PHY 408

ELECTRONICS IV

Course Team

Dr. K. C. Okafor (Course Developer/Writer) **-** Federal University of Technology, Owerri, Nigeria
Prof. A. C. Ugwoke (Course Developer/Writer) - Enugu State University of Science and Technology, Enugu State, Nigeria

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
University Village
Plot 91, Cadastral Zone
Nnamdi Azikiwe Expressway
Jabi, Abuja

Lagos Office
14/16 Ahmadu Bello Way
Victoria Island, Lagos
e-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Published by the National Open University of Nigeria
Printed by NOUN Press

Printed 2018

ISBN: 978-978-8521-

COURSE GUIDE

Course Introduction

In general, this course focused on the following topics viz: Modulation and detection; amplitude modulation: square law modulator and detector, switching modulator, envelope detector, double sideband suppressed carrier (DSBSC) modulation, generation of balanced modulator, ring modulation, coherent detector of DSBSC waves, double-balanced modulator; Single sideband modulation (SSB) and demodulator; vestigial sideband modulation (VSB); frequency modulated (FM) systems,; high frequency oscillators, high input impedance circuits.

You will learn about the four essential pillars of the course. The four modules includes: communication system model, amplitude modulation types, detection/demodulation strategies, and frequency modulation/ultra high impedance systems. to simplify the course modules, each module has been broken down into units.

We started with Module 1 which introduced the communication system model. Under this module, Unit 1 discussed the model of a communication system; Unit 2 presented noise analysis; Unit 3 discussed signals computation while unit 4 looked at the introduction to modulation /types of modulation.

Module 2 focused on the amplitude modulation types. Under this module, Unit 1 presented the amplitude modulation concepts. Unit 2 focused on the types of Amplitude Modulation (AM) techniques. Unit 3 presented amplitude modulated signal generation while Unit 4 presented Amplitude Modulation Techniques.

Module 3 of this course presented Frequency Modulation and Ultra High Impedance Systems. In Unit 1 of this module, you will learn about the concept off frequency modulation and its related concepts. In Unit 2, you will learn about Frequency Modulation Transmitters. In Unit 3 you will learn about ultra high frequency tubes and oscillators. Unit 4 focused on high impedance circuits

In Module 4, we introduced detection/demodulation strategies in communication systems. The emphasis on AM detection/demodulation.

As you have seen, these subjects are well understood within the context of the communication systems.

In each unit, we outlined the focused objectives, main content, conclusion, summary and references. This makes for a simplified reading experience.

We wish you success.

MODULE 1 COMMUNICATION SYSTEMS MODEL

Unit 1 Model of a Communication System

Content

1.0 Introduction
2.0 Objectives
3.0 Main Content
    3.1. Model of Communication System
    3.1.1. Subsystems of Communication Systems
    3.1.1.1. Information Source
    3.1.1.2. Transmitter Block
    3.1.1.3. Channel
    3.1.1.4. Receiver Block
    3.1.1.5. Destination Block
    3.2. High Carrier Frequency in Communication Systems
    3.3. Modulation requirement
    3.4. Bandwidth requirement
    3.5. Non-sinusoidal wave Frequency Spectra
    3.6. RF Spectrum
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0 References/Further Reading

1.0 INTRODUCTION

With advancements of every society regarding educational, business, manufacturing, computing, agricultural sectors etc, there became an intense desire for individuals (humans), as well as objects to communicate with other individuals such relatives, friends, associates, etc across the globe. Interactions such as such Human to human (H2H), Machine to Machine (M2M), or even human to machine (H2M) usually involves communication. The science of communication involving long distances is called Telecommunication derived from the word "tele", meaning long distance.

In technical context, communication defines transmission, reception and processing of information by electronic means. The original communication systems viz: line telegraphy was invented in the eighteen forties. But radio communication became a reality in the beginning of the 20th century with the invention of triode devices. Radio communication has evolved immensely from the 2nd world war till date. This is as a result of the explosion in the invention of transistors, integrated circuits and other semi-conductor devices. Moore's law of transistor count still drives the electronic industries supporting telecommunication till date. In the recent years, communication has become more widespread with the use of satellites, fiber optics, and long term evolution networks (3-5Gs). Telemetry and radar systems play vital role in our military defense, navigation and even scientific research engagements.

Generally, a comprehensive learning of telecommunication systems as highlighted above will stimulate critical thinking on how to match human problems with technological solutions. Here, you will learn about communication systems model, constituents of communication system models, examples of communications systems including bandwidth requirements. Towards the end of this unit, the concept of frequency spectra in communication systems will be explained.

2.0 OBJECTIVES

After going through this unit, you should be able to:

- ✓ understand the concept of communication systems
- ✓ differentiate between different elements of communication systems
- ✓ understand the importance of using high carrier frequencies
- ✓ compare the different RF spectrum
- ✓ Understand the limitations of electrical communication systems.

3.0 MAIN CONTENT

3.1.Model of Communication System

To transfer information from point *A* to point *B*, another form of link is required between the two pints. A communication system therefore specifies the totality of mechanisms that offers the needed information link. For example, an electrical communication system could serve this purpose being the information link. The main idea behind any communication system is to successfully replicate the source information at the destination. As such, for successful communication to be achieved, the message received at the destination must remain identical to the original message evolving from the source. The different steps involved in the transmission of information are enumerated below.

i. Origin of information in the mind of the person who wants to communicate.
ii. Generation of message signal carrying the information.
iii. Converting the message signal into electrical from using a suitable transducer.
iv. Processing the message signal such that it will have the capability to travel for a long distance.
v. Transmission of the processed message signal to the desire destination
vi. Reception of the processed message signal at the desired destination
vii. Processing the received message signal in such a way to recreate the original non-electrical form
viii. Finally, delivering the inform from the message signal to the intended person

A clear understanding of the basic issues in the above steps independent of the type of communication system will lead to comprehensive understanding of any communication system such as telephony, radio broadcasting, television broadcasting, radar communication, satellite communication, fiber optics communication, computer communication and wireless communication. These are typical examples of communication systems.

We can use a block diagram to represent all forms of communication systems. Let us consider a typical communication system model in Figure 1, and discuss the various constituents that functions to achieve a reliable communication sequence. Generally, any communication system will have five blocks. For the purposes of simplicity, we shall focus on three blocks only (transmitter, channel and receiver) since we have little or no control over information source and destination.

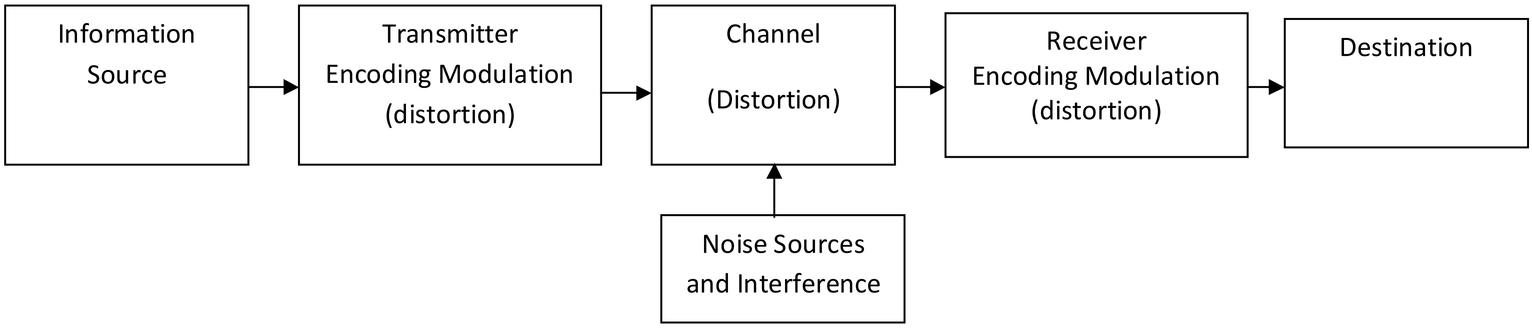


Figure 1.1. Block diagram of a general communication system

3.1.1. Subsystems of Communication Systems

From Figure 1.1, modern communication system can be summarized into:

i. Collation/sorting, processing and storage of information
ii. Actual transmission of information involving further processing and also reducing noise.
iii. Reception of information involving such processing steps as decoding, storage and interpretation.

In context, we shall define and discuss a few important terms such as information, message, and signal, channel, noise and distortion, modulation and demodulation etc.

3.1.1.1. Information Source

As previously explained, the aim objective of any communication system is to convey information form one point to another. The information comes from the information source which originates it. Messages originate in the information source. Essentially, there may be several messages in form of words, group words, code symbols, etc.

From various messages, other desired message is selected and transmitted. This is especially applied in telegraphy rather than entertainment broadcasting. It however, still applies to all forms of communications. The information in context represents the abstract level of anything intended for communication such as news, speech signal, written script or picture. It is converted in form a physical quantity.

Also, in electrical communication systems, we are interested in transmitting the information presented as the message signal to the receiving point as efficiently as possible. However, this message signal will be in non-electrical form.
For the purposes of electrical communication, this could be converted to electrical signal from using a suitable transducer. Transducers are devices used to convert energy in one form to another. For instance, if I choose to convey my thought that *I want to buy a new television* to my friend via speech mode, then the information will be manifested as the speech signal. *I*

want to buy a new television is the information and the speech corresponding to it is the message signal. The speech signal is the acoustic pressure variations plotted as a function of time.

These acoustic pressure variations are converted into electrical form using microphone as the transducer. The electrical version of the message is the actual input to the transmitter block of Figure 1.1.

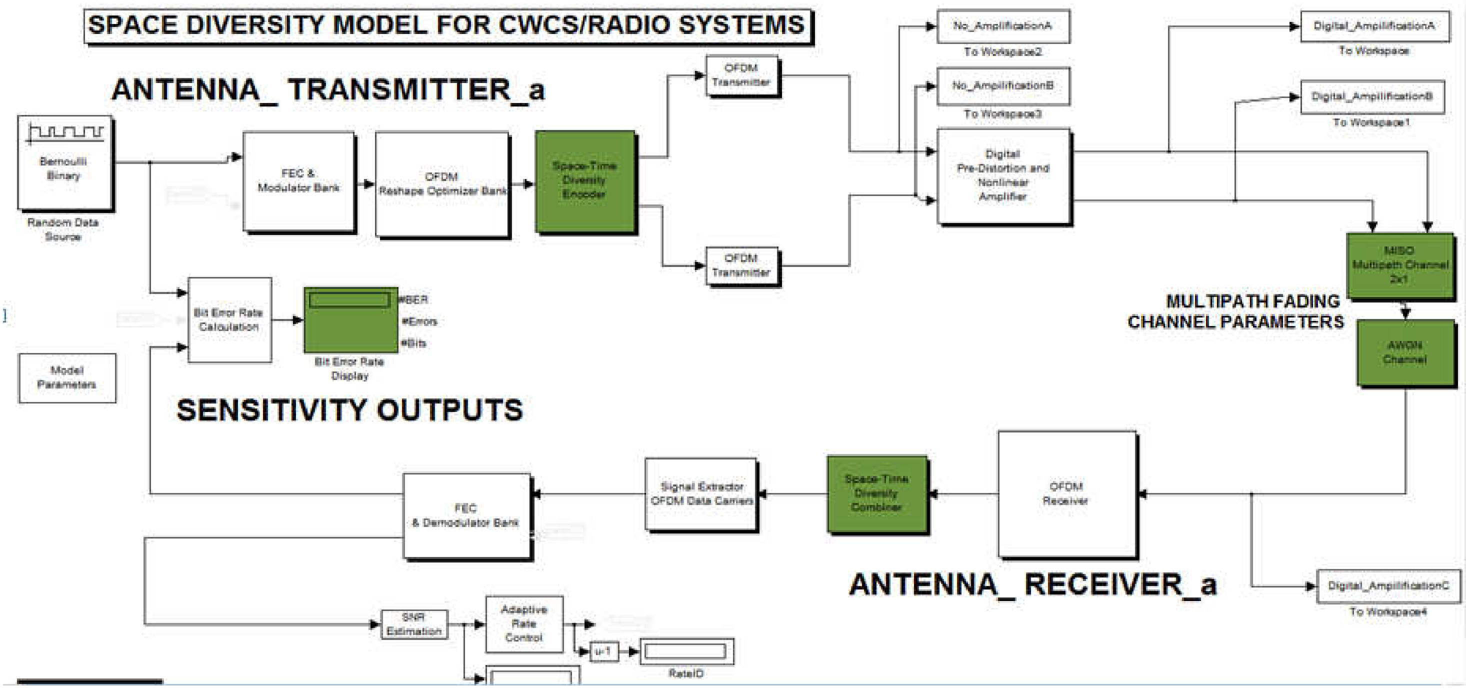


Figure 1.2a. Radio model design (message)

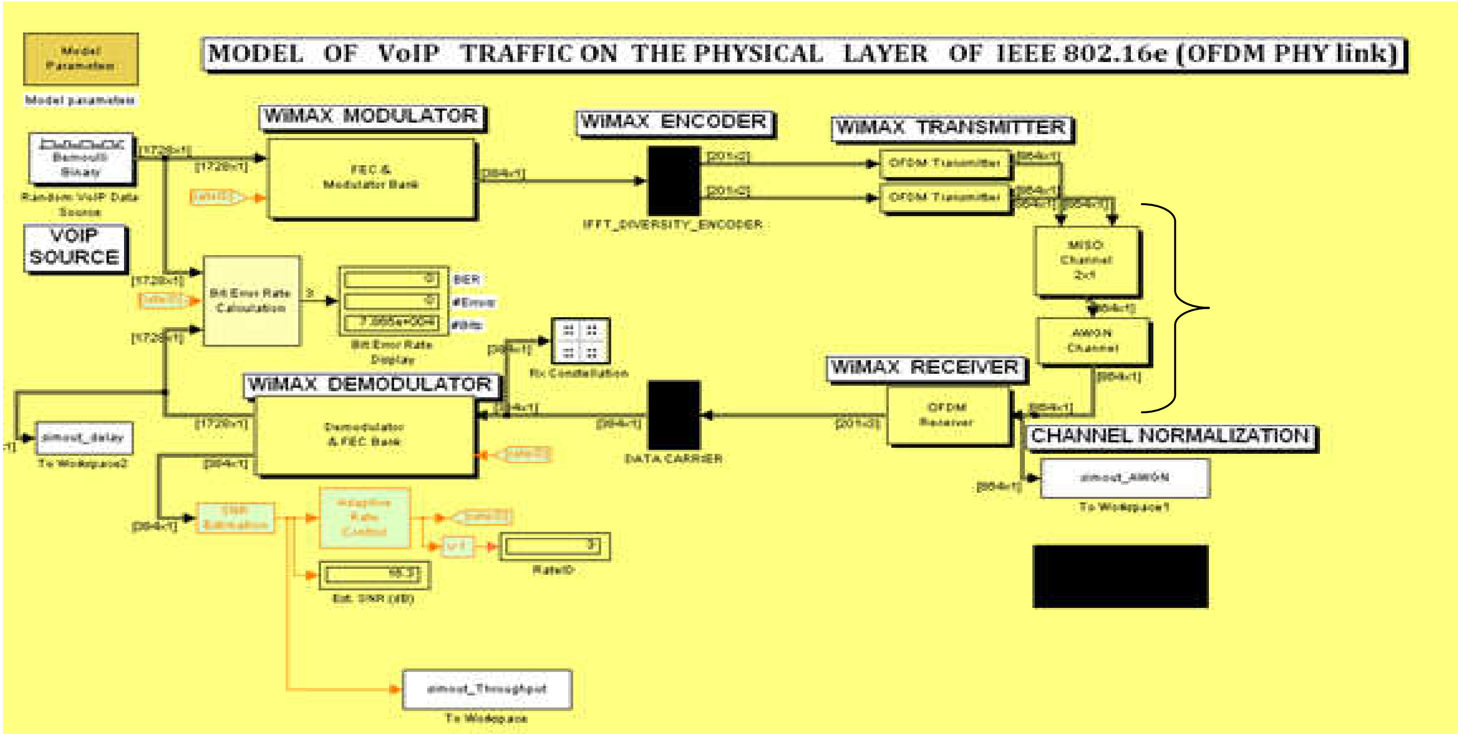


Figure 1.2b. WiMax Radio model design (Voice)

Figure 1.3 shows a physical antenna (as a transducer) used to transmit the modulated signal from the modulator into the atmosphere. It converts the signal energy in guided wave form to free space electromagnetic waves.

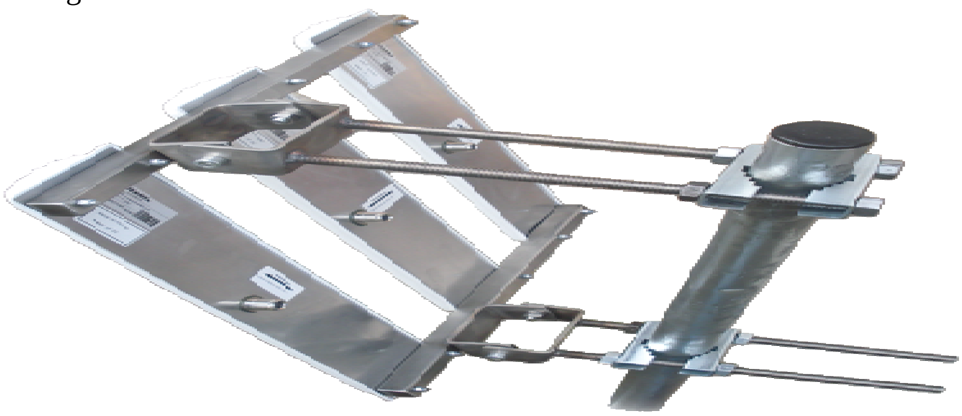


Figure 1.3. Radio antenna model

3.1.1.2. Transmitter Block

The main objective of the transmitter block is to collate the incoming message signal and modify it using a suitable signal conditioning scheme (transducer) such that it can be transmitted via the chosen channel to a receiving destination shown in Wimax setup in Figure 1.2b. Figure 1.4 shows the block diagram of a high level amplitude modulated broadcasting transmitter. The channel is the physical medium that connects the transmitter block with the receiver block. The functionality of the transmitter block is mainly determined by the nature and type of the channel chosen for the communication. In long distance radio broadcast communication, signal amplification is employed before modulation. The signal or message in a radio transmitter is amplified in several stages of small signal amplifiers (voltage amplifiers) and large signal amplifiers (power amplifiers) and may be encoded to make it suitable for transmission and subsequent reception.

