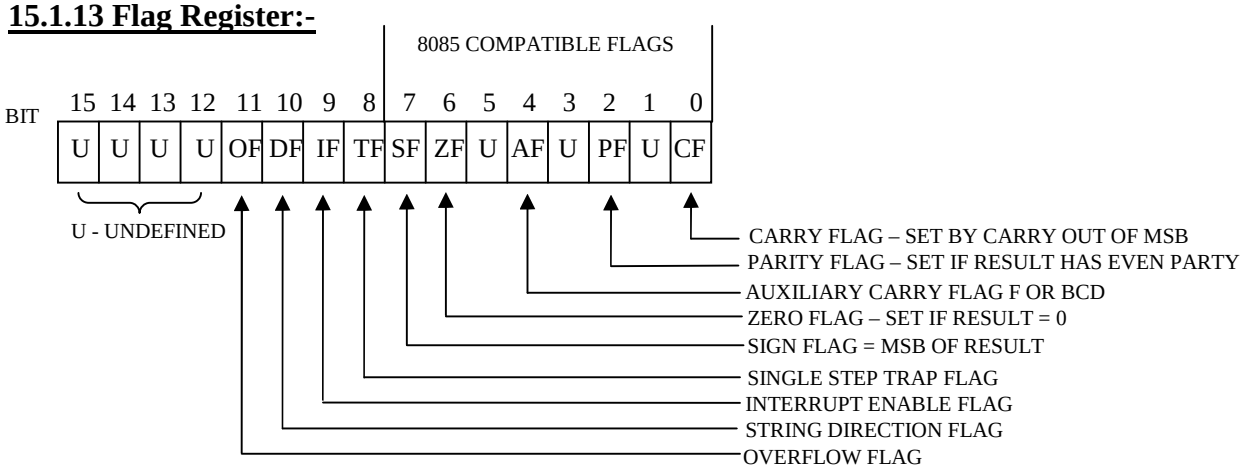


Fig 15.10. 8086 flag register format. (Intel Corp.)

A flag is a flip-flop which indicates some condition produced by the execution of an instruction or controls certain operations of the EU. A 16 bit flag register in the EU contains nine active flags.

Fig 15.9 shows the location of the nine flags used to indicate some condition produced by an instruction.

The six conditional flags are the carry flag (CF), the parity flag (PF), the auxiliary carry flag (AF), the zero flag (ZF), the sign flag (SF), and the overflow flag (OF). The three remaining flags in the flag register

are used to control certain operations of the processor. These flags are different from the six conditional flags described above in the way they are set or reset. The six conditional flags are set or reset by the EU on the basis of the results of some arithmetic or logic operations. The control flags are deliberately set or reset with specific instructions you put in your program. The three control flags are the trap flag (TF), which is used for single stepping through a program; the interrupt flag (IF), which is used to allow or prohibit the interruptions of a program; and the direction flag(DF), which is used with string instructions.

15.2 ADDRESSING MODES:

The way an instruction references its operands, is called its addressing modes, which may be one of the three basic types:

1. Register
2. Immediate
3. Memory

There are several basic elements that can be combined to create a memory operand.

Base Register	BX, BP
Index Register	SI, DI
Displacement	The name of a label or variable, which represents an offset from the beginning of a segment.
Direct Operand	The contents of memory at a location identified by the operand's name
Indirect Operand	An address stored in a register or a variable that is used to locate another variable.

Much of the CPU's power centers on its ability to handle a large variety of operand types, making it easy to access complex data structures.

15.2.1 Register Addressing Mode:-

In the register addressing mode, a register operand may be one of the following 16-bit registers:

AX, BX, CX, DX, SI, DI, SP, BP, IP, CS, DS, ES, SS

Alternatively, it may be one of the following 8-bit registers:

AH, AL, BH, BL, CH, CL, DH, DL

In general, the register addressing mode is the most efficient because registers are part of the CPU and no memory access is required.

Example:

```
MOV     AL, CH
MOV     AX, CX
```

are all examples of register addressing modes.

15.2.2 Immediate Addressing Mode:-

An immediate operand is a constant expression, such as a number, a character, or an arithmetic expression. The assembler must be able to determine the value of an immediate operand at assembly time. Its value is inserted directly into the machine instruction. For example, the following instruction moves 5 to AL.

```
MOV     AL,05
```

The corresponding machine code generated by the assembler is

```
B0 05
```

Where the first byte is the op code, and the second byte is the immediate value 05. Examples of immediate operands are shown below. The last one (2+3) / 5, is an expression that is evaluated at assembly time.

```
Examples:   MOV     AL, 10
            MOV     AL, 'A'
            MOV     AX, 'AB'
            MOV     AX, 64000
            MOV     AL, (2+3)/5
```

15.2.3 Direct Addressing Mode:-

A direct operand refers to the contents of memory at an address implied by the name of the variable. Assume COUNT was declared as a byte variable in the data segment, the following examples are valid:

```
MOV     COUNT, CL       ; move CL to COUNT
MOV     AL, COUNT       ; move COUNT to AL
JMP     LABELI          ; jump to LABELI
```

These are also called **as relocatable operands**. Their location depends on the offset of a label from the beginning of a segment. Depending on which segment the label is located in, the following segment registers are used by default:

Type of label	Default segment Register
Program code (instructions)	CS
Variable (data)	DS

In the above example, COUNT, by default will be assumed to be in data segment, while LABEL 1, will be assumed to be in code segment.

The other type of direct memory operands are non re-locatable. Their syntax is:

Segment : offset

Segment refers to either a segment register, or a segment name. Its value is unknown at assembly time, because it depends on where DOS will eventually load the program. Offset may be an integer, symbol table, or variable. Examples are:

```
MOV     AX, DS:5            ;segment register and offset
MOV     BX, CSEG:2Ch        ;segment name and offset
MOV     AX, ES:COUNT        ;segment register and variable
```

The offsets of these variables are calculated with respect to the segment name (register) specified in the instruction.

15.2.4 Indirect Addressing Mode:-

In indirect addressing modes, operands use registers to point to locations in memory. If a register is used in this way, we can change its value and access different memory locations at run time. Two types of registers are used: base register (BX, BP) and index registers (SI,DI). BP is assumed to contain offset from the stack segment. SI, DI, and BX contain offset from DS, the data segment register.