Finally, the signal or information amplitude modulates the carrier (high frequency sine wave) as shown in Figure 1.4. The actual modulation technique will of course differ from system to system. As such, the modulation may be done at high carrier level or low carrier level. Figure 1.4 illustrates the high level modulation system. In this case, the stable carrier voltage of desired frequency is generated via a master oscillator such as crystal oscillator.

The buffer amplifier stage follows the oscillator clock interface and it is used to isolate the master oscillator from the influence of modulation done at a later stage. The buffer amplifier is followed by a chain of R.F voltage and power amplifiers and finally fed to the RF output power amplifier where the carrier gets amplitude modulated. We can equally call Figure 1.4 a high level broadcast transmitter.

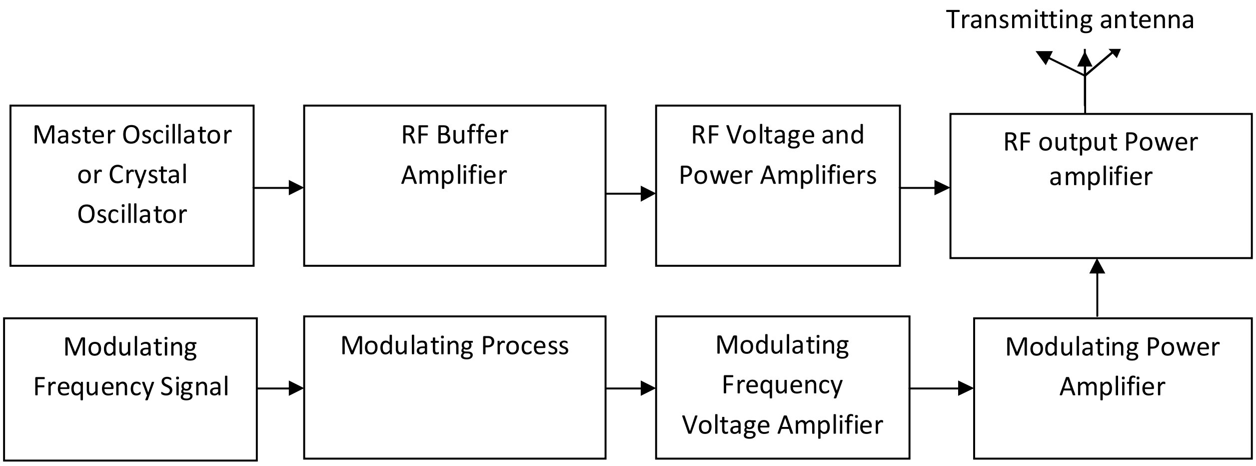


Figure 1.4: A Typical block diagram of a radio broadcast transmitter

In general, the transmitter block diagram of Figure 1.4 has several operations such as amplification, generation of high-frequency carrier signal, modulation and then radiation of the modulated signal. The amplification process essentially involves amplifying the signal amplitude values and also adding required power levels.

Now, the high-frequency signal is basically carrying out the modulation operation. Remember that this high-frequency signal commonly called the carrier is generated by the stable crystal oscillator. The carrier signal is characterized by these three parameters viz: amplitude, frequency, and phase. Again, the modulation process (which we shall discuss in details in Unit 2) involves varying one of these parameters in accordance with variation of the message signal.

3.1.1.3. Channel

As shown in Figure 1.1, the channel also called cloud or wave is the physical medium which connects the transmitter with the receiver. Examples of such physical medium include copper wire, coaxial cable, fiber optic cable, wave guide and free space or atmosphere. The choice of a particular channel depends on the feasibility and also the purpose of communication system. For instance, if the aim is to provide connectivity for speech communication among a localized segment of people, then copper wire may be the best choice. If the information needs to be sent to scattered geographically area like radio and television broadcasting, then free space or atmosphere will be the best option. The term channel is used to indicate the frequency range allocated to a given service or transmission. A television channel for example occupies a bandwidth of 7MHz while an amplitude broadcast channel occupies a bandwidth of 10Khz.

The nature of modification of message signal in the transmitter block is based on the choice of the communication channel. This is because the message signal should smoothly travel through the channel with least opposition so that maximum information can be delivered to the receiver. The message signal in the modified form travels through the channel to each the entry point of the receiver.

During the process of transmission and reception, the signal gets deteriorated due to i) distortion in the system and ii) noise introduced in the system.

This is illustrated in Figure 1.5. The noise so introduced is an unwanted energy, usually of a random character and may be caused by various sources. The noise gets superimposed on the signal. With sever noise; the signal-to-noise power ratio becomes so poor that the signal becomes unintelligible and hence useless. Noise added in Figure 1.1 has its greatest damaging effect when the signal is weakest. This implies that the noise in the channel or at the input to the receiver is most effective in deteriorating the signal-to-noise ratio.

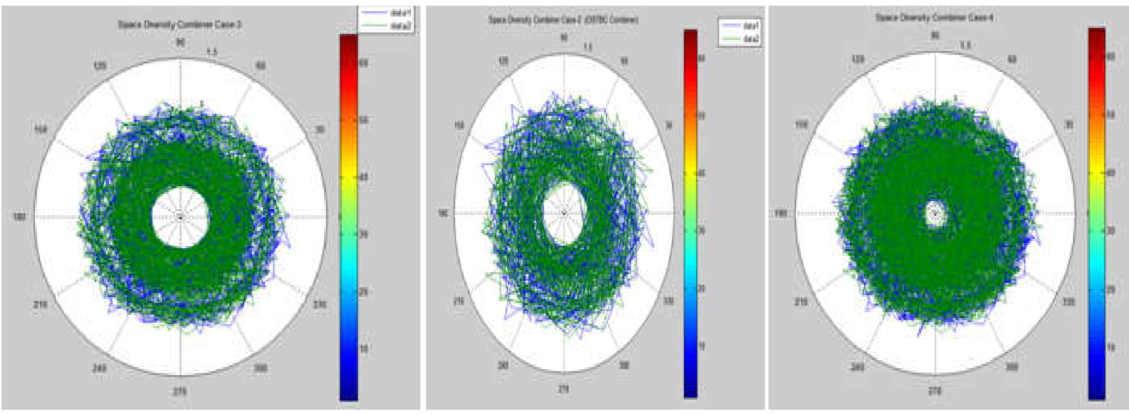


Figure 1.5. Channel noise signal deterioration impact on a communication system

3.1.1.3. Receiver Block

The receiver block receives the incoming modified version of the message signal from the channel and processes it to recreate the original (non-electrical) form of the message signal. There are a great variety of receivers in communication systems, depending on the processing required to recreate the original message signal and also final presentation of the message to the destination. Receivers in communication systems are of widely different varieties depending on the system requirements such as modulation system used, operating frequency, range of the system, type of display, etc.

Most of the receivers do conform broadly to the super heterodyne type, as does the simple broadcast receiver whose block diagram is shown in Figure 1.6.The super heterodyne receiver includes processing steps like reception, amplification, mixing, demodulation, and recreation of message signal. Among the different processing steps employed, demodulation is the most important one which covers the message signal available in the modified form to the original electrical version of the message.

Hence, demodulation is usually the inverse operation of modulation. The purpose of receiver and the form of output display influence its construction as much as the type of modulation system used. Accordingly, a receiver can be a very simple crystal receiver with headphones, to a far more complex radar receiver, with its involved antenna arrangements and visual display system. The output of a receiver may be fed to a loud speaker, video system unit, teletypewriter various radar displays, television picture tube, pen recorder, computer. In each instance, different arrangements must be made, each affecting the receiver design. Usually, the transmitter and receiver must be compatible with appropriate modulation methods, coding methods, time or synchronization and the receiver demodulation methods etc.

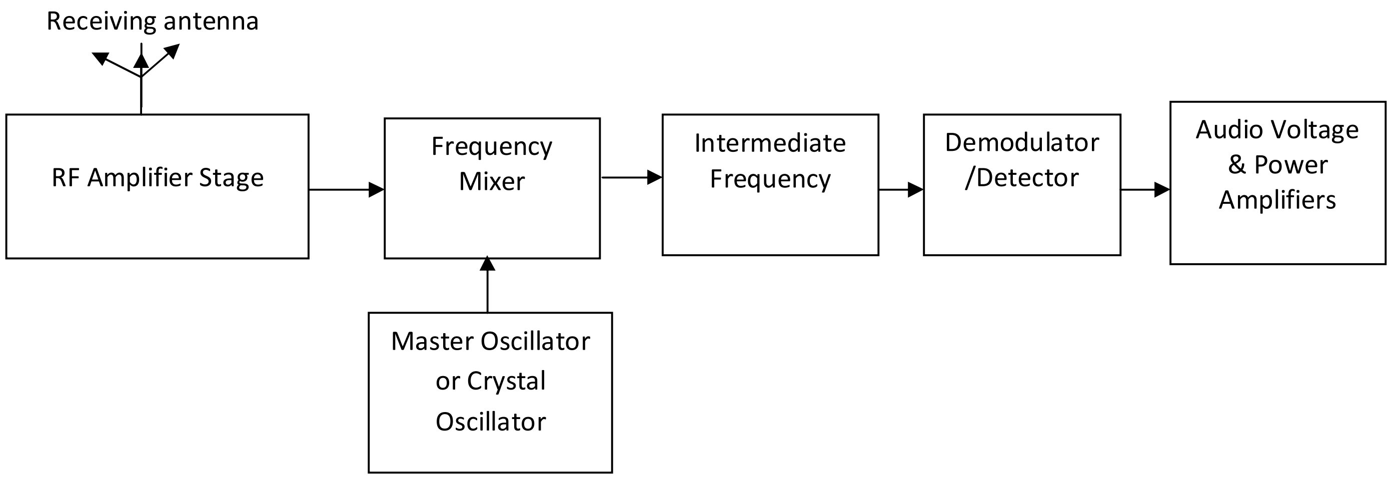


Figure 1.6. Channel noise signal deterioration impact on a communication system

Let's now discuss the pathology of Figure 1.6. Now, in the simple super heterodyne broadcast receiver shown in Figure 1.6, the voltage induced in the receiving antenna (such as Figure 1.3) is fed to R.F amplifier. The amplified carrier voltage from the R.F amplifier and the local oscillator voltage are heterodyned or mixed in the mixer stage resulting in intermediate frequency (I.F) voltage. This I.F voltage is amplified in I.F amplifier and then fed to the detector, at the output of

which we get the original signal voltage. This audio voltage is amplified and fed to the loud speaker or any other output interface as explained previously.
Figure 1.7 shows a simplified block diagram of a typical colored television receiver. We will refer to it throughout this book. It shows that any receiver must be capable of extracting information from the incoming channels to which it is tuned. The shaded blocks show examples of where demodulation or detection occurs for the video and audio signals.
Generally the sound uses frequency modulation (FM), while the video signal uses amplitude modulation (AM). The age of digital television and modern data communications will use other techniques. All these methods will be discussed later in this book.

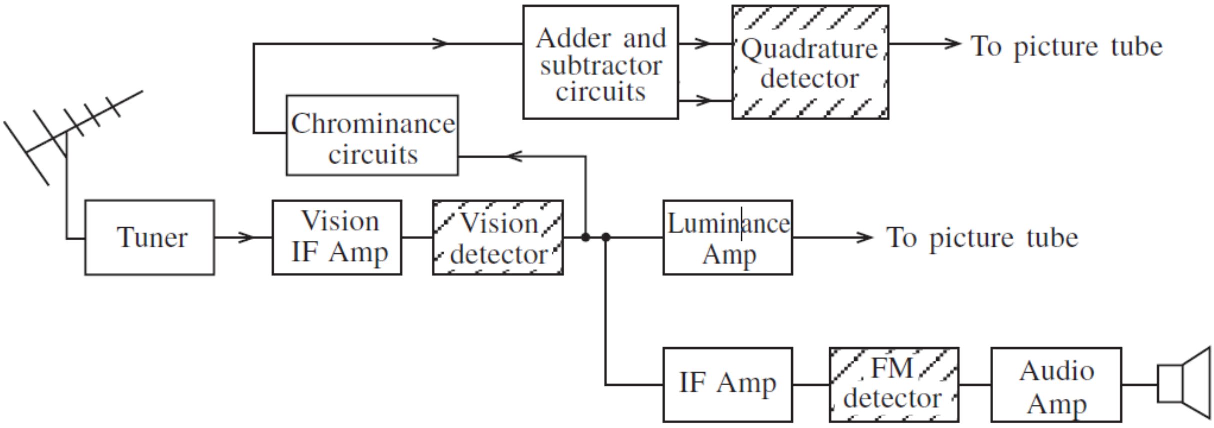


Figure 1.7: A simplified block diagram of a typical colored television receiver.

3.1.1.4. Destination Block

Recall that from Figure 1.1, it was established that the destination is the final block in the communication system which receives the message signal which is processed in order to comprehend the content information it carries.
The human beings are usually, the destination block. The incoming message signal via speech mode is processed by the speech perception system to comprehend the information. Similarly, the message signal via video or visual scene and written script is processed by the visual perception system to comprehend the information. The comprehension of the information from the message signal requires a robust human perception system in extracting the information even under very noisy condition otherwise it will amount to waste of efforts.

3.2. High Carrier Frequency in Communication Systems

A good question usually posed to students is why do we need high frequency carrier in Figure 1.1? We shall discuss few concepts in this regard below. We must understand that the communication carrier frequency is determined by the following factors such as:

- i. Radiation efficiency
- ii. Size of antenna
- iii. Ease of selection of radio signal in the receiver
- iv. Range of communication distance
- v. Operating frequency etc.

The transmitting antenna radiates efficiently, i.e., radiates a fraction of the total power fed to it if its dimensions are comparable with the carrier wavelength. The size of the antenna conductor (see Figure 1.3), is therefore inversely proportional to the carrier frequency. The implication is that the higher the carrier frequency, the smaller and hence economical is the antenna structure needed. In this regard, higher carrier frequencies are mostly preferred in communication systems. Furthermore, the higher the carrier frequency, the better is the selection of signal in the receiver. In this regard also, higher carrier frequency is preferred.

3.3. Modulation requirement

It is known that to transmit information some form of frequency transformation is necessary since we cannot send input signals directly over the channel. As such as carrier wave (radio frequency signal) whose attributes are well suited to the transmission medium is modified to represent the message signal.

By definition, modulation is the process of varying some characteristics of a high frequency sinusoidal voltage called carrier voltage in consonance with the instantaneous value of another voltage called the modulating voltage. The carrier voltage V_C is usually represented by

$V_C = V_a Cos(\omega_t + \phi)$ (1.1)

Where

$V_C =$ is the instantaneous value of carrier voltage

$V_a =$ is the amplitude of the carrier voltage

$\omega_t =$ is the angular frequency of the carrier voltage in radians/second

ϕ = is the phase angle relative to some reference

From Equ (1.1), we can vary any of the three parametric quantities viz: amplitude, frequency, and phase angle in consonance with the modulating voltage. The resulting modulations are referred to as the amplitude modulation (AM), frequency modulation (FM) and phase modulation (PM) respectively. The reasons for modulation and the various types of modulation is discussed in Section 3.3.1.

3.3.1. Reasons for Modulation

Are there reasons for modulation? Sure! Regardless of the type of modulation methods used (amplitude, or angle), analogue or digital, the reasons for modulation include:

1. For Ease of radiation since by utilizing the frequency translation property of modulation, signals can be impressed on a high frequency carrier, thereby permitting substantial reduction of antenna size.
2. For frequency Assignment. Modulation allows several radio and TV sources to broadcast simultaneously at different carrier frequencies and allows the owner of a radio or TV set the option of selecting one of the several stations even when all the stations are broadcasting similar program in the same transmission medium.
3. For Noise and Interference reduction. By using certain types of modulation schemes that enforces wide band noise reduction, this can address noise and interference issues at the expense of Bandwidth for noise suppression.
4. For Signal multiplexing involving several signals transmitting between the same two points. Modulation by signal multiplexing allows multiple-signal transmission on one channel such that signals can be picked out at the receiving end.
5. Equipment Limitation. Using modulation to achieve frequency translation eradicates the constraints of communication system equipment unavailability. Modulation can be used

to convert wideband signal to a narrowband signal thereby optimizing equipment availability issues.

3.4. Bandwidth requirement

The frequency bandwidth required for a given transmission depends on the maximum modulation frequency and the nature of modulation. The modulating voltage signal is hardly ever a single sine wave voltage. In practical context, it consists of several sine voltage waves having frequencies extending over the audio frequency band. Thus electrical signal produced by a speech or musical instrument may extend over the audio frequency range of 20Hz to 20Hz. This is rather a wastefully large frequency bandwidth. Hence, in practice, it is narrowed to range 0-3kHz for telegraph, 0-5kHz for AM broadcast and 0-15kHz for FM broadcast. Radio telephony or carrier current telephony uses a single sideband working using amplitude modulation. Hence, bandwidth requirement per channel is basically 3 kHz. AM broadcast uses double-sideband system thus requiring per channel bandwidth of 2x5 = 10kHz. FM produces sidebands extending theoretically up to infinity. However, the significant sidebands extend usually only up to fifth order sideband. Hence, with maximum modulating frequency of 15 kHz, the total bandwidth requirement per FM broadcast channel is 2x5x15 = 150 kHz.

From the above explanations, we have seen that the bandwidth of the modulating transmission depends upon the bandwidth of the modulating signal itself. In case the modulating signal consists of sinusoidal signals, the bandwidth occupied equals the frequency range between the lowest and highest sine wave signals. Instances of non- sinusoidal wave signals are quite complicated. The frequency spectrum occupied by non-sinusoidal wave signals can be determined by using Fourier series analysis.

3.5. Non-sinusoidal wave Frequency Spectra

It is important to understand that any non-sinusoidal single valued periodic wave such as square wave, triangular wave or saw-tooth wave may be divided via the process of Fourier series analysis into a series of cosine and sine waves consisting of a fundamental frequency (equal to the repetition rate of the non-sinusoidal wave signals) and harmonics as well. There are an infinite number of such harmonics.