There are five indirect addressing modes, identified by the type of operands used:

1. Register indirect
2. Based indirect
3. Indexed indirect
4. Based indexed
5. Based indexed with displacement

1. Register Indirect:-

Indirect operands are particularly powerful when processing list of arrays, because a base or an index register may be modified at run time. In the following example,

```
MOV   BX, OFFSET ARRAY  ;point to start of array
MOV   AL, [BX]                  ;get first element
INC   BX                        ;point to next
MOV   DL, [BX]                  ;get second element
```

The brackets around BX signify that we are referring to the contents of memory location, using the address stored in BX.

In the following example, three bytes in an array are added together:

```
MOV   SI, OFFSET ARRAY      ;address of first byte
MOV   AL, [SI]              ;move the first byte to AL
INC   SI                    ;point to next byte
ADD   AL, [SI]              ;add second byte
INC   SI                    ;point to the third byte
ADD   AL,[SI]               ;add the third byte
```

2&3 Based Indirect and Indexed Indirect:-

Based and Indexed addressing modes are used in the same manner. The contents of a register are added to a displacement to generate an effective address. The register must be one of the following: SI, DI, BX or BP. If the registers used for displacement are base registers, BX or BP, it is said to be base addressing, or else it is called indexed addressing. A displacement is either a number or a label whose offset is known at assembly time. The notation may take several equivalent forms.

Register added to an offset:

```
MOV   DX, ARRAY[BX]
MOV   DX, [DI + ARRAY]
MOV   DX, [ARRAY + SI]
```

Register added to a constant:

```
MOV   AX, [BP + 2]
MOV   DL,[DI – 2]           ;DI + (-2)
MOV   DX, 2[SI]
```

If BX, SI, or DI is used, the effective address is usually an offset from the DS register, BP on the other hand, usually contains an offset from the SS register.

4. Based Indexed:-

In this type of addressing the operand's effective address is formed by combining a base register with an index register. For example:

```
MOV   AL,[BP][SI]
MOV   DX,[BX + SI]
ADD   CX,[DI][BX]
```

Two base register or two index registers cannot be combined, so the following would be incorrect:

```
MOV   DL,[BP + BX]        ;error : two base registers
MOV   AX,[SI + DI]        ;error : two index registers
```

5. Based Indexed with Displacement:-

The operand's effective address is formed by combining a base register, an index register, and a displacement. Examples are as follows:

```
MOV   DX, ARRAY[BX][SI]
MOV   AX, [BX + SI + ARRAY]
ADD   DL, [BX + SI + 3]
SUB   CX, ARRAY [BP + SI]
```

Two base registers or two index registers cannot be combined, so the following would be incorrect:

```
MOV   AX,[BP + BX + 2]
MOV   DX, ARRAY [SI + DI]
```

15.3 Instruction Set:-

To run a program the microcomputer must have the program, in the form of logical sequence of instruction stored in binary form, in its memory. However writing programs as sequence of 1's and 0's can be very cumbersome, both to write and understand. Therefore, certain mnemonics are substituted for unique combinations of 1's and 0's. these mnemonics are called instructions, and this language is called assembly language. General format for an assembly language statement is as follows:

Example:-

Label	op-code	operand(s)	comment
Next	ADD	AL, BL	Add the contents of register AL to Register BL

In the above example, NEXT is the label field. It is used for giving all identity to any statement. It is purely an optional field, and is needed only when the normal flow of the program needs to be changed.

ADD is an op-code, or operation-code. It is the mnemonic for the sequence of 1's and 0's, which when executed, adds the registers AL and BL.

AL and BL are the two operands needed to execute this instruction. The number of operands are dependent upon the instructions.

There are certain instructions, which do not need any operand, while there are some which need one or even two operands. Again the operands can be:

1. name of the registers
2. name of some memory location
3. any constant, called the literal, or
4. the label to a statement in the program.

Comments always start with a semicolon, and end with a carriage return and a linefeed. If the comment is very long it can be extended to more than one line, by putting a semicolon at the beginning. There is no fixed column from which the comment should begin, it can begin from any where from column one, to the last few columns. Though the comments are purely optional, still it is always advisable to use them for better understanding of the program.

In the next few sections we look at the instructions set of the 8086 microprocessor. These instructions are grouped according to their functionality. The most frequently used instruction groups are I) Data transfer group ii) Arithmetical group iii)string instructions

iv) Conditional branch instructions.

15.3.1 Data Transfer Instructions:-

Table 15.2 lists the 14 data transfer instructions. These instructions move single bytes and words between a register and I/O ports. Let us explain some of the instructions in Table 15.2.

- MOV CX, DX copies the 16-bit contents of DX into CX, MOV AX, 2025H moves immediate data 2025H into the 16-bit register AX.
- MOV CH, [BX] moves the 8-bit contents of a memory location addressed by BX in segment register DS into CH. If [BX] = 0050H, [DS] = 2000H, and [20050H] = 08H, then, after MOV CH, [BX], the contents of CH will be 08H.

15.2 TABLE

8086 Data Transfer Instructions (Courtesy of Intel Corporation)

General Purpose	
MOV d,s	[d] ← [s] MOV byte or word
PUSH d	PUSH word into stack
POP d	POP word off stack
XCHG d,s	[d] ↔ [s] Exchange byte or word
XLAT	AL ← [20-bit address computed from AL, BX, and DS]

Input / Output	
IN A, DX or Port	Input byte or word
OUT DX or Port, A	Output byte or word
Address Object	
LEA reg, mem	LOAD Effective Address
LDS reg, mem	LOAD pointer using DS
LES reg, mem	LOAD Pointer using ES
Flag Transfer	
LAHF	LOAD AH register from flags
SAHF	STORE AH register in flags
PUSHF	PUSH flags onto stack
POPF	POP flags of stack

d = "mem' or "reg" or "segreg", s = "data" or "mem" or "reg" or "segreg", A= AX or AL

- MOV START [BP], CX moves the 16-bit (CL to first location and then CH) contents of CX into two memory locations addressed by the sum of the displacement START and BP in segment register SS. For example, if [CX] = 5009H,[BP] = 0030H, [SS] = 3000H, and START = 06H, then, after MOV START [BP], CX, physical memory location [30036H] = 09H and [30037H] = 50H. Note that the segment register SS can be overridden by CS using MOV CS: START [BP], CX.
- LDS SI, [10H] loads register and DS from memory. For example, if[DS] = 2000H, [20010] = 0200H, and [20012] = 0100H, then, after LDS SI, [10H], SI and DS will contain 0200H and 0100H, respectively.
- In the 8086, the SP is decremented by 2 for PUSH and incremented by 2 for POP. For example, consider PUSH[BX]. If[DS] = $2000_{16,}$ [BX] = $0200_{16,}$ [SP] = $3000_{16,}$ [SS] = $4000_{16,}$ and [20200] = $0120_{16,}$ then, after execution of PUSH[BX], memory location 42FFF and 42FFE will contain 01_{16} and $20_{16,}$ respectively, and the contents of SP will be $2FFE_{16.}$
- XCHG has two variations. These are XCHG reg, and XCHG mem, reg. For example, XCGH AX, BX exchange the contents of a 16-bit register such as BX with the contents of AX. XCHG mem, reg exchanges 8-bit or 16-bit data in reg or mem with 8 or 16-bit reg.
- XLAT can be used to employ an index in a table. This instruction utilizes BX to hold the starting address of the table in memory consisting of 8-bit data elements. The index in the table is assumed to be in the AL register. For example, if [BX] = $0200_{16,}$ [AL] = $04_{16,}$ and [DS] = $3000_{16,}$ then, after

XLAT, the contents of location $30204_{16,}$ will be loaded into AL. Note that the XLAT instruction is the same as MOV AL, [AL] [BX].

- Consider fixed port addressing in which the 8-bit port address is directly specified as part of the instruction. IN AL, 38H inputs 8-bit data from port 38H into AL. IN AX, 38H inputs 16-bit data from ports 38H and 39H into AX. OUT 38H, AL outputs the contents of AL to port 38H. OUT 38H, AX, on the other hand, outputs the 16-bit contents of AX to ports 38H and 39H.
- For variable port addressing, the port address is 16-bit and is specified in the DX register. Consider ports addressed by the 16-bit address contained in DX. Assume [DX] = $3124_{16,}$ in all the following examples:

  IN AL, DX inputs 8-bit data from 8-bit port $3124_{16,}$ into AL.

  IN AX, DX inputs 16-bit data from ports $3124_{16,}$ and $3125_{16,}$ into AX.

  OUT DX, AL outputs 8-bit data from AL into port $3124_{16,}$ data.

  OUT DX, AL outputs 16-bit data from AX into ports $3124_{16,}$ and $3125_{16.}$

Variable port addressing allows up to 65,536 port with addresses from 0000H to FFFFH. The port addresses in variable port addressing can be calculated dynamically in a program. For example, assume that an 8086-based microcomputer is connected to three printers via three separate ports. Now, in order to output to each one of the printers, separate programs are required if fixed port addressing is used. However, with variable port addressing, one can write a general subroutine to output to the printers and then supply the address of the ports for a particular printer in which data output is desired to register DX in the subroutine.

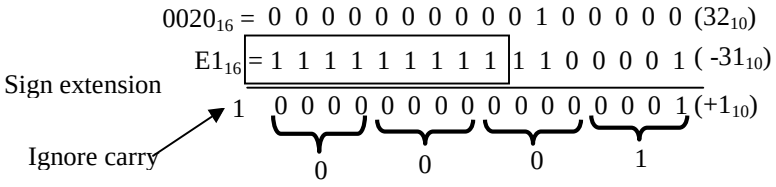
15.3.2 Arithmetic Instructions:-

Table 5-3 shows the 8086 arithmetic instructions. These operations can be performed on four types of numbers: unsigned binary, signed binary, unsigned packed decimal, and signed packed decimal numbers. Binary numbers can be 8 or 16 bits wide. Decimal numbers are stored in bytes – two digits per byte for unpacked decimal and one digit per byte for unpacked decimal and one digit per byte for unpacked decimal with the high 4 bits filled with zeros.

Let us explain some of the instructions in Table 5-3.

- Consider ADC mem/reg, mem/reg. This instruction adds dsta with carry from reg to reg or from reg to mem or from mem to reg. There is no ADC mem, mem instruction. For example, if [AX] = $0020_{16,}$ [BX] = $0300_{16,}$ CF = 1, [DS] = $2020_{16,}$ and [20500] = $0100_{16,}$ then after ADC AX, [BX], the contents of register AX = 0020 + 0100 +1 = $0121_{16.}$ All flags are affected.

- DIV mem/reg divides [AX] or [DX:AX] registers by reg or mem. For example, if [AX] = $0005_{16,}$ and [CL] = $02_{16,}$ then, after DIV CL, [AH] = $01_{16,}$ and [AL] = $02_{16.}$
- Consider MUL BL. If [AL] = $20_{16,}$ and [BL] = $02_{16,}$ then, after MUL BL, register AX will contain $0040_{16.}$
- Consider CBW. This instruction extends the sign from the AL register to AH register. For example, if AL = $F1_{16,}$ then, after execution of CBW, register AH will contain $FF_{16,}$ since the most significant bit of $F1_{16,}$ is 1. Note that the sign extension is very useful when one wants to perform an arithmetic operation on two numbers of different lengths. For example, the 16-bit number 0020_{16} can be added with the 8-bit number $E1_{16}$ by sign-extending E1 as follows:



Another example of sign extension is that, in order to multiply a signed 8-bit number by a signed 16-bit number, one must first sign-extend the signed 8-bit into a signed 16-bit number and then the instruction IMUL can be used for 16 x 16 signed multiplication.

- CWD sign-extends the AX register into the DX register. That is, if the most significant bit of AX is 1, then store $FFFF_{16}$ into DX.
- The distinction between the byte or word operations on memory is usually made by the assembler by using B for byte or W for word as trailing characters with the instructions. Typical examples are MULB or MULW.
- Numerical data received by an 8086-based microcomputer from a terminal is usually in ASCII code. The ASCII codes for numbers 0 to 9 are 30H through 39H. Two 8-bit data can be entered into an 8086 based microcomputer via a terminal. The ASCII codes for these data (with 3 as the upper nibble for each type) can be added. AAA instruction can then be used to provide the correct unpacked BCD. Suppose that ASCII codes

for 2 (32_{16}) and 5 (35_{16}) are entered into an 8086-based microcomputer via a terminal. These ASCII codes can be added and then the result can be adjusted to provide the correct unpacked BCD using the AAA instruction as follows: for 2 (32_{16}) and 5 (35_{16})

```
ADD CL,  DL        ; [CL] = 32H = ACSII for 2
                   ; [DL] = 35H = ASCII for 5
                   ; result [CL] = 67H
MOV AL,  CL        ; Move ASCII result
                   ; into AL since AAA
                   ; adjust only [AL]
AAA                ; [AL] = 07, unpacked
                   ; BCD for 7
```

Note that, in order to send the unpacked BCD result 07_{16} back to the terminal, [AL] = 07 can be ORed with 30H to provide 37H, the ASCII code for 7.