Hence, a non-sinusoidal periodic wave of repetition rate 100times per second will consist of terms of i) Fundamental frequency 100Hz and ii) harmonics at 200Hz, 300Hz, 400Hz, etc. It cannot contain any other frequency components. In some instances, only even harmonics or only odd harmonics may be present depending on the waveform of non- sinusoidal wave. Although the harmonics theoretically extend up to infinity, but in general, the higher the harmonic, the lower its relative amplitude. The presence of sine wave components as given by Fourier series analysis may be verified by the following two methods:

i. Graphical synthesis involving the drawing of sine wave components taken from the Fourier series analysis formula and adding up all the components
ii. Wave Analyzer which is a high gain tunable amplifier with a narrow pass band. In this case, we may tune to each component sine wave in turn and measure its amplitude.

We shall talk more on Fourier series analysis in Unit 2 of this lecture course material.

3.6. RF Spectrum

We shall discuss the classification of the RF spectrum considering the wide range of frequencies available within the communication systems model. The RF spectrum covers carriers of wide range of frequencies used in radio communication systems for different services. Table 1.1 shows the classification of electromagnetic spectrum used for radio communication. It presents the propagation characteristics and the type of services it supports.

Table 1.1: Standard Classification Spectrum of Frequency used in communication

Radio Carrier Frequency	Free Space Wave length (meters)	Class	Propagation Characteristics	Radio Services
10-30KHz	3×10^4-10^4	Very low Frequency (VLF)	Available at all times of day and year. Attenuation is Low	Long Distance Point to Point Communication
30-300KHz	10^4-10^4	Low Frequency (LF)	During day time absorption exceeds that with VLF. During night time, propagation is similar to VLF	Long Distance Point to Point Communication Navigation
300-3000KHz	10^3-10^2	Medium Frequency (MF)	Day time attenuation is Low (Ionospheric propagation)	Broadcasting; ship to shoe communication
3-30MHz	10^2-10	High Frequency (HF)	Ionospheric propagation	National and international broadcast; point-to-point telephone and telegraph communication, Aviation
30-300MHz	10-1	Very High Frequency (VHF)	Tropospheric propagation (typical range equals line of sight)	Radar; Television; F.M Broadcast, Short distance communication
300-3000MHz	1-0.1	Ultra High Frequency (UHF)	Same	Facsimile; Television Relay; Air navigation
3000-3000MHz	0.1-0.01	Super High Frequency (SHF)	Same	Radar Navigation, Radio Relay

4.0 CONCLUSION

In this introductory unit, you have learnt about the basic concepts of a general communication system. You have not only learnt about the principles of communication system, but also about the basic constituent of a communication systems, typical examples of communication systems, High Carrier Frequency in Communication Systems, Modulation requirement, Bandwidth requirement, Non-sinusoidal wave Frequency Spectra and the classification of RF spectrum.

5.0 SUMMARY

In this unit, we have learnt about the basic concepts of a communication system. Its basic elements include:

- The signal collation/modulation, encoding, processing and storage of information units
- Base band transmission of information through channel
- Signal reception involving decoding, storage and interpretations

We have learnt that transmitters consist of:

- Transducer which converts information into electrical signals.
- Modulation processing units with frequency voltage and high power amplifiers.
- Master oscillator.
- RF buffer amplifier cascade for isolation of the master oscillator.
- RF Voltage and power amplifiers.
- RF output power amplifier comprising the modulated amplifier.
- The transmitting antenna.

We explained that the channel denotes the medium used to propagate the message from the transmitter to the receiver while the receiver such as super heterodyne radio receiver normally consist of

- RF amplifier used to amplify deteriorated weak RF signals
- Local oscillator.
- Frequency mixer used to mix or heterodyne RF signals with the local oscillator output, hence producing an intermediate frequency signal (IF).
- IF amplifier.
- Detector circuits used to re-generate the original base band/ audio signals.
- Audio frequency voltage and power amplifiers.
- Speaker transducer to reproduce the original sound signal.

Also, we explained the need for high carrier frequency in relation to transmitting antenna and showed the classification of RF spectrum viz: Very low Frequency (VLF)- 10-30KHz, Low Frequency (LF)- 30-300KHz, Medium Frequency (MF)- 300-3000KHz, High Frequency (HF)- 3-30MHz, Very High Frequency (VHF)- 30-300MHz, Ultra High Frequency (UHF) - 300-3000MHz, Super High Frequency (SHF)- 3000-3000MHz.

6.0 TUTOR-MARKED ASSIGNMENT

1)Mention the elements of a communication system. Briefly describe their individual functionalities using a labeled diagram.

2) List the basic functions of a radio transmitter and the corresponding functions of a receiver.
3) Enumerate the typical frequency ranges for the following RF spectrum classification: MF, HF, VHF, and UHF.
4) Explain why radio wave carrier performs critical role in radio communication.
5) Explain the difference between the term “message” and “information”.
6) Draw the block diagram of a typical high-level broadcast A.M transmitter. Discuss the function of each stage.
7) Why is a high carrier frequency needed in a communication system?
8) List the factors that the bandwidth requirement of a communication system depends on.

7.0 REFERENCE/FURTHER READING

George Kennedy, “Electronic Communication Systems”, 3rd Edition, McGra Book Company, New York.

G.K. Mithal, “Radio Engineering”, 12th Edition, ISBN: 81-7409-014-2, Romesh Chander Khanna Publishers, Delhi, India.

James Flynn and Sharlene Katz(2010); Introduction to Communication Systems, https://www.csun.edu/~skatz/katzpage/sdr_project/sdr/comm_intro_07_01_2010.pdf; Date Accessed 10/10/2017.

John G. Proakis Masoud Salehi (2002) Communication Systems Engineering By Prentice-Hall, Inc. Upper Saddle River, New Jersey, ISBN 0-13-095007-6

Upamanyu Madhow (2014); Introduction to Communication Systems University of California, Santa Barbara January 17. Online URL - http://www.ece.ucsb.edu/wcsl/Publications/intro_comm_systems_madhow_jan2014b.pdf

Jossey-Bass/Pfeiffer (1998); Basic Communication Model, The Pfeiffer Library Volume 25, 2nd Edition.

UNIT 2 NOISE IN COMMUNICATION SYSTEMS

CONTENTS

1.0 Introduction
2.0 Objectives
3.0. Main Content
3.1. Classification of Noise
3.1.1. External noise
3.1.2. Internal noise
3.2. Atmospheric Noise
3.3. Extra-terrestrial Noise
3.3.1. Solar noise
3.3.2. Cosmic noise
3.3.3. Industrial/Man-made noise
3.3.4. Thermal Agitation noise
3.3.5. Short noise
3.3.6. Transit Time Noise
3.4. Internal Miscellaneous Noises
3.5. Noise cascade in Complex Networks
3.5.1. Derivation of Additive Noise
3.5.2. Practical Additive Noise
3.5.3. Noise in Reactive Circuits
3.6. Signal-to-Noise Ratio
3.7. Noise Figure
3.7.1. Noise Figure Computation
3.7.2. Determination of Noise Figure in Terms of Equivalent Noise Resistance
3.7.3. Measurement of Noise Figure
3.8. Noise Temperature
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0 References/Further Readings

2.0 INTRODUCTION

Noise is a very important topic in communication systems that must be properly studied in order to understand its underlying implications in communication systems. Electrical disturbances normally interfere with systems thereby producing noise in the output of the electrical system. It is always present and in most cases affects the operational performance of critical systems.
Its measurement and evaluation seems to be quiet complex owing to its nature. For example, in an A.M broadcast receiver, noise may cause hiss in the loudspeaker output.

In a Television receiver, noise may appear in the picture tube in the form of “snow” (in black and white TV receiver) or colored snow (in color TV receiver). In pulse communication system, noise may result in unwanted pulses or may cancel one or more wanted pulse thereby causing serious errors in reproduction. Similarly, in Radar system, noise produces spiky pulses in the output reproduced on the indicator. It extensively affects the sensitivity of a receiver by placing a limit on the weakest signal that can be amplified. In most electronic communication system such as Radar system, noise tends to impose a limit on the bandwidth of the system.

2.0 OBJECTIVES

At the end of this unit, you should:

- ✓ be familiar with the concept, types and sources of noise.
- ✓ understand the methods of calculating noise produced by various sources.
- ✓ demonstrate the understand noise metrics such as signal-to-noise ratio, noise figure and noise temperature.
- ✓ be able to solve problems involving resistance and temperature induced noise.

3.0. MAIN CONTENT

3.1. Classification of Noise

A good question is what is Noise?

In electronic communication systems, noise may be defined as any form of energy which seeks to interfere and distort with proper reception and reproduction of wanted signal. Numerous disturbances of electronic or electrical format yields noise in the output of the system. There are different forms of noise whose categories are given below.

3.7.1. **External noise-** its sources are external to the receiver of the communication. External noises may be classified into the following three types:

i. Atmospheric noises
ii. Extraterrestrial noises
iii. Man-made noises or industrial noises

3.7.2. **Internal noise-**Here, its sources are obtained within receiver of the communication system. Internal noise may be classified into the following categories:

i. Shot noise
ii. Transit time noise
iii. Miscellaneous internal noises.

It is difficult to handle external noises quantitatively and in a given geographical location, these forms of noise are uncontrollable. Therefore, to reduce the adverse effect of external noise, the best option is to shift the system to another location having smaller noise level. For this reason, radio telescopes are always located away from industrial areas where various electrical processes produce large electrical noise.

Similarly, satellite earth stations are located in noise free valleys. Internal noises on the other hand can be handled quantitatively and can also be reduced by proper receiver design. Since this noise is randomly distributed over the entire frequency spectrum, the noise present in a given bandwidth *B* is the same at any frequency in the frequency spectrum. Thus the random noise power is proportional to the bandwidth over which it is measured. Considering the above enumerations, we shall now discuss them in details.

3.8. **Atmospheric Noise**

This is also referred to as static noise and is caused by lightning discharges in thunderstorms and other natural electrical disturbances occurring in the atmosphere. These electrical impulses are random in nature. Hence, the energy is spread over the complete frequency spectrum used for radio communication. Atmospheric noise basically consists of spurious radio signals with components spread over a wide frequency range. These spurious radio waves constituting the noise get propagated over the earth in the same fashion as the desired radio waves of the same frequency. In context, at a given receiving point, the receiving antenna picks up not only the signal but also the static from the entire thunderstorm, local or remote.

The field strength of atmospheric noise varies approximately inversely with the frequency. Thus large atmospheric noise is generated in low and medium frequency (broadcast) bands while very little noise is generated in the VHF and UHF bands. Further VHF and UHF components of noise are limited to the line-of-sight (less than about 80km) propagation. For these two reasons, the atmospheric noise becomes severer at frequencies exceeding about 30MHz.

3.9. **Extra-terrestrial Noise**

There are numerous types of extraterrestrial noises or space noises depending on their sources. However, these may be classified into two major groups viz: Solar noise and cosmic noise

3.9.1. **Solar noise**

This is the electrical noise emanating from the sun. Under quiet conditions, there is a steady radiation from the sun. This results because the sun is a large body at a very high temperature (exceeding 6000 deg C on the surface), and radiates electrical energy in the form of noise over a wide frequency spectrum including the spectrum used for radio communication. This solar cycle creates electrical disturbances with an increase in intensity.

3.9.2. **Cosmic noise**

This is found in distant stars or suns having very high temperatures. These stars therefore radiate noise in the same way as sun. The noise received from these distance stars is thermal noise or black body noise) and is distributed almost uniformly over the entire sky. Noise is also received from the center of our own galaxy (the Milky way), from other galaxies and from other virtual point sources such as quasars' and pulsers. This galactic noise is very intense but since it comes from very distance sources, the angle subtended by the earth is very small. Hence, the strength of galactic noise received on earth gets diminished.

At this point, it is feasible to conclude that the space noise is significant at frequencies in the range from about 8MHz to somewhat above 1.43giga-hertz (GHz), where this frequency 1.43GHz corresponds to 21cm hydrogen line. In the frequency range of 20 to 120MHz, the space noise forms the strongest noise component next only to the man-made noise.

However, below 20MHz, not much of the space noise penetrates through the ionosphere to reach the earth. Also, the space noise disappears at frequencies in excess of 1.5GHz probably due to its absorption by hydrogen in interstellar space.

3.9.3. **Industrial/Man-made noise**

This refers to the electrical noise produced by sources such as auto-mobile and air craft ignition, electric motors, and switch gears, leakage from high voltage lines, fluorescent lights, and

numerous other heavy electrical machines. Such noises are produced by the arc discharge taking palace during operation of these machines. These noises exceed all other sources in the frequency range extending from about 1MHz to 600MHz.
The nature of industrial noise is highly variable; hence it can be analyzed only statistically. Hence, man-made noise increases as the receiver bandwidth increases.

3.9.4. **Thermal Agitation noise**
The noise generated in a resistance or the resistive component of complex impedance is called thermal agitation, white or Johnson noise. It occurs due to the rapid and random motion of the molecules (atoms and electrons) inside the component itself.
From the perspective of thermodynamics, kinetic theory explains that the temperature of a particle is a way of expressing its internal kinetic energy. Thus the "temperature" of a body is the statistical root mean square (rms) value of the velocity of motion of the particles in the body. From this theory, the kinetic energy of these particles becomes approximately zero (i.e. zero velocity) at the temperature of absolute zero, i.e., 0 deg K (nearly -273^{o}C). Hence, it valid to say that the noise generated by a resistor is proportional to its absolute temperature, in addition to being proportional to the bandwidth over which the noise is to be measured.

$$P_n \alpha T \Delta B \qquad (2.1)$$

Therefore, we can obtain the following mathematical model for the maximum power output P_n of a resistor.

$$P_n = \overline{K * T}B \qquad (2.2)$$

$\overline{K}$ = Boltzmann constant (1.38$*$ 10^{-23} joule/deg.K)
T= Absolute temperature, K =273^{0}C + deg.C
B = Band with in hertz.

Note that random noise is always expressed in terms of rms value and not its instantaneous value. The peak noise voltage hardly ever exceeds 10times the rms value.
From Equ (2.1), we can draw an equivalent circuit of resistor as a noise generator as shown in Figure 2.1.

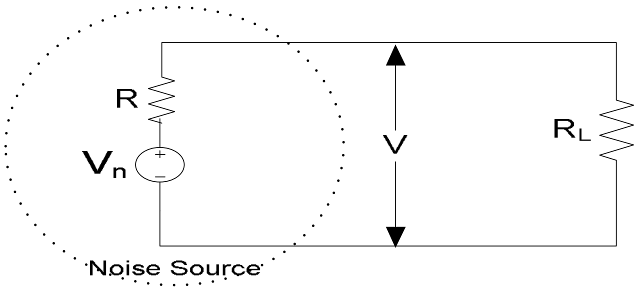


Figure 2.1. Resistor as a noise generator.

From this equivalent circuit, we may compute the resistor's equivalent noise voltage V_n. Now let a noiseless load resistor R_L be connected across the noise generator as shown in Figure.2.1. For maximum power transfer from noise source V_n to load resistor R_L, it is pertinent to maintain $R_L = R$. Then, the maximum noise power so transferred is given by Equ. 2.3.

$$P_n = \frac{V^2}{R_L} = \frac{V^2}{R} = \frac{(V_n/2)^2}{R} = \frac{V_n^2}{4R} \qquad (2.3)$$

$$V^2 = 4RP_n = 4R\overline{K * T}B \qquad (2.4)$$

$$V_n = \sqrt{4R\overline{K*}TB} \qquad (2.5)$$

From Equ. 2.4, we conclude that the square of the rms noise voltage associated with a resistor is proportional to the absolute temperature *T* of the resistor, value *R* of resistor and the bandwidth B over which the noise is measured. This is because noise is random and as such on an average evenly distributed over the spectrum.

Example 2

A resistor of value 20kΩ is connected at the input of an amplifier operating over the frequency range 10 to 11MHz. Compute the rms noise voltage at the input of the amplifier if the ambient temperature is 24^0C.

Solution:

Now T = K =273^0C + deg.C = 273^0C +24^0C = 397^0C.

$\overline{K}$ = Boltzmann constant (1.38∗ 10^{-23} joule/deg.K)

R= 20kΩ.

B =11-10MHz

From Equ.2.5,

$V_n = \sqrt{4R\overline{K*}TB}$

$V_n = \sqrt{4x1.38x10^{-23}x(273+24)x20x10^3x(11-10)10^6}$ volt = 18.1μV.

Example 3

The noise output of a resistor id simplified by a noiseless amplifier having gain of 40 and bandwidth of 40 KHz. A meter connected to the output of the amplifier reads 4mV rms. a) if the resistor is operated at 27^0C, what is its resistance? b) If the bandwidth of the amplifier is reduced to 10KHz, its gain remaining constant, what will the meter read now?

Solution:

$V_n = \sqrt{4R\overline{K*}TB}$

Hence, $R = \frac{Vn^2}{4R\overline{K*}TB}$

The *rms* noise voltage generated in the resistor V_{rms} = $\frac{4mv}{40}$=100μV

$R = \frac{(100x10^{-6})^2}{4x1.38x10^{-23}x(273+27)40x10^3}\ \Omega$

$R = 15.1x10^6\ \Omega$

b) Initially, $B = 40kHz$

Then, V_n=$\sqrt{4R\overline{K*}TB}$,

$$V_0 = A\sqrt{4R\overline{K*}TB}$$

Where *A* is the amplifier again. The subsequent bandwidth is dropped to 10kHz, ie. B' = B/4.

Hence, $V_0 = A\sqrt{4R\overline{K*}T\left(\frac{B}{4}\right)} = \frac{1}{2}A\sqrt{4RB} = \frac{1}{2}x4mV = 2mV.$

Example 3

A parallel tuned circuit having Q=10resonances at 10MHz with 10pF capacitor. If this circuit is maintained at 27^0C, what noise voltage will a wide band voltmeter measure when placed across it?

Solution.

$$\frac{1}{\omega_0 C} = \frac{R_0}{Q_n}$$

This implies that,

$R_0 = \frac{Q_0}{\omega_0 C} = \frac{10}{(2\pi x 10^7)x10x10^{-12}}\,\Omega$

$R_0 = 15.9k\Omega.$

$B = \frac{f}{Q_0} = \frac{10^7}{10} = 10^6\text{Hz}$

Hence, $V_0 = \sqrt{4x1.38x10^{-23}x(273+27)x15.9x10^3x10^6V}$ =16.22μV.