- DAA is used to adjust the result of adding two packed BCD numbers in AL to provide a valid BCD number. If after the addition, the low 4 bits of the result in AL is greater than 9 (or if AF = 1), then the DAA adds 6 to the low 4 bits of AL. On the other hand, if the high 4 bits of the result in AL is greater than 9 (or if CF = 1), then DAA adds 60H to AL.
- DAS may be used to adjust the result of subtraction in AL of two packed BCD numbers to provide the correct packed BCD.
  While performing these subtractions, any borrows from low and high nibbles are ignored. For example, consider subtracting BCD 55 in DL from BCD 94 in AL:

```
SUB AL, DL      ; [AL] = 3FH low nibble = 1111
DAS             ; CF = 0              -6 = 1010
                                 Ignore 1   1001
                ; [AL] = 39 BCD
```

- **IMUL** mem / reg provides signed 8 x 8 or signed 16 x 16 multiplication. As an example, if [CL] = FDH = - 3_{10}, [AL] = FEH = - 2_{10}, then, after IMUL CL, register AX contains 0006H.
- Consider 16 x16 unsigned multiplication, MUL WORD PTR [BX]. If [BX] = 005H, [DS] = 3000H, [30050H] = 0002H, and [AX] = 0006H, then, after MUL WORD PTR [BX], [DX] = 0000H and [AX] = 000CH.
- Consider DIV BL. If [AX] = 0009H and [BL] = 02H, then after DIV BL,
  [AH] = remainder = 01H
  [AL] = quotient = 04H

- Consider IDIV WORD PTR [BX]. If [BX] = 0020H, [DS] = 2000H, [20020H] = 0004H, and [DX] [AX] = 00000011H, then, after IDIV WORD PTR [BX].

  [DX] = remainder = 0001H

  [AX] = quotient = 0004H

- AAD converts two unpacked BCD digits in AH and AL to an equivalent binary number in AL. AAD must be used before dividing two unpacked BCD digits in AX by an unpacked BCD byte. For example, consider dividing [AX] = unpacked BCD 0508 (58 decimal) by [DH] = 07H. [AX] must first be converted to binary by using AAD. The register AX will then contain 003AH = 58 decimal. After DIV DH, [AL] = quotient = 08 unpacked BCD, [AH] = remainder = 02 unpacked BCD.
- AAM adjusts the product of two unpacked BCD digits in AX. If [AL] = BCD3 = 00000011_2 and [CH] = BCD8 = $0000\ 1000_2$, then, after MUL CH, [AX] = 0000000000011000_2 = 0018H, and after using AAM, [AX] = $00000010\ 00000100_2$ = unpacked 24. The following instruction sequence accomplishes this:

```
MUL CH
AAM
```

Table 15.3 Arithmetic Instructions (courtesy of Intel Corporation)

Addition	
AD a,b	Add byte or word
ADC a,b	Add byte or word with carry
INC reg / mem	Increment byte or word by one
AAA	ASCII adjust for addition
DAA	Decimal adjust [AL]to be used after ADD or ADC
Subtraction	
SUB a, b	Subtract byte or word
SBB a, b	Subtract byte or word with borrow
DEC reg / mem	Decrement byte or word by one
NEG reg / mem	Negate byte or word
CMP a, b	Compare byte or word
AAS	ASCII adjust for subtraction
DAS	Decimal adjust [AL] to be used after SUB or SBB

Multiplication

MUL reg / mem	Multiply byte or word unsigned	for byte
MUL reg / mem	Integer multiply byte or word (signed)	[AX] ← [AL].[mem / reg for word [DX][AX] ← [AX].[mem / reg]

Division

DIV reg / mem	Divide byte or word unsigned	16 ÷ 8 bit
IDIV reg / mem	Integer divide byte or word (signed)	[AH] ← remainder [AX] ← -------- [AL] ← quotient [mem / reg] 32 ÷ 16 bit
AAD	ASCII adjust for division	
CBW	Convert byte to word	[DX] ← remainder [DX:AX]
CWD	Convert word to double word	[AX] ← quotient [mem / reg]

a = “reg” or “mem”, b = “reg” or “mem” or “data”.

Note that 8086 does not allow multiplication of two ASCII codes. Therefore, before multiplying two ASCII bytes received from a terminal, one must mask the upper 4 bits of each one of these bytes and then multiply them as two unpacked BCD digits and then use AAM for adjustment. In order to convert the unpacked BCD product back to ASCII for sending back to the terminal, one must OR the product with 3030h.

- If CMPS is prefixed with REPE or REPZ, the operation is interpreted as “compare while not end-of-string (C X ≠ 0) and strings are equal (ZF = 1).” If CMPS is preceded by REPNE or REPNZ, the operation is interpreted as “compare while not end-of-string (CX ≠ 0) and strings not equal (ZF = 0).” Thus, repeated CMPS can be used to find matching or differing string elements.

15.3.5 Conditional Branch Instructions:- All 8086 conditional branch instructions use 8-bit signed displacement. That is, the displacement covers a branch range of -128_{10} to $+127_{10}$ with ‘0’ being positive. The structure of a typical conditional branch instruction is as follows

If condition is true,

Then PC ← PC + disp8

otherwise PC ← PC + 2 and execute next instruction

There are two types of conditional branch instructions. In one type, the various relationships that exist between two numbers such as equal, above, below, less than, or greater than can be determined by the appropriate conditional branch instruction after a COMPARE instruction. These instructions can be used for both signed and unsigned numbers. While comparing signed numbers, terms such as “less than” and “greater than” are used. On the other hand, while comparing unsigned numbers, terms such as “below zero” or “above zero” are used.

Table (15.4) lists the 8086 signed and unsigned conditional branch instructions. Note that in table 15.4, the instructions for checking which two numbers are “equal”or “not equal” are the same for both signed and unsigned numbers. This is because when two numbers are compared for equality, irrespective of whether they are signed or unsigned, they will provide a zero result (Z = 1) if equal or a nonzero result (Z = 0) if not equal. Therefore, the same instructions apply for both signed and unsigned numbers for “equal to” or “not equal to” conditions.

The second type of conditional branch instructions is concerned with the setting of flags rather than the relationship between two numbers. Table 15-5 lists these instructions.

Now, in order to check whether the result of an arithmetic or logic operation is zero, nonzero, positive or negative, did or did not produce a carry, did or did not produce parity, or did or did not cause overflow, the following instructions should be used: JZ, JNZ, JS, JNS, JC, JNC, JP, JNP, JO, JNO.

TABLE 15.4
8086 Signed and Unsigned Conditional Branch Instructions

Signed		Unsigned	
Name	**Alternate Name**	**Name**	**Alternate Name**
JE disp 8 (JUMP if equal)	JZ disp8 (JUMP if result zero)	JE disp8 (JUMP if equal)	JZ disp8 (JUMP if zero)
JNE disp8 (JUMP is not equal)	JNZ disp8 (JUMP if not zero)	JNE disp8 (JUMP if not equal)	JNZ disp8 (JUMP if not zero)
JG disp8 (JUMP if greater)	JNLE disp8 (JUMP if not less or equal)	JA disp8 (JUMP if above)	JNBE disp8 (JUMP if not below or equal)
JGE disp8 (JUMP if greater or equal)	JNL disp8 (JUMP if not less)	JAE disp8 (JUMP if above or equal)	JNB disp8 (JUMP if not below)
JL disp8 (JUMP if less than)	JNGE disp8 (JUMP if not greater or equal)	JB disp8 (JUMP if below)	JNAE disp8 (JUMP if not above or equal)
JLE disp8 (JUMP if less or equal)	JNG disp8 (JUMP if not greater)	JBE disp8 (JUMP if bellow or equal)	JNA disp8 (JUMP if not above)

Table 15.5
8086 Conditional Branch Instructions Affecting Individual Flags

JC disp	JUMP if carry, i.e., CF = 1
JNC disp8	JUMP if no carry, i.e., CF = 0
JP disp8	JUMP if parity, i.e., PF = 1
JNP disp8	JUMP if no parity, i.e., PF = 0
JO disp8	JUMP if overflow, i.e., OF = 1
JNO disp8	JUMP if no overflow, i.e., OF = 0
JS disp8	JUMP if sign, i.e., SF = 1
JNS disp8	JUMP if no sign, i.e., SF = 0
JZ disp8	JUMP if result zero, i.e., Z = 1
JNZ disp8	JUMP if result not zero, i.e., Z = 0

15.4 Summary:

- In this lesson we have studied about one of the most popular series of microprocessors, viz. Intel 8086.
- 8086 serves as a base to all its successors, 8088, 80186,80286,80386 and 80486.
- All the programs on 8086 can be directly run on any of its successors.
- To summarize, the features of 8086, we can say, 8086:

- a 16-bit data bus
- a 20-bit address bus
- CPU is divided into bus Interface Unit and Execution unit
- 6 byte instruction prefetch queue
- Segmented memory
- 4 several purpose registers
- Instruction pointer and a stack pointer
- Set of index Registers
- Powerful instruction set
- Powerful addressing modes
- Designed for multi processor environment
- Available in versions of 5 MHz and 8MHz clock speed

15.5 Key Terminology:

Accumulator: The primary data register within the CPU that holds the initial and final results of a number of processor operations

Address: A block of data specifying a memory or port location

Addressing mode: A particular way of specifying the operand of our instruction.

Bus: A set of conducting elements forming a common connection between many

circuit elements

Byte: 8-bits

Data transfer: A process that moves information between CPU and external device by use of a where/when operation

Instruction set: The list of instruction types recognized by a given microprocessor.

Microprocessor: On LSI component usually integrating the CPU and ALU on a single-chip microprocessors place all components on the chip.

Segmented address: Physical address equals sum of base segment address and offset into segment.

Stack: A block of successive memory locations used to store return address and register and flag information during subroutine processing.

Stack pointer: A register used to point to the most recent data stored in the stack or the next piece of data to be removed.

Status register: A register used to hold information relative to the present state of the system.

15.6 Self assessment Questions:-

1. How many functional units does 8086 contain? Discuss them in brief.
2. Discuss the register organization of 8086. Explain the function of each register.
3. Discuss the function of instruction pointer and stack pointer in 8086.
4. How many operating modes does 8086 have? Discuss them in brief.
5. Discuss the various addressing modes of 8086.

15.7 References:-

1. **Microcomputer systems:** The 8086/8088 Family Architecture, Programming and Design – Yu – Cheng. Liu Glenn A. Gibson, PHI
2. **Microprocessor:** Theory and Applications – M. Rafiquzzaman, PHI
3. **Microprocessor and Interfacing –** Programming and Hardware – 2nd edition, D.V. Hall – McGraw – Hill International editions.
4. **Intel Microprocessor – Hardware, Software and Applications –** Roy . W. Goody **McGraw –** Hill International Editions.

Name of the Author:

S.V. RangaNayakulu, M.Sc., M.Phil., M.Ed.,

Dept. of Physics & Electronics

T.J.P.S. COLLEGE (P.G.COURSES)

GUNTUR-522 006

UNIT IV

LESSON-16

Assembly Language Programming of 8086

Objectives:-

This lesson introduces assembly language Programming and provides some sample programs to get some basic background knowledge of writing Programs.

Structure of the lesson:-

16.1.0 Introduction:-

Use: Assembly language teaches you about the way the computer's hardware and operating system work together and how, the application programs communicate with the operating system. Assembly language, unlike high level language, is machine dependent. Each microprocessor has its own set of instructions, that it can support. Here we will discuss, only the IBM-PC assembly language. It consists of the Intel 8086 / 8088 instruction set.

16.1.1 Why learn Assembly language?

You must learn assembly language for various reasons.

1. It helps you understand the computer architecture and operating system.
2. one of the most important advantages of assembly language, is that the Programs written in assembly language are easier than the same program written in high level language. The reason for this is that, as of today the compilers are still not so intelligent to take advantage of some of the complex instructions of the assembly language.
3. Assembly language has very few restrictions or rules, nearly every thing is left to the discretion of the Programmer. High level languages, out of necessity, impose rules about what is allowed in a program.