3.9.5. **Short noise**

Another form of noise is the short noise which results from short effect present in all amplifying devices and virtually all active devices. This is caused by random variations in the arrival of electrons or holes at the output electrode (or collector in the case of CE amplifier) of an amplifying device. Hence, the noise voltage appears as a randomly varying noise current superimposed on the direct current in the output circuit of the amplifying device. The noise voltage gets amplified by the subsequent amplifying stages and when fed to a loudspeaker causes noise as though a shower of lead shots were falling on a metal sheet. This substantiates the concept of shot noise.

In an active device, the average output current is governed by the different bias voltages. Obviously, the number of electrons migrating at the output electrode varies randomly with time. In bipolar junction transistor, this result mainly due to random drift of the discrete charge carriers (electrons or holes) across the junctions. The paths taken by different carriers are random and hence unequal. Hence although the average collector current is constant, minute variations result. Several variables are involved in the generation of shot noise. Hence usually approximate equations are used for shot noise except in the case of a diode. For a diode, the *rms* shot noise current in amperes is given by current in amperes in Equ. (2.6),

$$I_n = \sqrt{2qI_pB} \qquad (2.6)$$

Where q = Magnitude of the charge of an electron (1.6 x 10^{-19} coulomb)

I_p = Direct diode current, *Amp*.

B= Bandwidth of the system, *Hertz*

3.9.6. **Transit Time Noise**

In the very high frequency (VHF) range, the transit time (i.e., the time taken by an electron to travel from say the emitter to the collector in a transistor) becomes comparable to the periodic time of the signal being amplified. In this regard, Transit time effect is said to taken place. As a result of this, the input admittance of the transistor increases. The thermal noise generated in this input resistance then represents the noise contribution of the transit time VHF domain. You must understand that in a typical transistor, at frequencies beyond the one at which transit time noise becomes perceptible, this noise goes on increasing with the increase of frequency at a rate which soon approaches 6dB/octave. At such high frequencies, this noise predominates over other noise. Hence at such high frequencies, it is preferable to measure the noise rather than make an attempt to calculate the input equivalent noise resistance. Essentially, an RF transistor, however, have remarkably low noise.

3.10. **Internal Miscellaneous Noises**

We shall look at few of these types of noises below.

- Flicker Noise: Flicker noise otherwise known as modulation noise is the type that often appears in transistors operating at low audio frequencies. Flicker noise is proportional to the emitter current and junction temperature of a transistor. However, this noise is inversely proportional to the frequency. As such often neglected at frequencies above about 500 Hz, and therefore presents serious problem in practical circuits.
- Transistor Thermal Noise: Within the bipolar junction transistor, thermal noise is caused by the emitter, base and collector internal resistances. Out of these three regions, the base region contributes maximum thermal noise. Essentially, in a BJ transistor operating in the low frequency region from about 500Hz to about 0.2 F_a (alpha cut off frequency); the transistor noise is essentially constant. Hence, noise contributions by both shot noise and transistor thermal noise may be represented by an equivalent input resistance, *Req*
- Partition Noise: In transistor, partition noise is caused by the random fluctuation in the division of current between the collector and the base.
- Frequency Mixer Noise: Generally, Frequency mixers have more noise than amplifiers using identical devices, except at microwave frequencies. The higher noise in mixers have been attributed to the following two reasons:

i. Conversion transconductance $g_{cn}$of mixers is much lower than the transconductance $g_m$of the amplifiers

ii. Whenever, an image frequency rejection is inadequate, the noise associated with the image frequency also gets accepted. This creates higher noise in frequency mixers.

3.11. **Noise cascade in Complex Networks**

3.11.1. **Derivation of Additive Noise**

This refers to the additional noise elements due to several sources. Now, consider several thermal noise sources in series producing noise voltages $V_{n_1}, V_{n_2}, V_{n_3},$ etc., as derived in the previous sections.

Now, Let,

$$V_{n_1} = \sqrt{4\bar{k}TBR_1} \qquad (2.7)$$

$$V_{n_2} = \sqrt{4\bar{k}TBR_2} \qquad (2.8)$$

Then the sum of those rms noise voltages in series is given by the square root of the sum of their squares then the resultant rms noise voltages is given by,

$$V_{n_r} = \sqrt{V_{n_1}^2 + V_{n_2}^2 + V_{n_3}^2 + \cdots} = \sqrt{4\bar{K}TB(R_1 + R_2 + R_3 + \ldots\ldots)}$$
$$= \sqrt{4\bar{k}TBR} \qquad (2.9)$$

Where R is the sum of the individual resistances and is given by,

$$R = R_1 + R_2 + R_3 + \ \ldots.. \qquad (2.10)$$

The same procedure applies if one of these resistances is the equivalent input noise resistances amplifying devices. For computing the resultant noise voltage due to several resistors in parallel, we find the total resistances by standard method and then substitute this value of R in Equ. (2.9). obviously the total noise voltage is less than that caused by any of the individual resistance. However, the noise power remains the same.

Example 4.
Compute the noise voltage at the input of a video amplifier using a device having 300Ω equivalent noise resistances and 400Ω input resistors. The bandwidth of the amplifier is 7MHz and the ambient temperature is 27°C.
Solution.
T = 273 + 27 = 300k
R = (300 + 400) = 700Ω
Then the noise voltage is,

$$V_n = \sqrt{4\bar{k}TBR} = \sqrt{4x\ 1.38\ x10^{-23} x\ 300\ x\ 7x10^6 x\ 700\ volt.} = 9\ \mu V.$$

3.11.2. **Practical Additive Noise**
Figure 2.2 shows a good scenario for addition of noise due to several amplifiers in cascade. It illustrates a number of amplifier stages in cascade each having a resistance at its input and output. One might be required to find their combined effect on the noise of the overall system. Recall that in a super heterodyne receiver, the first stage is usually the R.F. amplifier while the second stage is a frequency mixer.
Let us look at finding the equivalent input noise voltage and its equivalent noise resistance for the complete receiver. This is the resistance which will produce the same amount of random noise in the output of the complete circuit as done the receiver. Accordingly, let us replace the actual receiver by an ideal noiseless receivers with an equivalent noise resistance located as its input. This method enables us to compare several receivers regarding noise and permits us to compute the lowest input signal which the receiver may amplify without drowning it under noise.
Consider the first two stages of the multi-stages amplifier depicted in Figure 2.2. The amplifier have gains A_1 and A_2respectively; with input resistances R_1 and R_2respectively. R_3Forms the output resistance

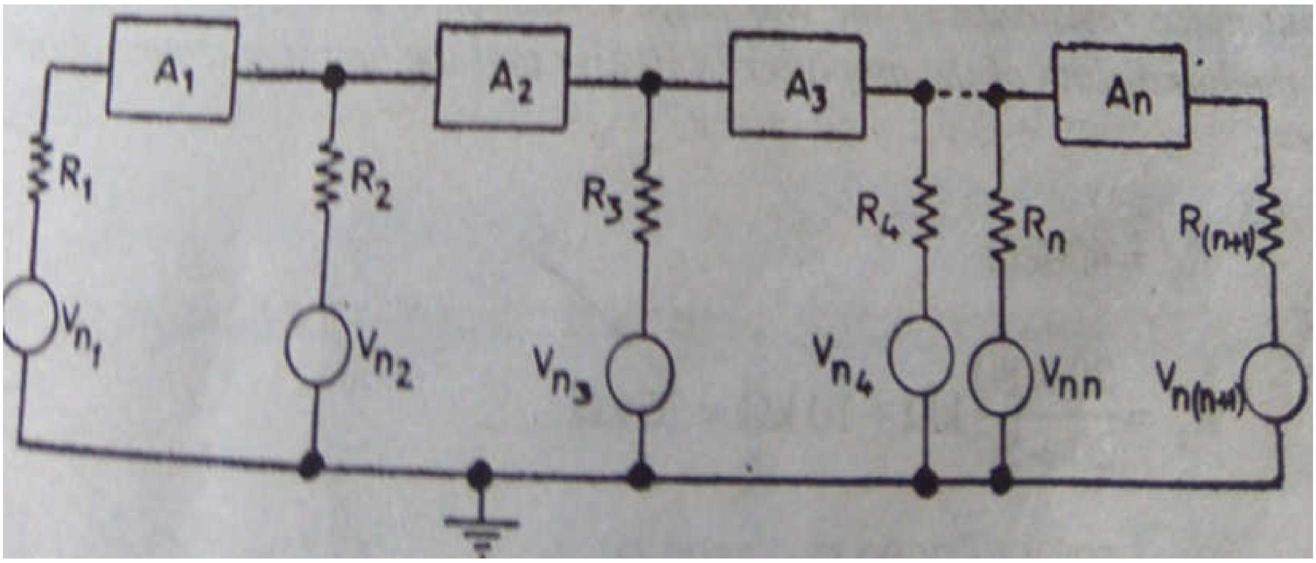


Figure 2.2. Noise of several amplifier stages in cascade.

The *rms* noise voltage at the output due to R_3 is given by,

$V_{n_3} = \sqrt{4\bar{k}TBR_3}$ (2.11)

The same noise voltage is present in the output if instead of R_3we have resistance R_3at the input of stage 2.

Then $V_{n_3} = \frac{V_{n_3}}{A_2} = \frac{\sqrt{4\bar{K}TBR_3}}{A_2} = \sqrt{4\bar{K}TBR_3}$ (2.12)

Where $R_3{}'$ is given by$R_3' = \frac{R_3}{A_2^2}$

Equ. (2.12) shows that a noise resistance on being transferred from the output of a stage to its input gets divided by the square of the voltage gain of the stage. The noise resistance already present at the input of the second stages is R_2. Hence the net noise resistance at the input of the second stage including the contribution made by R_3 is given by,

$R_{2t} = R_2 + R_3 = R_2 + \frac{R_3}{A_2^2}$ (2.13)

This resistance R_{21} may be transferred to the input of the first stages to contribute equivalent noise resistanceR_2' given by,

$R_2 = \frac{R_{2t}}{A_1^2} = \frac{R_2 + R_3/A_2^2}{A_1^2} = \frac{R_2}{A_1^2} + \frac{R_3}{A_1^2 A_2^2}$ (2.14)

Hence, the net noise resistance at the input of the first stage is given by,

$R_{eq} = R_1 + R_2' = R_1 + \frac{R_2}{A_1^2} + \frac{R_3}{A_1^2 A_2^2}$ (2.15)

With Equ. (2.15), it is possible to expand up to$n$-stage amplifier.

Usually, it does not make sense to consider more than three stages since the noise resistance at the input of the first stage is the most significant contribution to noise while the stage beyond the third stage contribute insignificantly to noise. This is true since the denominators in Equ. (2.15) have terms likeA_1^2, A_1^2, A_2^2, etc.

Example 5

The first stages of a two stage amplifier has output resistance of 20kΩ, voltage gain of 10, input resistance of 500 Ω and equivalent 2000 Ω. The second stage has output resistance of 400 kΩ, voltage gain of 20, input resistance of 80 kΩ and equivalent noise resistance of 10 kΩ. Compute the equivalent input noise resistance of the two stage amplifier. Also compute the equivalent input noise voltage given that the bandwidth of the amplifier is 10 kHz and the ambient temperature is 300 K.

Solution.

Given that R_3= 400kΩ,R_2 = $\frac{20x80}{20+80} k\Omega + 10k\Omega = 26k\Omega$, R_1 = 500 Ω +2000 Ω = 2500 Ω.

$R_{eq} = R_1 + \frac{R_2}{A_1^2} + \frac{R_3}{A_1^2 A_1^2}$

That is, R_{eq} = *2500* + $\frac{26{,}000}{(10)^2} = \frac{400{,}000}{(10\times 20)^2} ohms = 2670\Omega$

$V_{neq} = \sqrt{4\bar{k}TBR_{eq}} = = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 10^4 \times 2670\ volt}$ =0.677.μV

Example 6

A three stage amplifier operates over a bandwidth of 10 kHz at a temperature of 300 deg. K. the stages have respectively

i. voltage gain of 20, 25 and 25
ii. input resistor of 600 Ω, 40kΩ and 80 kΩ
iii. equivalent noise resistance of 1500 Ω, 6 kΩ and 10 kΩ and
iv. Output resistor of 30 100 kΩ and 1 MΩ. Compute the equivalent input noise resistance of the overall 3-stage amplifier. Also company the equivalent noise voltage of the first stage.

Solution.

$$R_1 = 600\Omega + 1500\Omega = 2100\Omega$$

$$R_2 = \frac{40x30}{40+30} k\Omega + 6k\Omega = 23.14k\Omega$$

$R_3 = \frac{80x\ 100}{80+100} k\Omega + 10\ k\Omega = 54.44\ k\Omega$= 1MΩ

Hence,

$$R_{eq} = R_1 + \frac{R_2}{(A_1)^2} + \frac{R_3}{(A_1 A_2)^2} + \frac{R_4}{(A_1 A_2 A_3)^2}$$

= 2100 + $\frac{23140}{400} + \frac{54.44\times 10^3}{(20\times 25)^2} + \frac{10^6}{(20\times 25\times 25)^2}$

= 2100 + 57.85 + 0.217 + 0.006 = 2158

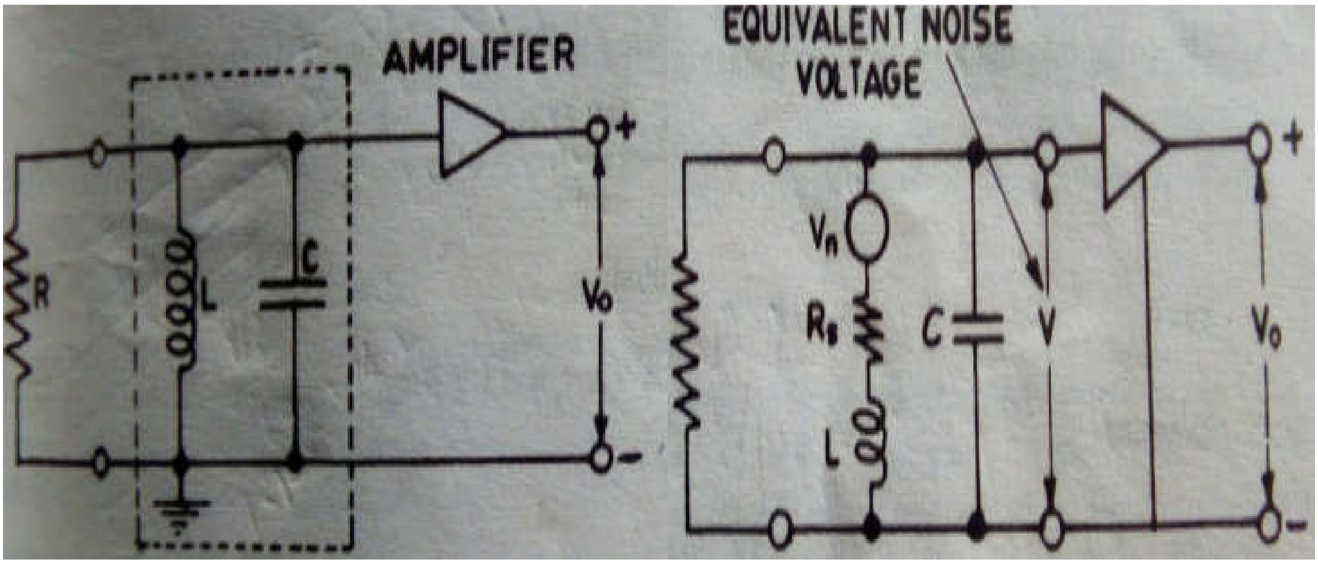
Hence, $V_{neq} = \sqrt{4\bar{K}TBR_{neq}} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 10^4 \times 2158}$volt

=0.598μV

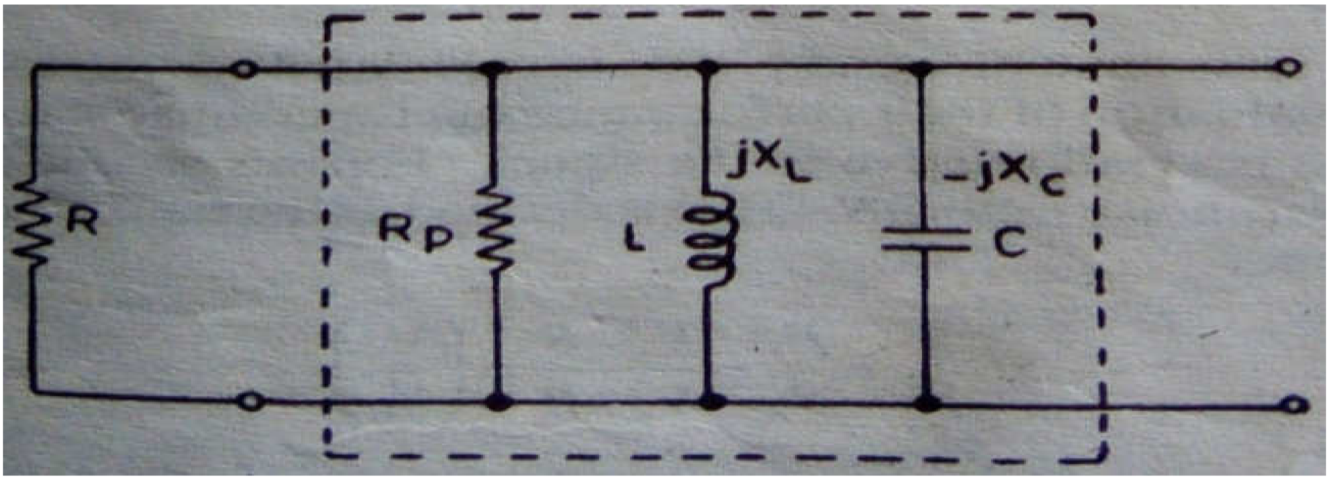
3.5.3. Noise in Reactive Circuits

Figure 2.3a) shows a reactive circuit in which a physical resistor R is placed in parallel with a parallel tuned circuit. The tuned is theoretically noiseless. Then, the presence of this tuned circuit does not affect the noise generated by the resistor$R$.

However, on either side of the pass band, the tuned circuit attenuates both the signal and the noise in the same fashion. Thus the tuned circuit limits the bandwidth of the noise source by disallowing noise outside its own pass band.



a) Amplifier circuit with ideal tuned circuit in the input
b) Amplifier circuit with non-ideal tuned circuit in the input



c) Noise Equivalent circuit

Figure 2.3. Noise in a tuned circuit.