16.1.2 Assembly Language Applications:-

The assembly language programs presented in this unit, are all trivial. The language requires a great deal of attention to learn in detail. Most programmers don't write large application programs in assembly language, instead they write short, specific routines. Often we write short sub-routines in assembly language, and call them from high level language. You can take advantage of the strengths of high level languages by using them to write applications. Then you can write assembly language subroutines to handle operations that are not available in high level language.

For example, suppose you are writing a word processor in Pascal but find that this language performs badly when updating the screen display. You can write the routines to handle the screen in assembly language. Similarly, you can write other critical parts of the Program in assembly language to speed up the performance of the Program.

16.1.3 Machine Instructions:-

A machine instruction is a binary code that has a special meaning for a computer's CPU; it tells the CPU, to perform a task. The task could be to move a number from one memory location to

another, or to add two specific numbers. Each machine instruction is precisely defined when the CPU is constructed, and it is specific to that type of CPU.

16.1.4 Assembly Instructions:-

Assembly language makes the task easier, to program directly in machine language using numbers. The assembly language instruction to move the number 04 to AL, would be

MOV AL, 04

Assembly language is called the low-level language, because it is close to the machine language in structure and function.

16.2.0 Assembly Language Fundamentals:-

The best way to learn to write assembly language program, is to first study a simple assembly written Program.

16.2.1 A Simple Program:-

```
ABSTRACT : This Program adds two 8-bit words in the memory.
         : Locations called NUM1 and NUM2.
         : The result is stored in the memory location called RESULT.
         : If there was a carry from the addition, it will be stored as 0000 0001
           in the location CARRY.
ALGORITHM :
         get  NUM 1
         add  NUM 2
         Put sum into memory at SUM
         Position carry in LSB of byte registers
         Mask off upper seven bits
           Store the result in the carry location
REGISTERS        Uses CS, DS, AX
DATA  SEGMENT
         NUM DB    16h ;  first number stored here
         NUM DB    19h ;  second number stored here
         RESULT    DB     ?     ; put sum here
         CARRY     DB     ?     ; put any carry here
DATA ENDS
CODE  SEGMENT
```

```
        ASSUME    CS: CODE,    DS :   DATA
START:MOV  AX,    DATA         ;      Initialize data segment
           MOV    DS,     AX          ;      register
           MOV    AL,     NUM1        ;      Get the first number
           ADD    AL,     NUM2        ;      add it to 2nd number
           MOV    RESULT, AL          ;      Store the Result
           RCL    AL,     01          ;      Rotate carry into LSB
           AND    AL,  00000001B      ;      mask out all but LSB
           MOV    CARRY,  AL          ;      store the carry result
           INT    3h
CODE       ENDS
           END START
```

This program contains, certain additional mnemonics, in addition to the instructions you have studied so far. These are called as **assembler directives or pseudo operations.** These are the directions for the assembler. Their meaning is valid only till the assembly time. There is no code generated for them. We shall now study the complete program step by step.

16.2.2 SEGMENT and ENDS Directive:-

These directives are used to identify a group of data items or a group of instructions, called the segment. A group of data statement or the instructions, that are put in between the SEGMENT and ENDS directives are said to constitute a logical statement. This segment is given a name. In our example CODE and DATA are the name given to code and data segments respectively. The segments should have a unique name, can be no blanks within the segment name, the length of the segment name can be up to 31 characters. Name of the mnemonics or any other reserved word is not allowed as the segment name or label.

16.2.3 Data Definition Directives:-

In assembly language, we define storage for variables using data definition directives. Data definition directives create storage at assembly time, and can even initialize a variable string to a starting value. The directives are summarized in the following table.

Directive	Description	Number of Bytes	Attribute
DB	Define byte	1	Byte
DN	Define word	2	Word
DD	Define double word	4	Double word
DQ	Define quad ward	8	Quad ward
DT	Define 10 bytes	10	Ten bytes

DUP directive is used to duplicate the basic data definition 'n' number of times.

Example:-

```
ARRAY    DB        10DUP(0)
```

Define an arrange ARRAY of 10 data bytes, each byte initialized to '0'. The initial value can be anything acceptable to the basic data type.

EQU directive is used to define a name to a constant.

16.2.4 The ASSUME Directive:-

8086 has four type of segments. In the program there can be more than one code segments, data segments, or extra segments defined. ASSUME directive is used to tell the assembler, which segment is to be used as an active segment at any instant, and with respect to which it has to calculate the offsets of the variables or instructions. It is usually placed immediately after the SEGMENT directive, in the code segment, but you can have as many additional ASSUMEs as you like. Each time an ASSUME is encountered, the assembler starts to calculate the offset with respect to that segment.

16.2.5 Initializing Segment Registers:-

ASSUME is only a directive, which is used to calculate the offset of variables, instructions or stack element, with respect to a specific segment of its type. It does not initialize the segment registers. Initialization of the segment registers has to be done explicitly using MOV instructions as follows:

```
MOV     AX,   DATA
MOV     DS,    AX
```

The above statements are used to initialize the data segment register. The segment registers cannot be directly loaded with memory variable, therefore, the segment name is first moved into

some general purpose register, which then is moved into the segment register. All segment registers can be initialized in the same manner: code segment register is initialized automatically by the loader.

16.2.6 END Directive:-

The END directive tells the assembler to stop reading and assembling the program from there an any statement after the END will be ignored by the assembler. There can be only one END in the program, which is the last statement of the program.

16.2.7 Input / output Services:-

On the machines based on Intel 8086 and 8088 series, running on DOS, the input / output is carried out in the form of services provided by the hardware and the operating systems. They are called as ROM-BIOS, and DOS services respectively, and are in the form of interrupts.

16.2.8 Interrupts:-

An interrupt occurs when a currently executing program is interrupted. Interrupts are generated for a variety of reasons, usually related to the services, related to the external devices connected to the machine, example:- Keyboard, printer, monitor, etc.

8086 recognizes two kinds of interrupts, hardware interrupts and software interrupts. Hardware interrupts are generated when the peripheral, connected to the CPU requests for some service. A software interrupt is a call to a sub routine located in the operating system, usually the input-output routine.