Figure 2.3 b) shows a practical case when this tuned circuit is non-ideal, i.e. the inductor L of the tuned circuit possesses a small resistive element $R_S$as shown. This resistive element $R_S$generates noise.

Recall that in preceding sections, we have considered physical input resistor as the source of noise. It may, however, be clearly understood that the noise producing resistance need not necessarily be a physical resistor. Thus, in Figure. 2.2, if we use parallel tuned circuits having equivalent parallel resistances equal to R_1, $R_2$and R_3respectively, then the noise contribution remains the same. This is established below using the practical tuned circuit of Figure. 2.3 (b).Thus, the series resistance $R_s$of the coil forms the noise source generating a noise voltage V_n as shown in Figure. 2.3(b).
From Figure. 2.3, we are required to calculate the noise voltage across the capacitor *C*, i.e., across the input to the amplifier. From this, we may calculate the resistance which may be said to be generating noise. Now, the noise current in the *RLC* circuit is given by I_n
Where$I_n = \frac{V_n}{Z}$

$$Z = R_s + j\ (X_L - X_c) \qquad (2.20)$$

At resonance,$X_L = X_c$ so that, $I_n = \frac{V_n}{R_s}$.
Hence, the magnitude of the noise voltage appearing across the capacitor *C*is given by,

$$V = I_n\ .\ X_c = \frac{V_n . X_v}{R_s} = \frac{V_n Q R}{R_s} = QV_n$$

Since $X_c = Q . R_s$ at resonance, Then

$$V^2 = V_n^2 = Q^2 4\ \overline{k}TBR_p \text{ and } V = \sqrt{4\ \overline{k}TBR_p}$$

Where V = the noise voltage is across the tuned circuit due to the internal resistance Rs
Rp= Equivalent parallel impedance of the tuned circuit at resonance and equals $R_s Q^2$.

Equ. (2.20) shows that the equivalent parallel impedance of a tuned circuit equals its equivalent resistance for noise.

3.12. **Signal-to-Noise Ratio**

Usually, the equivalent noise resistance of an amplifier is calculated for two major reasons:

i. For comparing two circuits for evaluation of their performance
ii. For comparison of noise and signal at the same point to ensure that the noise is small relative to signal.

For this second purpose, it is advantageous to compute the so called signal- to noise ratio *S/N*. this is defined as the ratio of signal power to the noise power at the same point in the system.

Thus, $$\frac{S}{N} = \frac{P_s}{P_n} = \frac{V_s^2/R}{V_n^2/R} = \left(\frac{V_s}{V_n}\right)^2 \qquad (2.21)$$

In Equ. (2.21), it is assumed that both the signal and the noise are developed across the same resistor$R$. It is desirable to keep the S/N ratio as high as possible.

3.13. **Noise Figure**

This is a metric benchmark used to evaluate the performance of a system regarding noise at a glance. This factor is referred to as *Noise Figure F* of the system. By simple definition, the noise figure *F* is defined as the ratio of the signal-to-noise power supplied to the output terminals of the system (amplifier or receiver) to the signal-to-power supplied by the system to the output load impedance. Mathematically, this is given by Equ. (2.22)
Hence, Noise Figure,

$$F = \frac{\frac{S}{N} attheinput}{\frac{S}{N} attheoutput} \quad (2.22)$$

In electronic communication systems, most elements such as amplifier or receivers have different input impedances. This makes it practically impossible to compare their relative performance regarding noise from the knowledge of their equivalent noise resistance.

Assuming that a receiver have impedance of 60Ω and R_{eq}= 100Ω while another have input impedance of 400 Ω and R_{eq} =450 Ω.

Now, although the second receiver has higher $R_{eq,}$ but still, it is better regarding noise.
Any practical receiver generates some noise and the S/N will deteriorate as we move towards the output. Hence, in a practical receiver, the output S/N will be lower than the input S/N and hence the noise figure exceeds unity. In an ideal receiver, no additional noise is introduced by the various stages of the receiver so that the noise figure is unity. Hence, we may find noise figure in an alternative manner. Thus noise figure *F* equals the S/N of an ideal system divided by the S/N at the output of the receiver or system under test, both systems working at the same temperature, over the same bandwidth and fed from the same source. In addition, both the systems must be linear.
Noise figure *F* may be expressed either as a ratio or in decibels. Noise figure of a practical receiver should be as small as possible. It may be kept low by a suitable choice of the transistor, by proper circuit design and by use of low noise resistors. Thus, in a practical receiver operating at frequencies up to a few Giga-hertz, the noise-figure may be kept below a couple of decibels. Noise figure generally deteriorates with increase in operating frequency. However, even at frequencies exceeding a few giga-hertz, noise figure may still be kept below a couple of decibels (dbs) by use of devices which themselves use the transmit-time effect or are relatively independent of it.

3.7.1. Noise Figure Computation

The noise figure of an receiver or amplifier may be calculated by treating the entire system as a two port network having input impedanceR_t, output impedance $R_L$and overall voltage gain *A*. Let the system be supplied from a source (or antenna) of internal impedance $R_a$which may or may not be equal to R_r. Figure 2.4 gives a contextual the block diagram of circuit arrangement. We shall now look at the calculation of an amplifier (or receiver) noise in Figure 2.4.

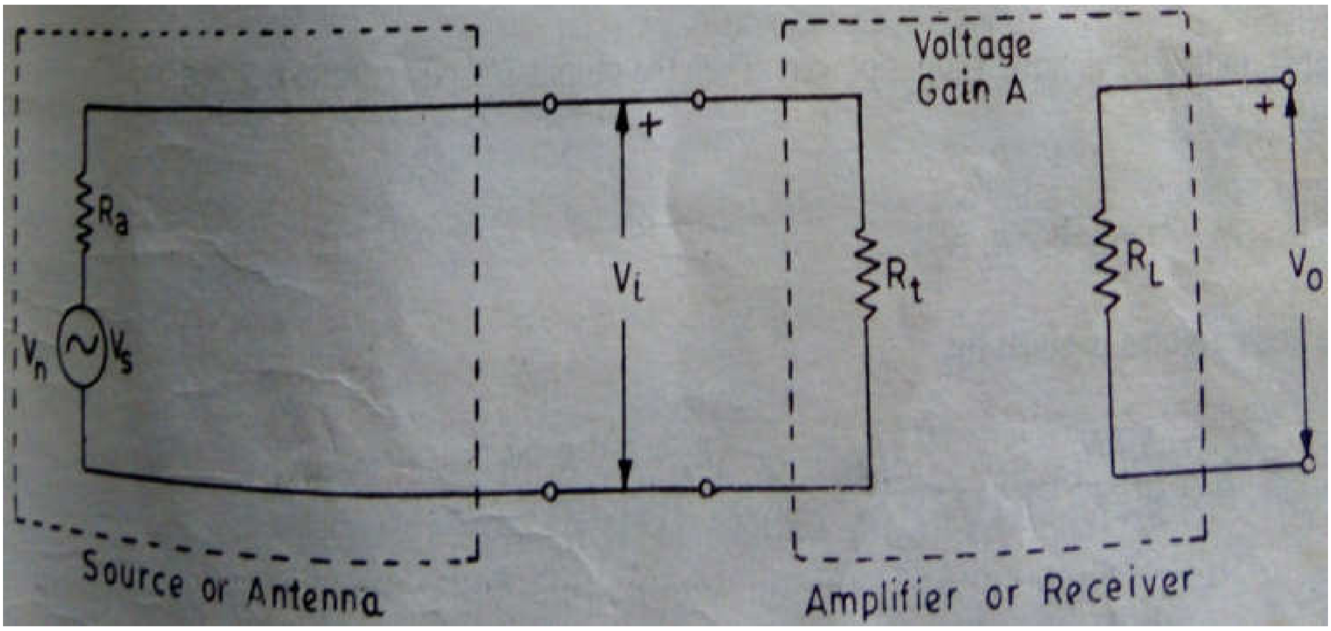


Figure.2.4. Noise Figure Block diagram.

Considering Figure 2.4, we want to calculate the *F*of an amplifier (or receiver). In this case, the following procedure has been enumerated to be adopted for calculating the noise figure of this system, viz:

i. Determine the input signal power S_i
ii. Determine the input noise power N_i
iii. Calculate the input signal- to- noise power ratio $S_i/_{\mathrm{i}}$
iv. Determine the output signal power S_o
v. Determine the output noise power N_o
vi. Calculate the output signal-to-noise power ratio$S_{o/No}$
vii. From steps (iii) and (vi) calculate the noise figure *F*.

From Figure. 2.4, it is worthy to derive that,

$$V_{si=} \frac{VR1}{Ra+R1} \tag{2.23}$$

Hence, $$S_i = \frac{V_{si}^2}{R_t} = \left(\frac{V_{s.R_t}}{R_{a+R_t}}\right)\frac{1}{R_t} = \frac{V_s^2\, R_t}{(R_{a+R_t})} \tag{2.24}$$

Input noise voltage is given by,

$$V^2_{ni=}\ 4KTB\frac{R_aR_t}{R_{a+}\,R_t} \tag{2.25}$$

Hence input noise power is,

$$N_{i=}\frac{V_{ni}^2}{R_t}=\frac{4KTBRa}{R_{a+}\,R_t} \tag{2.26}$$

The input signal power is given by,

$$\frac{S_i}{N_i} = \frac{V_s^2R_t}{(R_{a+}\,R_t}\frac{(R_{a+R_t)}}{4\bar{K}TBR_a} = \frac{V_s^2R_t}{4\bar{k}TBR_{a\,(R_a+R_t)}} \tag{2.27}$$

The output signal power is given by,

$$S_{\circ} = \frac{V_{S\circ}^2}{R_L}\frac{(AV_{si})}{R_L} = \frac{(AV_s R_t}{R_a + R_t}\frac{1}{R_L}$$

$$= \frac{A^2 V_s^2 R_t^2}{(R_a+R_t)^2 R_L} \qquad (2.28)$$

Let $N_{\circ}$ indicate the noise output power. Then the output $S/_N$ power ratio is given by,

$$\frac{S_{\circ}}{N_{\circ}} = \frac{A^2 V_s^2 R_t^2}{(R_a+R_t)^2 R_L N_o} \qquad (2.29)$$

Hence the noise figure is given by,

$$F = \frac{S_i/N_i}{S_{\circ}/N_{\circ}} = \frac{V_{s^2} R_t}{4\bar{k}TBR_a R_a + R_t} \cdot \frac{(R_a+R_t)^2 R_L N_o}{A^2 V_s^2 R_1^2}$$

$$= \frac{R_L N_O (R_a + R_t)}{4\bar{k}TBA^2\, R_a T_t} \qquad (2.30)$$

Equ. (2.30) gives an estimate theoretical results, the real formula for the noise figure *F* may be obtained by substituting for the output noise power or from the value of equivalent resistance or from measurement.

3.7.2. Determination of Noise Figure in Terms of Equivalent Noise Resistance

Recall from Equ. (2.15), the equivalent noise resistances of an amplifier or receiver equal the sum of i) input resistance of the first stage ii) the equivalent noise resistance of the first stage and iii) noise resistance of the subsequent stage referred to the first stage.

All these resistances get added up to constituteR_{eq}given a lumped resistance which corresponds to all the noise making of circuit. The rest of the circuit may now be assumed to be noiseless. With reference to Figure. 2.4, this implies that all these noise resistances get added up to the parallel combination of $R_a$and $R_{t.}$In order to correlate the noise figure and the equivalent noise resistance we define the term R_{eq}as the noise resistance not including R_t. Then,

$$Req = R_{eq} - R_t \qquad (2.31)$$

The total equivalent noise resistance of this receiver is then given by Equ. (2.32)

$$R = R_{eq+} \frac{R_a R_t}{R_a + R_t} \qquad (2.32)$$

The equivalent noise voltage effective at the input of the receiver is then given by,

$$V_{ni} = \sqrt{4\bar{k}TBR} \qquad (2.33)$$

With all the noise components included into the total equivalent noise resistance R, the amplifier could now be assumed as noiseless with voltage gain of A. In this regard, the output noise power is now given by Equ. 2.34

$$N_o = \frac{V_{no}^2}{R_L} = \frac{(AV_{n_i})^2}{R_L} = \frac{A^2 4\bar{K}TBR}{R_L} \qquad (2.34)$$

On substituting the value of N_o from Equ. (2.34) into Equ. (2.30), we now obtain

$$F = \frac{R_L(R_a+R_t)}{4kTBA^2 R_a R_t}\frac{A^2 4\bar{k}TBR}{R_L}$$

$$= R\frac{R_a+R_t}{R_a R_t} = \left(R'_{eq} + \frac{R_a R_t}{R_a+R_t}\right)\cdot\frac{R_a+R_t}{R_a R_t} = 1 + R_{eq}{'}\frac{R_a+R_t}{R\cdot R\cdot} \qquad (2.35)$$

It is pertinent to state that Equ. (2.35) gives the noise figure F in terms of the equivalent noise resistanceR_{eq}. Also, from Equ. (2.35), we conclude that for minimizing the noise figure for a given value of antenna resistance R_a, the ratio $(R_a + R)/R_a R_t$ must be minimum, i.e. R_t must be made much larger than R_a.

The inequality $R_a \neq R_t$ represent a condition of impedance mismatch, under which condition, the transfer of power from source (antenna) to the system (receiver) is not maximum. However, this mismatch is often used for reducing noise. Under extreme condition $R_t \gg R_{a,}$the ratio $(R_a + R_t)/R_t$ approaches unity and Equ. (2.35) for noise figure reduces to the following simple form:

$$F = 1+ \frac{R_{eq}}{R_a} \tag{2.36}$$

Example 7.

Compute the noise figure for the amplifier of Example 5 if the amplifier is driven by a generator whose output impedance is 40 Ω.

Solution.

$R_{\boldsymbol{eq=}} R_{\boldsymbol{eq}} - R_{\boldsymbol{t}}$ =2670 -500= 2170 Ω

F = 1+$R_{eq} \frac{R_a + R_t}{R_a x R_1} = 1 + 2170 * \frac{40+500}{40+500} = 59.6.$

Example 8.

The first stage of two stage samplifiers has output resistance of 25 kΩ, voltage gain of 12, input resistance of 600 kΩ, voltage gain of 25, input resistance of 100 kΩ and equivalent noise resistance of 8 kΩ. The amplifier is driven by a generator of output resistance 50 Ω. Compute for this two stage amplifier: (i) equivalent input noise resistance (ii) equivalent input noise voltage given that the bandwidth of the amplifier is 10 kHz and the ambient temperature is 330K and (iii) noise figure of the complete system.

Solution.

$$R_3 = 300 k\Omega$$

$$R_2 = \frac{100 \times 25}{100 + 25} \, k\Omega + 8 \, k\Omega = 28 k\Omega$$

$$R_1 = 600\Omega + 2400 \, \Omega = 3000\Omega$$

$$R_{eq} = R_1 + \frac{R_2}{A_1^2} \frac{R_3}{(A_1 A_2)} = 300 \, \Omega + \frac{28{,}000}{(12)} + \frac{300 \times 10^3}{(12 \times 25)} = 3198 \, \Omega$$

$$V_{neq} = \sqrt{4kTBR_{eq}} = \sqrt{4 \times 1.38 \times 10^{-23} \times 330 \times 10^4 \times 3198}\text{Volt}$$

$$= 0.727 uV$$

$$R_{eq} = R_{eq - R_t = 3198 - 600 = 2598 \, \Omega}$$

$$F = 1 + R_{eq} \times \frac{R_a + R_t}{R_a \cdot R_t} = 1 + 2598 \times \frac{50 + 600}{50 \times 600} = 49.5.$$

3.7.3. Measurement of Noise Figure

Basically, Noise figure may without loss of generality be calculated using of Equ. (2.35), once the equivalent noise resistance, R_a and R_t are known. But at very high frequencies where transit time becomes prevalent, using Equ. 2.35 become impossible. When this becomes the case, the best alternative is to use a measurement testbed to determine the noise figure.

An example of this type of testbed is a diode noise generator circuit model shown in Figure 2.5 depicting the circuit arrangement in such a case.

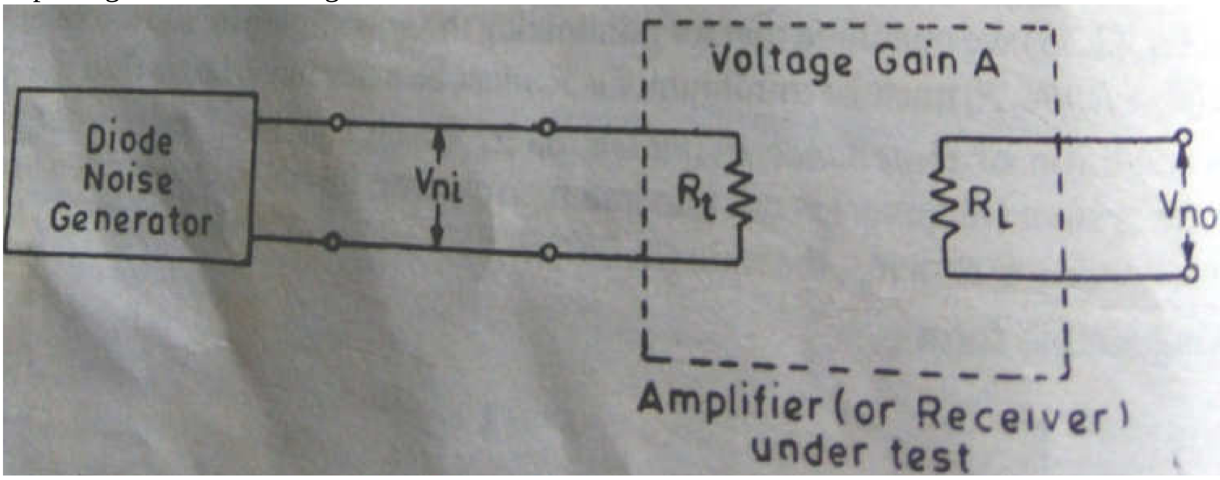


Figure. 2.5. Circuit arrangement for noise figure measurements.

From Figure 2.5, the diode noise generator makes use of shot noise generated by a diode operated under temperature limited condition. Equ. (2.5) gives the plate noise current of a vacuum diode. This plate noise current is controlled by varying the filament voltage by means of a potentiometer. The output capacitance of the diode and its associated circuit is made to resonate at the operating frequency of the receiver by using a variable inductance. The capacitance may be neglected and the output impedance of the noise generator is simply a resistorR_a. Then the noise voltage supplied by the noise generator to the input circuit of the receiver is given by,

$$V_n = I_n Z_n = I_n \frac{R_a R_t}{R_a+R_t} = \frac{R_a R_t\sqrt{2qI_pB}}{R_a+R_t} \quad (2.37)$$

The noise generator is connected to the receiver (or amplifier) under test and the noise output power of the receiver is measured with zero diode plate current, i.e. with noise diode plate supply switched off. Next the diode plate supply is switched on and the filament potentiometer is adjusted to allow the plate current to assume such as a value that the noise power developed across R_L becomes twice as large as the noise power in the absence of diode plate current. This value of diode plate current is measured by a millimeter. Under this condition, the additional noise power provided by the diode equals the normal noise power output of the receiver. This enables us to express the noise power output of the receiver in terms of the diode plate current.
Thus, we have

$$N_o = \frac{V_{no}^2}{R_L} = \frac{(A\,V_n)^2}{R_L} = \frac{A^2 R_a^2 R_t^2 2qI_pB}{R_L\,(R_a + R_t)^2} \quad (2.38)$$

Now the value of N_o as given by Equ. (2.38) may be substituted into Equ. (2.30) to get

$$F = \frac{R_L(R_a + R_t)N_o}{A^2\, 4\bar{k}TBR_aR_t} = \frac{R_L\,(R_a + R_t)}{A^2\, 4\bar{k}TBR_aR_t}\frac{A^2R_a^2R_t^2 2_qI_{pB}}{R_L\,(R_a + R_t)^2}$$

$$= \frac{q_p^I R_aR_t}{2kT\,(R_aR_t)} \quad (2.39)$$

As usual, from the consideration of minimizing noise, the system is kept mismatched. Thus, let $R_t \gg R_t$ so that Equ. (2.39) reduces to the following simple form:

$$F = \frac{q.R_a I_p}{2\bar{k}T} \qquad (2.40)$$

If the above described procedure is repeated from the very beginning for a matched system, it may be proved that Equ. (2.40) applies exactly to such a matched system instead of being a good approximation. Thus Equ. (2.40) is valid for both matched and mismatched conditions. This is the merit of this noise diode measurement.

Finally, from Equ (2.40), we can (i) substitute the value of q ($1.6 \times 10^{-19} C$) (ii) substitute the value of $\bar{k}$ (1.38×10^{-23} joules/deg K) and (iii) put T = 300 K. then Equ. (2.40) reduces to,

$$F = \frac{R_a I_a\,(1.6 \times 10^{-19})}{2 \times 1.38 \times 10^{-3} \times 300}\; 19.3\, R_a I_p \qquad (2.41)$$

Where R_a is in ohms and I_p is in amperes.

3.8. Noise Temperature

A very useful indicator of noise performance of a receiver system is the noise figure. However, when dealing with UHF and microwave low noise antennas, receivers or devices, noise figure is not always the most convenient measure of noise. Noise temperature is extensively used for antennas and low noise microwave amplifiers. A part from the ease of measurement, the concept of noise temperature permits addition similar to noise power. As such we can now rewrite Equ. (2.1) below while giving the total noise power from several sources,

$$P_{t=\bar{k}T_t B} \qquad (2.42)$$

This may be put as,

$$P_t = P_1 + P_2 = \bar{k}BT_1 + \bar{k}BT_2 \qquad (2.43)$$

$$\bar{k}BT_t = \bar{k}B(T_t + T_2) \qquad (2.44)$$

$$T_t = T_1 + T_2 \qquad (2.45)$$

Where p_1 and p_2 are the two individual noise power which may respectively be the noise powers received by the antenna and the power generated by the antenna.

T_1 and T_2 are the individual noise temperature corresponding to P_1 and P_2 and respectively and T_t is the total noise temperature.

The use of noise temperature has additional advantages for use at low noise level in that for any given noise level, it shows greater variation than does the noise figure.The term equivalent noise resistance has been defined earlier. This resistance is a fictions one having no physical presence. But the term is frequently used because of its convenience. In the same manner, the equivalent noise temperature of a receiver or amplifier may be used, if convenient. In defining the equivalent noise temperature T_{eq} of a receiver, it is assumed that $R_{eq} = R_{a.}$ if this assumption result in correct value of noise output power, then R_{eq} must be at a temperature other than that of all the components including $R_{a.}$ then we may use Equ.(2.36) to relate the noise figure F and equivalent temperature T_{eq} as below:

$$F = 1 + \frac{R_{eq}}{R_a} = 1 + \frac{kT_{eq}\,BR_{eq}}{kT_o\,BR_a} = 1 + \frac{T_{eq}}{T_a} \qquad (2.46)$$

Where $R_{eq} = R_a$ as assumed in the definition of T_{eq}

$T = 27^o\ C = 300^o\ K$

T_{eq}= equivalent noise temperature of the receiver or amplifier under consideration.

From Equ. (2.46). $T_{o^{F} = T_o} + T_{eq}$

Hence, $T_{eq} = T_o\ (F\ - 1)$ (2.47)
Knowing the noise figure F, Equ. (2.47) enables us to calculate the equivalent noise temperature $T_{eq.}$

Example 9.
A receiver having equivalent noise resistance of 2500 Ω and input resistance of 500 Ω is connected to an antenna of resistance 50 Ω. Compute the noise figure (in dBs) and equivalent noise temperature for the receiver.

Solution.

$$F = 1 + \frac{R_{eq}}{R_a}$$

$$R_{eq} = R_{eq} - R_t = 2500 - 500 = 2000\,\Omega$$

$$F = 1 + \frac{2000}{50} = 41$$

$$F\ in\ dB = 10\ log_{10}\ 41 = 16.12\ dB$$

$$T_{eq} = T_o(F - 1) = 300\ (41 - 1) = 1200^o\ K.$$

4.0 Conclusion

Reliable electrical communication systems depend on how accurate the receiver can determine and distinguish between wanted and unwanted signals. The presence of noise makes it possible difficult to have a perfect signal identification. Noise is will always be found in communication systems since there is a default thermal noise associated with conduction. If the signal strength is insufficient, adding more stages of amplification at the receiver is useless since noise will be distributed alongside with the amplified signal.

5.0 Summary

This unit has explained the concepts, types and sources of noise. A good understanding of the methods of calculating noise produced by various sources has been presented. We have discussed noise metrics such as signal-to-noise ratio, noise figure and noise temperature and have equally solved some simple problems involving resistance and temperature induced noise even in cascaded mode.

6.0 Tutor-Marked Assignment (TMA)

1. What is meant by the term “noise” as used in communication system?
2. Name the different sources of random noise and impulse noise external to a receiver. How can these be avoided or at least minimized? Which is the strongest extra- terrestrial noise?
3. Name the different noise which may be created within a receiver or amplifier. Discuss the effects of these noises on the performance of a receiver.
4. Give reasons for the following noise in a transistor (a) flicker noise (b) transistor thermal noise device, (c) partition noise.

5. State reasons why frequency mixer stage produces more noise than the amplifier using the same device.
6. In a multistage amplifier, derive expression for the net equivalent noise resistance in terms of resistance $R_{1,}R_{2,}R_{3,}$etc at the input of first, second, third etc. stage respectively and the voltage gains of the individual stages.
7. Show that in a parallel tuned circuit, the equivalent parallel impedance at resonance equals its equivalent resistance for noise generation.
8. Describe the method of measurement of noise figure of a receiver using a diode generator.
9. Defined the equivalent noise temperature of a receiver or amplifier. Under what condition is the quantity noise temperature more useful quantity than the noise figure? Give reasons.
10. Defined thermal noise. Given the mechanism of generation of thermal noise. Write expression for thermal noise power in a resistor at T deg K.

7.0 REFERENCES/FURTHER READING

Sturley, K.R.(1965);Radio Receiver Design; Part 1, 3rd ed., p.p. 193-198: Chapman & Hall Ltd London.

Blake, L.V: Antenna' p. 388, John Wiley & Sons, Inc., New York, 1966.

Mumford, W.W. and Elmer H. Scheibe (1966); Noise Performance Factors in Communication System,' Horizon House Microwave Inc., Dedham Mass, 1968.

George Kennedy: Electronic(1984); Communication System, 3rd ed., McGraw Hill Book Company,.

Stremler G.: Introduction (1977). Communication System,' Addison- Wesley Publishing Company, Reading Mass, p. 175.

Schwartz, M.: Information, Transmission, Modulation and Noise', 2nd ed., McGraw Hill Book Company, New High input impedance circuits; High frequency oscillators;

UNIT 3 SIGNALS ANALYSISAND COMPUTATION

1.0 Introduction
2.0 Objectives
3.0 Main Content
3.1. Signal Definition
3.2. Signal Classification
3.2.1. Continuous-Time and Discrete-time signals
3.2.2. Analog and Digital Signals
3.3.2.1. Analog signal
3.3.2.2. Digital signal
3.2.3. Periodic and aperiodic signals
3.4. Periodic Signal Determination and Fundamental frequency
3.2.4. Energy and power signals
3.2.5. Deterministic and Random signals
3.2.6. Odd and even signals
3.3. Analysis of Periodic Signal Theories
3.3.1. Period
3.3.2. Frequency
3.3.3. Radial Frequency
3.3.4. Amplitude
3.3.5. Continuous Signal
3.3.6. DC Offset
3.4. Signal processing
3.4.1. Analog signal processing (ASP)
3.4.2. Continuous-time signal processing (CTSP)
3.4.3. Discrete-time signal processing (DTSP)
3.4.4. Digital signal processing (DSP)
3.4.5. Non-linear signal processing (NSP)
3.5.Mathematical methods of Signal processing
3.5.1. Fourier analysis
3.5.2. Fourier Application in signal processing
3.5.3. Fourier analysis methods
3.5.3.1. Continuous-time Fourier series (CTFS)
3.5.3.2. Discrete-time Fourier series (DTFS)
3.5.3.3. Continuous-time Fourier transform (CTFT)
3.5.3.4. Discrete-time Fourier transform (DTFT)
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
8.0 References/Further Reading

1.0 INTRODUCTION

In electronic communication systems, the study of signals is very important.Signals are detectable quantities used to convey information about time-varying physical phenomena. Common examples of signals are human speech, temperature, pressure, and stock prices. Electrical signals, normally expressed in the form of voltage or current waveforms, are some of the easiest signals to generate and process. Mathematically, signals are modeled as functions of one or more independent variables. Examples of independent variables used to represent signals are time,Frequency, or spatial coordinates. Signals are applied in so many areas of electronic communication systems. This unit will lay the theoretical foundation for future dealings with electrical signals.

3.0 Objectives

After going through this unit, you should be able to:

- ✓ understand the concept of signals in communication systems
- ✓ differentiate between different types and classification of signals
- ✓ understand the different application of signals
- ✓ compare the different signal Fourier analysis
- ✓ Understand the importance of Fourier analysis
- ✓ know the formulation of the four Fourier analysis models in communication system.

3.0 Main Content

3.1. Signal Definition

In electronic communication systems, and signal processing, a signal is a function that "conveys information about the behavior or attributes of some phenomenon. In the physical world, any quantity exhibiting variation in time or variation in space (such as an image) is potentially a signal that might provide information on the status of a physical system, or convey a message between observers, among other possibilities. The term signal includes audio, video, speech, image, communication, geophysical, sonar, radar, medical and musical signals.

In science and human engineering, signals are typically provided by a sensor, and often the original form of a signal is converted to another form of energy using a transducer. For example, a microphone converts an acoustic signal to a voltage waveform, and a speaker does the reverse.

The formal study of the information content of signals is the field of information theory. The information in a signal is usually accompanied by noise as discussed in the previous chapter. Since noise is an undesirable random disturbance, or unwanted signals conflicting with the desired signal, prevention of noise deals with signal recovery which is an aspect of estimation theory, a probabilistic approach to suppressing random disturbances.

In the context of signal processing, arbitrary binary data streams are not considered as signals, but only analog and digital signals that are representations of analog physical quantities. Recall in our previous study in *communication systems*, a *transmitter* encodes a *message* to a signal, which is carried to a *receiver* by the communications *channel*. For instance, the words "Obi is a boy” might be the message spoken into a telephone. The telephone transmitter converts the sounds into an electrical voltage signal. The signal is transmitted to the receiving telephone by wires; at the receiver it is reconverted into sounds. In telephone networks, signaling, for example common-channel signaling, refers to phone number and other digital control information rather than the actual voice signal. This unit will focus on signal classifications, applications, analysis and mathematical formulations of signals including Fourier analysis.

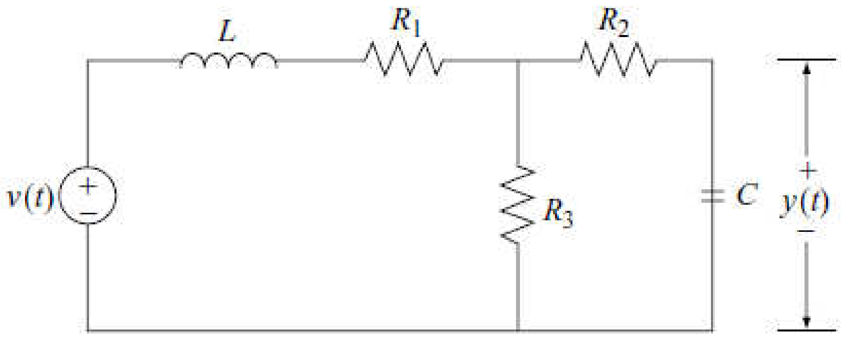
3.2. Signal Classification

A signal is classified into several categories depending upon the criteria used for its classification. In this section, we cover the following categories for signals:

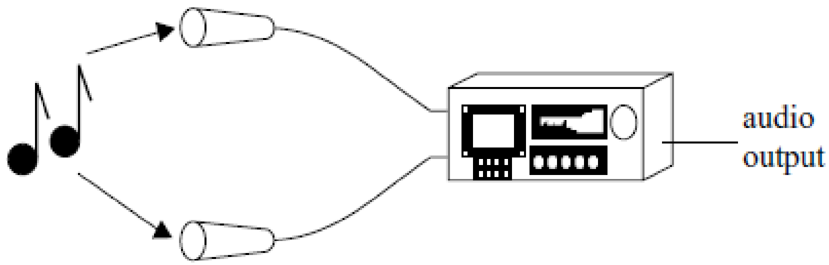
i. Continuous-time and discrete-time signals;
ii. Analog and digital signals;
iii. Periodic and a periodic (or non-periodic) signals;
iv. Energy and power signals;
v. Deterministic and probabilistic signals;
vi. Even and odd signals.

3.2.1. Continuous-Time and Discrete-time signals

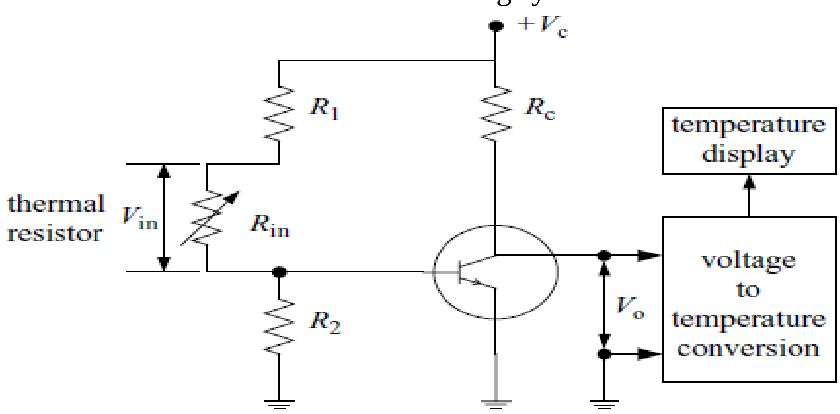
If a signal is defined for all values of the independent variable *t*, it is called a *continuous-time* (CT) signal. Consider Figure 1.1 as examples of signals and systems with important attributes below.



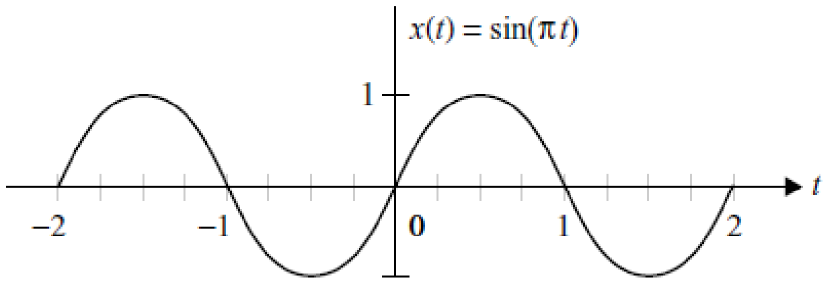
a. An electrical circuit



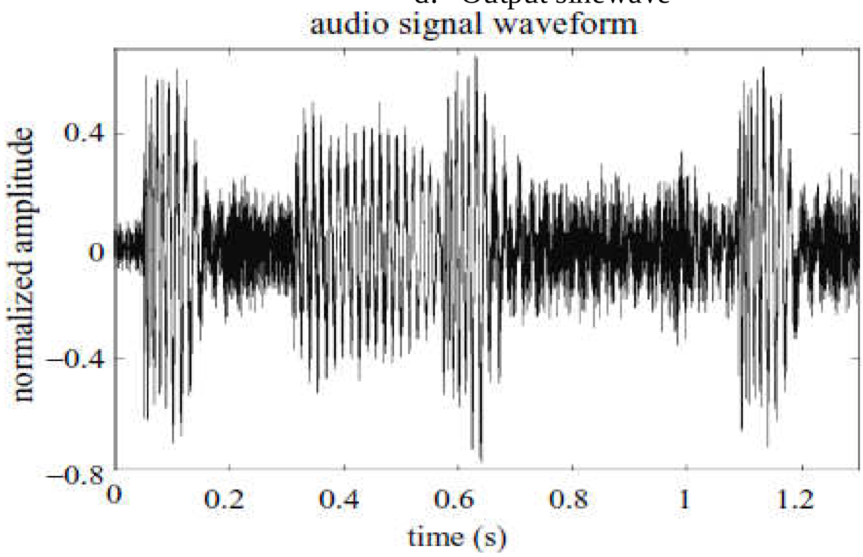
b. an audio recording system



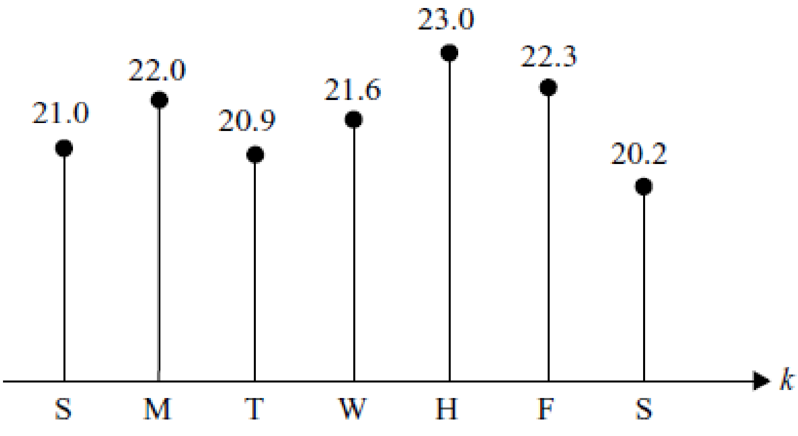
c. digital thermometer



d. Output sinewave



e. Audio Signal Waveform



f. Output Signal

Figure. 3.1. Examples of signals and systems. Plots (d),(e), and (f) are output signals generated, respectively, by the systems shown in (a), (b), and (c).