INT (interrupt) instruction is used with in application programs to request the services of DOS, or ROM-BIOS. The INT instructions calls an operating system sub-routine, identified by a number, in the range 0-FFh. The syntax is

INT number

The CPU processes an interrupt instructions using the interrupt vector table (IVT). It is situated in the first 1K bytes of memory, and has a total of 256 entries, each of 4 bytes. The entry in the interrupt vector table is identified by the number given in the interrupt instruction, and in turn points to, an operating system sub-routine. The actual address in this table varies from machine to machine.

16.2.9 DOS Function calls:-

INT 21h is called a DOS function call. There are some 100 different functions supported by this interrupt, identified by a function number placed in the AH register, prior to calling INT. we shall explain the DOS routines in the following format.

Function number	Description
	Call with

Returns

Example

01h Reads the character from the Keyboard, and echoes the character on the monitor. If no character is ready, waits until one is available.

Calls with AH = 01

Returns AL = 8 bit data input

Example:-1.

Read one character from the keyboard into register AL, with echo, and store it in the variable VAR 1

```
MOV   AH,     01
INT   21h
MOV   VAR 1, AL
VAR 1 DB0
```

Example:-2

Outputs the character on the monitor

calls with: A + 1 = 02

DL = 8 bit – data

Returns: nothing

Example: transmit the character '*' in the screen

```
MOV     AH,     02
MOV     DL,     '*'
INT     21h
```

16.3 Assembly Language Program Development tools:-

There are some developmental tools required for writing assembly language programs and try them out on the machine. Let us study them now.

Editor:-

An editor is a program which, when run on a system, lets you type in text, and store in a file. This text could also be your assembly language program. There are a number of editors available on PC. Some of the more popular ones are: EDLIN, WORDSTAR, TURBO etc., The editor helps you type the program in required format. This form of the program is called as the source program. The editor gives you all the flexibility, to insert lines, insert words, characters, delete words, characters etc.

Assembler:-

An assembler program is used to translate assembly language mnemonics to the binary code for each instruction. After the complete program has been written, with the help of an editor, it is then assembled with the help of an assembler. The assembler generates two files, one is object file and the other is list file. The object file contains the binary code for each instruction in the program. It is created only when your program has been successfully assembled, with no errors. The errors that are detected by the assembler, are called the syntax errors.

List file is optional and contains the source code, the binary equivalent of each instruction, and the offsets of the symbols in the program. This file is for purely documentation purposes.

Some of the assemblers available on PC are MASM, TURBO etc.,

Linker:-

The file containing the common routines, can be linked to your other programs also. The program that links your programs is called the linker.

The linker produces a link files which contains the binary codes for all compound modules. The linker also produces a link map which contains the address information about the linked files. The linker, however, does not assign absolute address to your program. The linker available on your PC is LINK. TURBO has a built in linker.

Loader:-

Loader is a program, which assigns absolute addresses to the program. These addresses are generated, by adding to all the offsets, the address from where the program is loaded into the memory. Loader comes into action, when you execute your program. This program is brought from the secondary memory, like disk, or floppy diskette into the main memory at a specific address.

Debugger:-

Debugger allows you to load your program into just like a loader, and trouble shoot your program. Debugger helps you detect the logical errors, that could not be detected by the assembler.

16.4 Sample Programs:-

1. Write an assembly language program to calculate the average of two temperatures :-

```
ABSTRACT:    This program averages two temperatures.
         ;   Named HI_TEMP and LO_TEMP and puts the result in the
         ;   Memory locations AV_TEMP
REGISTERS:   Uses DS, CS, AX, BL
   PORTS:    None used
```

```
DATA        SEGMENT
        HI_TEMP DB  92h     ;       max     temp    storage
        LO_TEMP DB  52h     ;       low     temp    storage
        AV_TEMP DB  ?       ;       store   average here
DATA    ENDS
CODE        SEGMENT
        ASSUME      CS:     CODE,       DS: DATA
START:      MOV     AX,     DATA        ;       Initialize  data  segment
            MOV     DS,     AX          ;       Register
            MOV     AL,     HI_TEMP     ;       Get first temperature
            ADD     AL,     LO_TEMP     ;       Add second to it
            MOV     AH,     00h         ;       Clear all of AH register
            ADC     AH,     00h         ;       Put carry in LSB of AH
            MOV     BL,     02h         ;       Load divisor in BL register
            DIV     BL                  ;       Divide AX by BL.  Quotient in AL,
                                                and remainder in AH
            MOV     AV_TEMP, AL;        Copy result to memory
CODE        ENDS
END         START
```

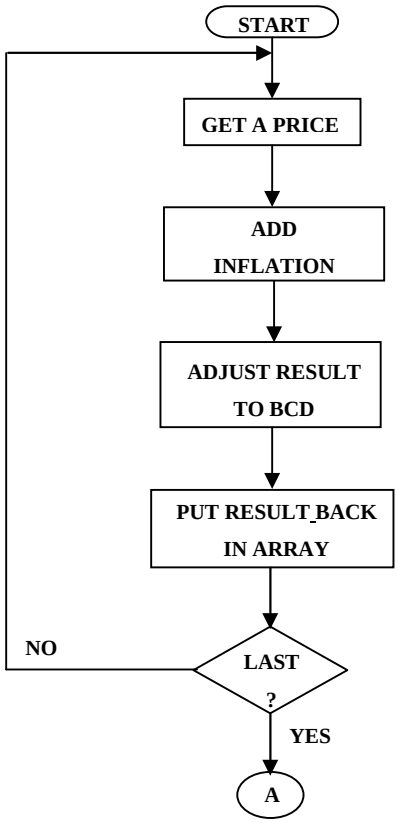
2.Write a program to add an inflation factor to a series of prices in memory and copies the new price over the old price:-

ABSTRACT: This programs adds an inflation factor to a series of prices in memory. It copies the new price over the old price.