Consider the signals shown in Figures3.1(d) and (e). Since these signals vary continuously with time *t* and have known magnitudes for all time instants, they are classified as CT signals. On the other hand, if a signal is defined only at discrete values of time, it is called a *discretetime*

(DT)signal. Figure 3.1(f) shows the output temperature of a room measured at the same hour every day for one week. No information is available for the temperature in between the daily readings. Figure 3.1(f) is therefore an example of a DT signal. In our notation, a CT signal is denoted by $x(t)$with regular parenthesis, and a DT signal is denoted with square parenthesis as follows:

$x[kT]$, $k = 0,\pm1,\pm2,\pm3, \ldots,$

Where T denotes the time interval between two consecutive samples. In the example of Figure. 3.1(f), the value of T is one day. To keep the notation simple, we denote a one-dimensional (1D) DT signal x by $x[k]$. Though the sampling interval is not explicitly included in $x[k]$, it will be incorporated if and when required. Note that all DT signals are not functions of time.

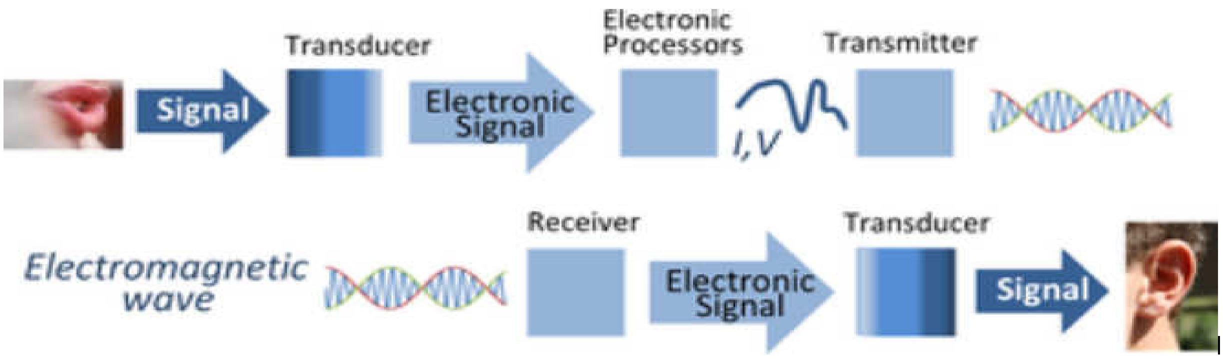


Figure 3.2. Signal Transmission using electronic signals

An analog signal, denoted *x(t),* is a continuous function of time and is uniquely determined for all t. When a physical signal such as speech is converted to an electrical signal by a microphone, we have an electrical analog of the physical waveform. An equivalent discrete-time signal, denoted as *x(kT),* exists only at discrete instants.

It is characterized by a sequence of values that exist at specific times, *kT*, where k is an integer and *T* is normally a fixed time interval. On the other hand, a continuous time signal may be restricted to a set of discrete amplitudes. A signal that is discrete in both time and amplitude is referred to as a digital signal. Furthermore, these discrete digital signal amplitudes can be represented by a set of numbers (codes) and, as such, can be stored in a computer memory. Pulse code modulation (PCM) is an example of a digital signal. These categorizations are illustrated in Figure 3.3.

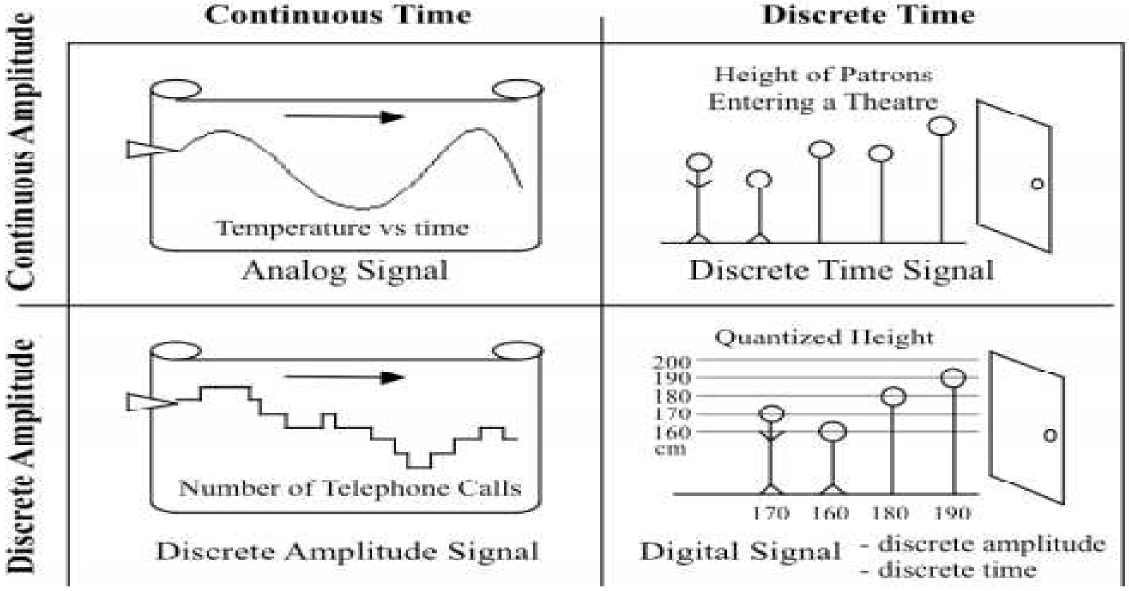


Figure 3.3. Concept of Analog and Digital Signals

Please note that signals moves with energy. But its energy waveform has finite amplitude and it either exists for a finite duration or it decays to zero over time. An energy signal has finite energy but zero average power. A power waveform has finite amplitude and semi-infinite duration, thus, it has finite average power but semi-infinite energy. The signal is considered to be continuous over the observation period. Energy and power classifications are mutually exclusive; a signal must be one or the other. Periodic signals are classified as power signals.

As shown in Figure 3.3, Signals can be categorized in various ways. The most common distinction is between discrete and continuous spaces that the functions are defined over, for example discrete and continuous time domains. Discrete-time signals are often referred to as time series in other fields. Continuous-time signals are often referred to as continuous signals even when the signal functions are not continuous; an example is a square-wave signal.
A second important distinction is between discrete-valued and continuous-valued. Particularly in digital signal processing a digital signal is sometimes defined as a sequence of discrete values that may or may not be derived from an underlying continuous-valued physical process.

In other contexts, digital signals are defined as the continuous-time waveform signals in a digital system, representing a bit-stream. In the first case, a signal that is generated by means of a digital modulation method is considered as converted to an analog signal, while it is considered as a digital signal in the second case. Another important property of a signal is its entropy or information content which has to do with statistical signal class definition. Signals in nature can be converted to electronic signals by various sensors as seen in Figure 3.2.

3.2.2. Analog and Digital Signals

A second classification of signals is based on their amplitudes. The amplitudes of many real-world signals, such as voltage, current, temperature, and pressure, change continuously, and these signals are called *analog* signals. For example, the ambient temperature of a house is an analog number that requires an infinite number of digits (e.g., 24.763578. . .) to record the readings precisely. Digital signals, on the other hand, can only have a finite number of amplitude values. For example, if a digital thermometer, with a resolution of 1˚C and a range of [10 ˚C, 30 ˚C], is used to measure the room temperature at discrete time instants, $t = kT$, then the recordings constitute a digital signal. An example ofa digital signal was shown in Figure3.1(f), which plots the temperature readings taken once a day for one week. This digital signal has an amplitude resolution of 0.1˚C, and a sampling interval of one day. Figure 3.4 shows an analog signal with its digital approximation. The waveform for the analog signal is shown with a line plot the quantized digital approximation is shown with a stem plot.

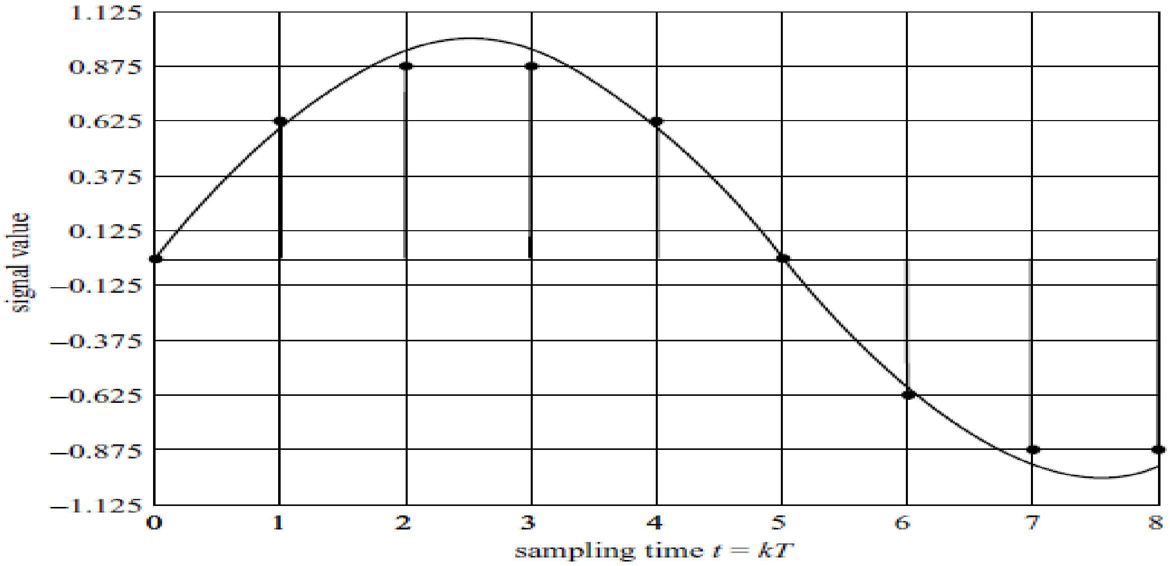


Figure 3.4. Analog signal with its digital approximation.

The analog signal has a limited dynamic range between [−1, 1] but can assume any real value (rational or irrational) within this dynamic range. If the analog signal is sampled at time instants $t = kT$ and the magnitude of the resulting samples are quantized to a set of finite number of known values within the range [−1, 1],the resulting signal becomes a digital signal. Using the following set of eight uniformly distributed values,[−0.875, −0.625, −0.375, −0.125, 0.125, 0.375, 0.625, 0.875],within the range [−1, 1], the best approximation of the analog signal is the digital signal shown with the stem plot in Figure 3.4. .As we can see From Figure 3.3, the analog and digital signals are the major two main types of signals encountered in practice. We shall further discuss both analog and digital signals as individual subsystem to place them in perspective.

3.3.2.1. Analog signal

An analog signal is any continuous signal for which the time varying feature (variable) of the signal is a representation of some other time varying quantity, i.e., *analogous* to another time varying signal. For example, in an analog audio signal, the instantaneous voltage of the signal varies continuously with the pressure of the sound waves. It differs from a digital signal, in which the continuous quantity is a representation of a sequence of discrete values which can only take on one of a finite number of values. The term analog signal usually refers to electrical signals; however, mechanical, pneumatic, hydraulic, human speech, and other systems may also convey or be considered analog signals.

An analog signal uses some property of the medium to convey the signal's information. For example, an aneroid barometer uses rotary position as the signal to convey pressure information. In an electrical signal, the voltage, current, or frequency of the signal may be varied to represent the information.

Any information may be conveyed by an analog signal; often such a signal is a measured response to changes in physical phenomena, such as sound, light, temperature, position, or pressure. The physical variable is converted to an analog signal by a transducer. For example, in sound recording, fluctuations in air pressure (that is to say, sound) strike the diaphragm of a microphone which induces corresponding fluctuations in the current produced by a coil in an electromagnetic microphone, or the voltage produced by a condenser microphone. The voltage or the current is said to be an "analog" of the sound.

3.3.2.2.Digital signal

We need to understand the meaning of a digital signal at this point. A digital signal is a signal that is constructed from a discrete set of waveforms of a physical quantity so as to represent a sequence of discrete values. A logic signal is a digital signal with only two possible values and describes an arbitrary bit stream. Other types of digital signals can represent three-valued logic or higher valued logics.

Conversely, a digital signal may be considered to be the sequence of codes represented by such a physical quantity. The physical quantity may be a variable electric current or voltage, the intensity, phase or polarization of an optical or other electromagnetic field, acoustic pressure, the magnetization of a magnetic storage media, etcetera. Digital signals are present in all digital electronics, notably computing equipment and data transmission.
With digital signals, system noise, provided it is not too great, will not affect system operation whereas noise always degrades the operation of analog signals to some degree. Digital signals often arise via sampling of analog signals, for example, a continually fluctuating voltage on a line that can be digitized by an analog-to-digital converter circuit, wherein the circuit will read the voltage level on the line, say, every 50 microseconds and represent each reading with a fixed number of bits. The resulting stream of numbers is stored as digital data on a discrete-time and quantized-amplitude signal. Computers and other digital devices are restricted to discrete time. We shall introduce you to various forms of signal occurrences and their representations in graphical formats.

3.2.3. Periodic and aperiodic signals

A CT signal $x(t)$ is said to be *periodic* if it satisfies the following property:

$$x(t) = x(t + T_0), \qquad (3.1)$$

at all time t and for some positive constant T_0. The smallest positive value of T_0 that satisfies the periodicity condition, Equ. (3.2), is referred to as the *fundamental period* of $x(t)$.
Likewise, a DT signal $x[k]$ is said to be *periodic* if it satisfies $x[k] = x[k + K_0]$ (3.2)
at all-time$k$ and for some positive constant K_0.

The smallest positive value of K_0 that satisfies the periodicity condition, Equ. (3.3), is referred to as the fundamental period of x[k]. A signal that is not periodic is called an aperiodic or non-periodic signal. Figure 3.5 shows examples of both periodic and aperiodic signals. The reciprocal of the fundamental period of a signal is called the fundamental frequency.
Mathematically, the fundamental frequency is expressed as follows

$f_0 = \frac{1}{T_0}$, for CT signals, or$f_0 = \frac{1}{k_0}$, for DT signals, (3.3),

Where T_0 and K_0 are, respectively, the fundamental periods of the CT and DT signals. The frequency of a signal provides useful information regarding how fast the signal changes its amplitude. The unit of frequency is cycles per second(c/s) or hertz (Hz). Sometimes, we also use radians per second as a unit off requency. Since there are 2π radians (or 360^0) in one cycle, a frequency of $f_0$hertz is equivalent to $2\pi f_0$ radians per second. If radians per second are used as a unit of frequency, the frequency is referred to as the angular frequency and is given by Equ.(3.4).

$\omega_0 = \frac{2\pi}{T_0}$for CT signals, or $\Omega_0 = \frac{2\pi}{K_0}$ for DT signals, (3.4),

A familiar example of a periodic signal is a sinusoidal function represented mathematically by the following expression:

$x(t) = A\sin(\omega_0 t + \theta)$. (3.5),

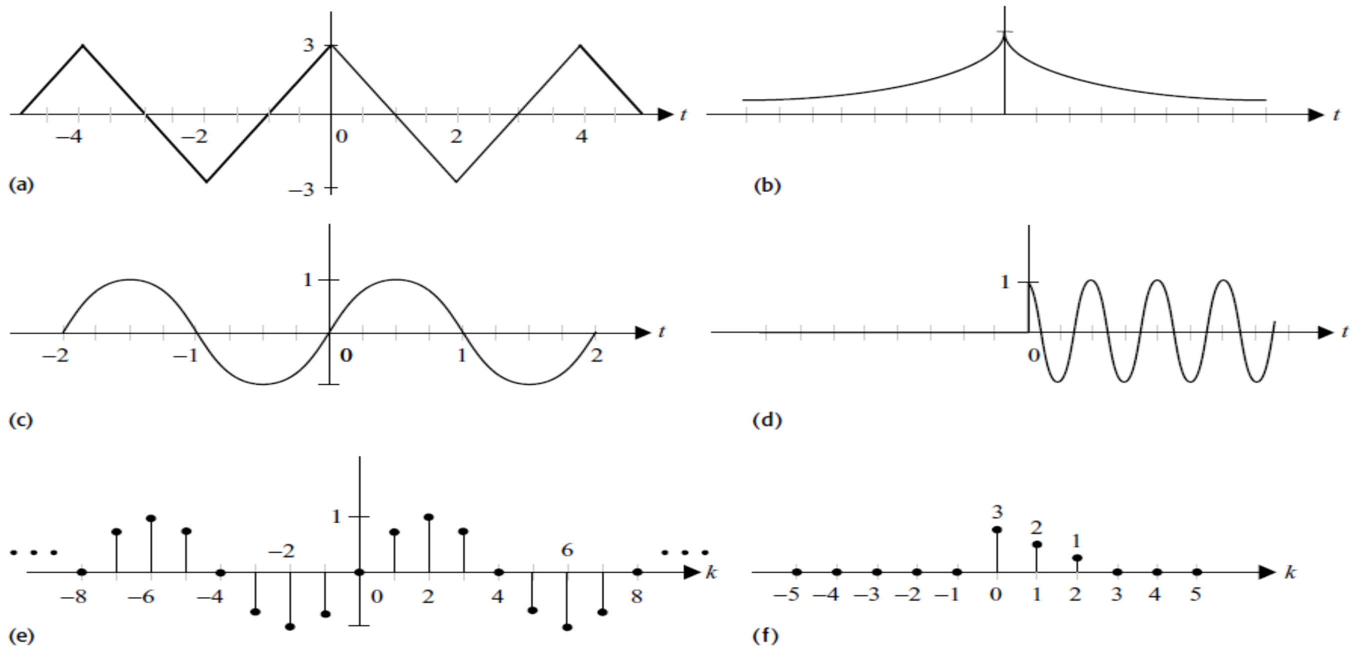


Figure 3.5.Examples of periodic((a), (c), and (e)) and aperiodic((b), (d), and (f)) signals. The line plots (a) and (c) represent CT periodic signals with fundamental periods T_0 of 1.5and 3, while the stem plot (e)represents a DT periodic signal with fundamental period K_0= 10.