REGISTERS: uses DS, CS, AX, BX, CX

PORTS: None used

```
ARRAYS      SEGMENT
COST   DB    20h, 28h, 15h, 26h, 19h, 27h,
             16h, 29h
PRICE DB     36h, 55h, 27h, 42h, 38h, 41h,
             29h, 39h,.
ARRAYS      ENDS
```



Flow Chart

```
CODE        SEGMENT
ASSUME      CS: CODE,    DS: ARRAYS

START:      MOV   AX, ARRAYS; Initialize data
            Segment
MOV   DS, AX        ; register
      LEA   BX, PRICES   ;      Initialize pointer
      MOV   CX, 0008h    ;      Initialize center
DO_NEXT: MOV AL, [BX] ;  Copy a price to AL
         ADD      AL, 03H       ;      Add inflation factor
         DAA
         MOV [BX], AL    ;      Copy result back to memory
         INC      BX             ;      Point to next Price
         DEC      CX             ;      Decrement counter
         JNZ DO_NEXT    ;      If not last, go get next
CODE     ENDS
END      START
```

3.Write a Program to find the largest and the smallest of the Array:-

The program uses the JGE (Jump greater than or equal to) instruction, because we have assumed the array values as signed. JAE instruction works correctly only in the case of unsigned numbers.

The smallest and the largest variables are first initialized to the first number in the array. They are then compared with the other array values one by one. If the value happens to be smaller than the assumed smallest number or larger than the assumed largest value, the smallest and the largest variables are changed with the new values respectively. DI points to the current array value and LOOP instruction is used to scan the array elements.

```
CODE        SEGMENT
            MOV   AX,    DATA
            MOV   DS,     AX            ;  initialize DS
            MOV   DI,    OFFSET ARRAY;  DI points to the arrange
            MOV   AX,    [DI]          ;  AX contains the first element
            MOV   LARGEST,    AX      ;  initialize largest
            MOV   SMALLEST,   AX      ;  initialize smallest
```

```
        MOV   CX,    6              ; loop counter
A1:     MOV   AX,    [DI]           ; get array value
        CMP   AX,    SMALLEST       ; [DI] = smallest?
        JGE   A2                    ; yes: skip
        MOV   SMALLEST,   AX        ; no: more [DI] to smallest
        JMP   A3                    ; as it is smallest, thus no need to
                                        compare it with the largest.
A2:     CMP   AX,    LARGEST        ; [DI] = largest
        JLE   A3                    ; yes:  skip
        MOV   LARGEST,    AX        ; no:  move [DI] to largest
A3:     ADD   DI,    2              ; point to next number
        LOOP  A1                    ; repeat the loop until CX = O
        INT   3h                    ; halt, return to DOS
CODE    ENDS
DATA    SEGMENT
        ARRAY     DW   -1, 2000, -4000, 32767, 500,0
        LARGEST   DW    ?
        SMALLEST  DW    ?
DATA    ENDS
END
```

4. Write a program to add two 5-byte numbers in the given two different arrays:

ABSTRACT:

This program adds a 5-byte number in one array, called NUM1 to a 5-byte number in another array called NUM2. The sum is stored in an array called SUM, with the state of the carry flag stored in byte 6 of Sum. The first value in each array is the LSB of that number.

REGISTERS : Uses CS, DS, AX, CX, BX, DX

```
DATA    SEGMENT
NUM1    DB    0FFh, 10h, 01h, 11h, 20h
NUM2    DB    10h,  20h, 30h, 40h, 0FFh
SUM     DB    6 DUP(0)
DATA    ENDS
    LEN EQU   05H   ;       constant for length of the array
```

```
CODE    SEGMENT
    ASSUME    CS:    CODE, DS:    DATA

START:  MOV   AX,    DATA        ; Initialize data segment
        MOV   DS,    AX          ; register
        MOV   BX,    00          ; Load displacement of 1st number
        MOV   CX,    00          ; clear counter
        MOV   CL,    LEN         ; set up count
        CLC                      ; clear carry. Ready for addition.
AGAIN:  MOV   AL,    NUM1[BX]    ; Get a byte from NUM1
        ADC   AL,    NUM2[BX]    ; add to byte from NuM2
        MOV   SUM[BX],    AL     ; store in Sum array
        INC   BX
        LOOP  AGAIN              ; continue unit 1 no more bytes
        RCL   AL,    01h         ; move carry into bit 0 of AL
        AND   AL,    01h         ; mask all but bit 0 of AL
        MOV   SUM[BX],    AL     ; put carry into 6th byte
FINISH  NOP
        NOP
CODE    ENDS
        END   START
```

16.5 SUMMARY:-

- In this lesson, we have covered some basic aspects of assembly language programming. We started with some elementary arithmetic problems, types of loops and slightly complex arithmetic.
- Assembly language gives you an access to the most of the hardware features of the machine, which might not be possible with high level language. We have also seen that some kind of applications can be efficiently executed in assembly language.
- A few sample programmes were given to understand the assembly language.

16.6 Key Terminology:-

Assembler: A program that translates all English – like source Program into machine – language, (binary code).

Assembly – language program: A program written using mnemonics and listed in four Fields: code, operand, label and comment.

Algorithm: A precisely defined set of steps for solving a problem

Execution: The part in the processing of instruction in which a command is carried out.

File: A named program, usually stored on disk.

Flow chart: A graphical representation of a program, using block symbols to represent functions.

Index Register: A register holding a memory address and used by certain instructions as a reference to locate a specific memory location.

Instruction: A computer command that can be decoded and used to direct a process

Instruction set: The list of instruction types recognized by a given microprocessor

Interrupt: An unscheduled request for special CPU action.

Machine language: The lowest – level computer language, written in binary or hexadecimal.

Mnemonic: A shorthand English – like symbol for an instruction type.

Programming: Placing a sequential list of instructions into the computer's memory.

16.7 Self assessment Questions:-

1. Write assembly program to add two bytes in the given two differed arrays.
2. Write assembly program to find the largest and the smallest of the array values.
3. Write assembly program for the average of two temperatures .
4. Write assembly language program, to add an inflation factor to a series of prices in memory and copies the new price over the old price .

16.8 References:-

1. Microprocessors and Interfacing – programming and hardware (2nd edition)
   – D.V. Hall.
   McGraw – Hill International edition.
2. INTEL Microprocessors – Hardware, software and Applications – Ray W. Goody
   McGraw – Hill International edition.
3. Micro computer systems: The 8086 /8088 family – Architecture, Programming
   And design - Yu – Cheng Liu & Glenn A. Gibson – PHI.
4. Introduction to Microprocessors for engineers and scientists (2nd edn) – Gosh &
   Sridhar, Eastern Economy Edition.

Name of the Author:

S.V. RangaNayakulu, M.Sc., M.Phil., M.Ed.,

Dept. of Physics & Electronics,
T.J.P.S. COLLEGE (P.G.COURSES)

GUNTUR-522 006