3.4. Periodic Signal Determination and Fundamental frequency

As we can see in Figure 3.6, periodic signals are a bit like time loop movies. We can allow a periodic signal to do anything it wants for a limited interval of time (i.e. the period). The length of the period will be referred to as T_0 seconds. When the periodic signal reaches the end of its period, it must immediately jump back to the voltage it had at the start of the period, and trace out the same path. This pattern repeats every T_0 seconds. This is called a Sisyphean signal. A periodic signal can be observed by making a graph of the voltage, and looking for this repetitive pattern. The signal below is periodic, with a period of T_0 = 3 seconds.

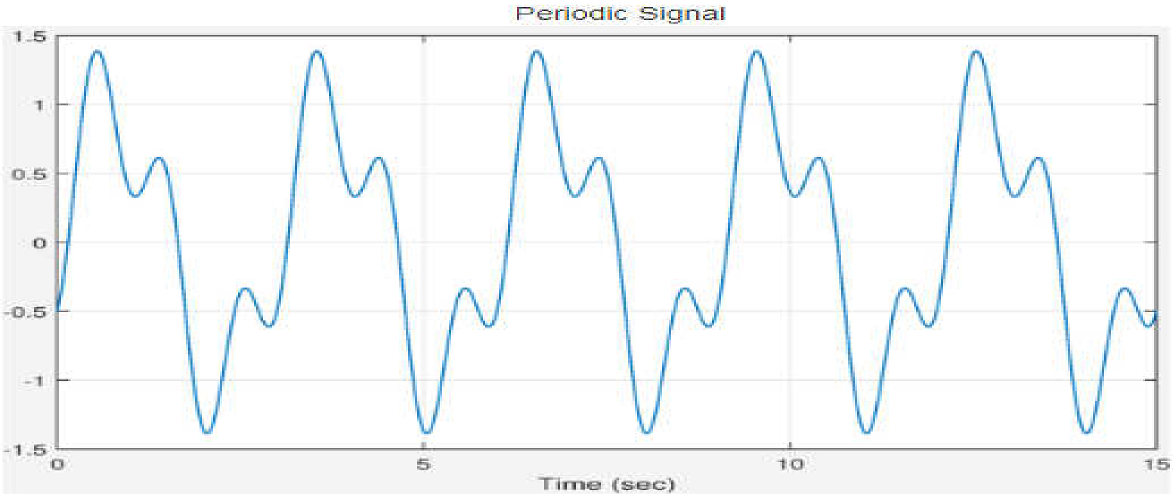


Figure 3.6. Periodic Signal waveform

Sometimes it is difficult to spot the period from a graph. At first glance, the signal in Figure 3.7 appears to have a period of 3 seconds again. But if you lay a straight edge across the peaks, you will find the one at 3.5 seconds is a little bit lower than the peaks at 0.5 and 6.5 seconds. This means this signal doesn't have a period of 3 seconds, it has a period of 6 seconds. That's one of the problems with just looking at graphs. You may miss a small variation in the signal, which has a big impact on the period.

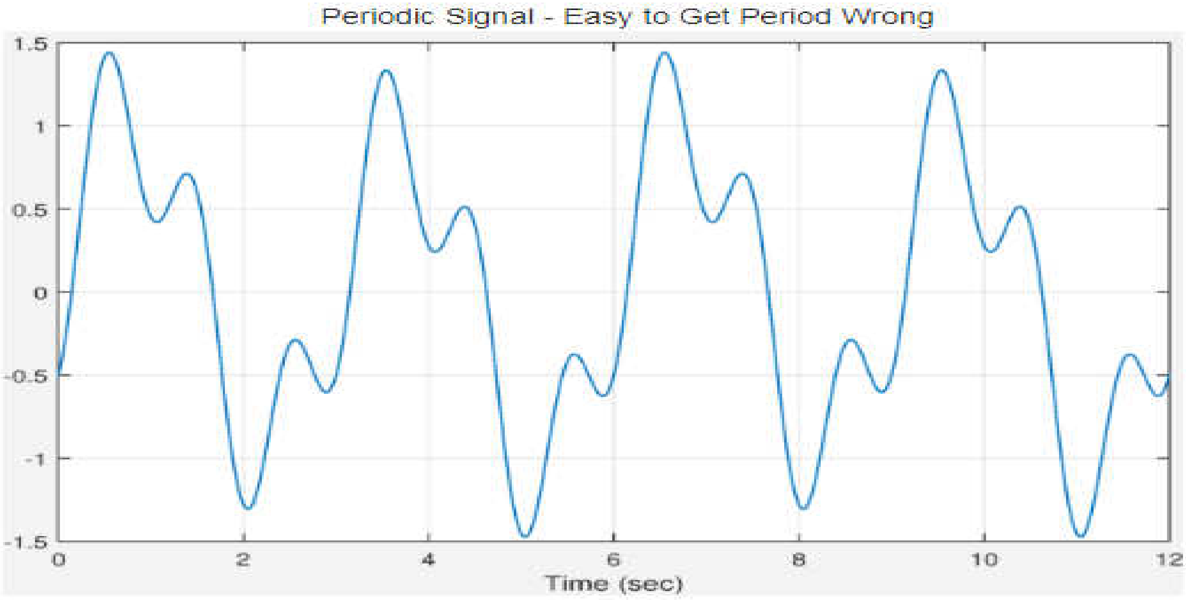


Figure 3.7. Periodic Signal waveform variations at 6 Seconds.

Complex periodic signal is most challenging to work it. Sometimes, it is most impossible to figure out the period from a graph. The signal in Figure 3.8 has a period of 1 second. If you look at the pattern of unusually high peaks, you might guess this was the period. However, the rapid variations in the signal makes it look like a blur most of the time, and it's impossible to tell from the graph if the blur from 0 to 1 second exactly matches the blur from 1 to 2 seconds.

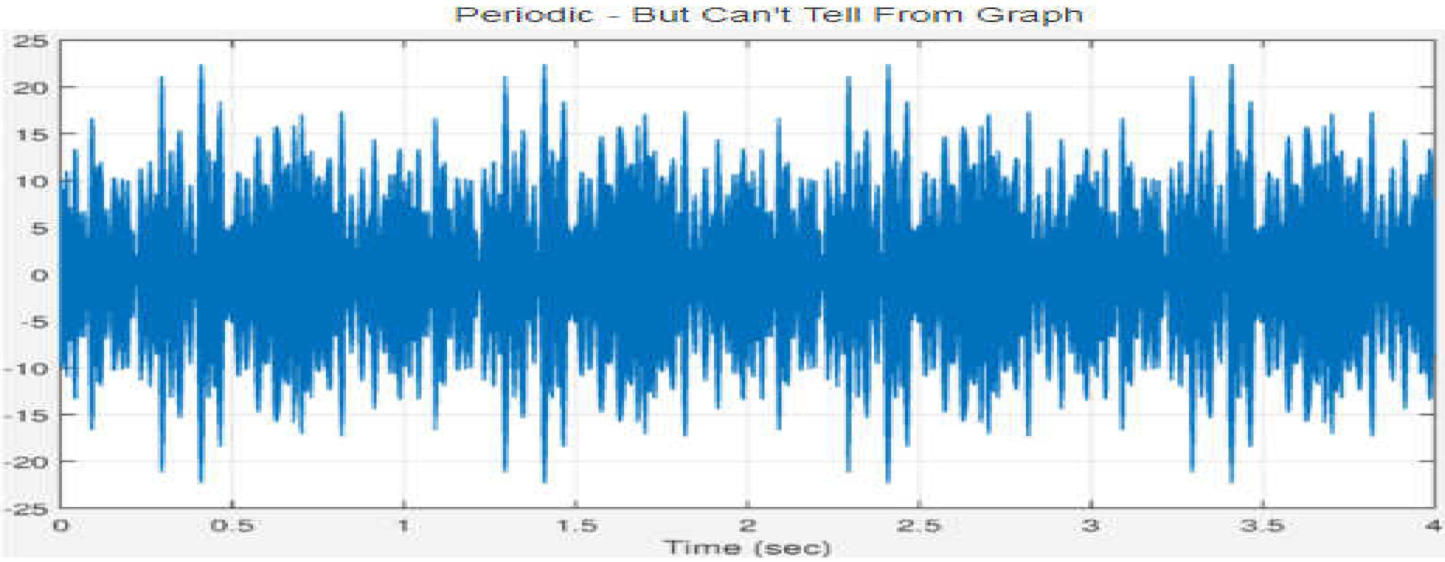


Figure 3.8. Complex periodic signal.

In some case, signals are clearly not periodic, or aperiodic. For instance, polynomials are aperiodic, like the one shown in Figure 3.9.

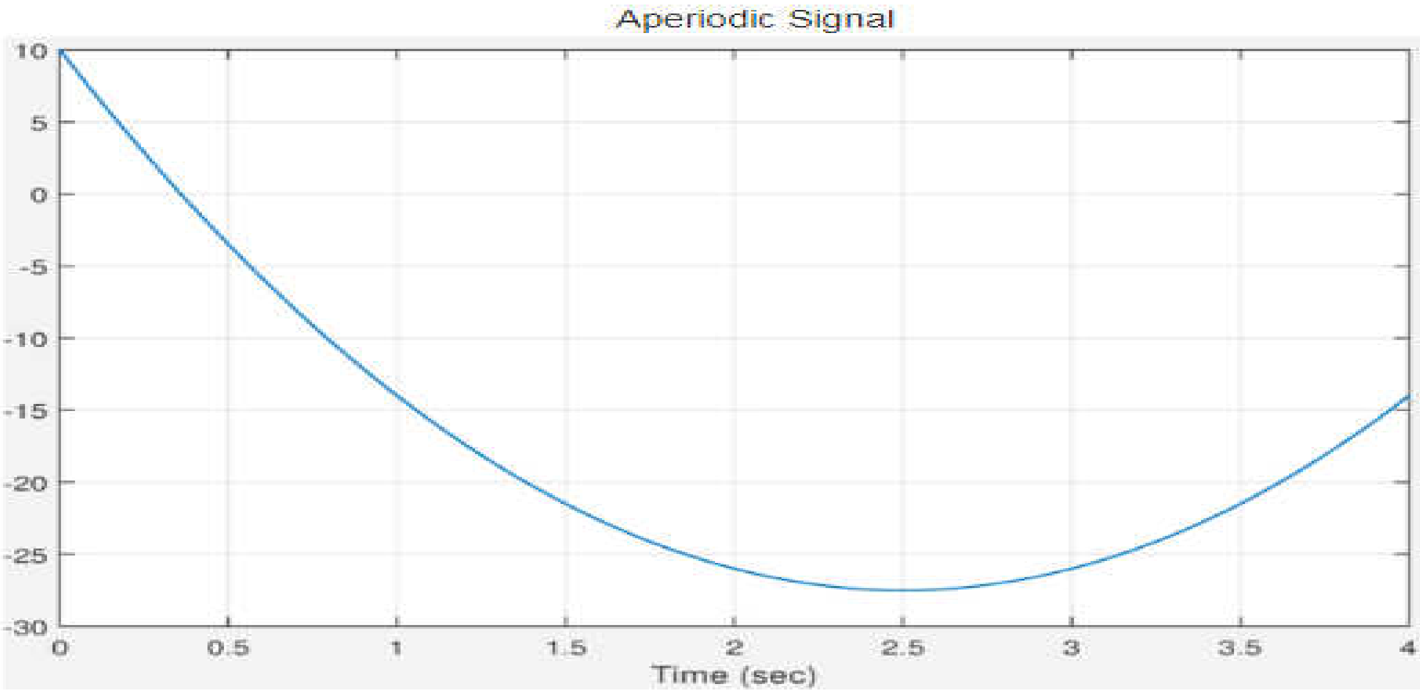


Figure 3.9. Aperiodic Signal representation.

Finally, there are signals that at first glance look periodic, but when you try to carefully compare voltages and times, you find out they are aperiodic. For a signal to be periodic, each period must be identical to every other period. The signal shown in Figure 3.10 ripples back and forth in very similar ways, but it never exactly repeats a previous time interval, so it is aperiodic

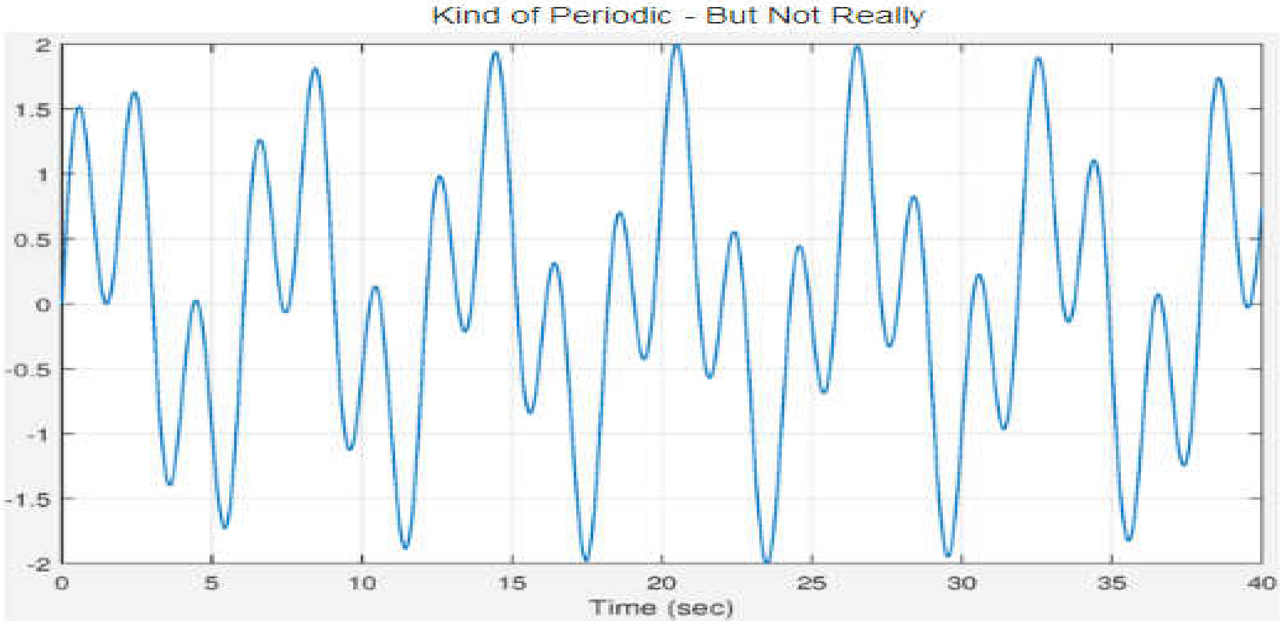


Figure 3.10. Aperiodic Oscillatory Signal.

3.2.4. Energy and Power Signals

Before presenting the conditions for classifying a signal as energy or a power signal, let us now present the formulas for calculating the energy and power in a signal. The *instantaneous power* at time $t = t_0$ of a real-valued CT signal $x(t)$ is given by $x^2(t_0)$. Similarly, the instantaneous power of a real-valued DT signal $x[k]$ at time instant $k = k_0$ is given by $x^2[k]$. If the signal is complex-valued, the expressions for the instantaneous power are modified to $|x(t_0)|^2$ or $|x[k_0]|^2$,where the symbol$|\cdot|$ represents the absolute value of a complex number.
The *energy* present in a CT or DT signal within a given time interval is given by the following: CT signals in Equ. (3.6)

$$E_{(T_1,T_2)} = \int_{T_1}^{T_2} |x(t)|^2 dt \text{ in interval } t = (T_1, T_2) \text{ with} T_2 > T_1); \qquad (3.6)$$

DT sequences in to

$$E_{(N_1,N_2)} = \sum_{K=N_1,}^{N_2} |x[k]|^2 \text{ in interval } K = [N_1, N_2] \text{ with } N_1 > N_2 \qquad (3.7)$$

The *total energy* of a CT signal is its energy calculated over the interval $t = [-\infty,\infty]$. Likewise, the total energy of aDT signal is its energy calculated over the range $k = [-\infty,\infty]$. The expressions for the total energy are therefore given by the following:
CT signals

$E_x = \int_{-\infty}^{\infty} |x(t)|^2 dt$;

DT sequences

$E_x = \sum_{-\infty}^{\infty} |x[k]|^2 \; dt$;

Since power is defined as energy per unit time, the *average power* of a CT signal $x(t)$ over the interval $t = (-\infty,\infty)$ and of a DT signal $x[k]$ over the range $k = [-\infty,\infty]$ are expressed as follows:
CT signals

$$P_x = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} |x[k]|^2 dt \qquad (3.8)$$

DT sequences

$$P_x = \frac{1}{2K+1} \sum_{K=-K}^{K} |x[k]|^2 (3.9)$$

Equations (3.8) and (3.9) are simplified considerably for periodic signals.
Since a periodic signal repeats itself, the average power is calculated from one period of the signal as follows:

CT signals

$$P_x = \frac{1}{T_0} \int_{T_0}^{t} |x[t]|^2 \; dt = \frac{1}{T_0} \int_{t_1}^{t_1+T_0} |x[t]|^2 \; dt \; (3.10)$$

DT sequences

$$P_x = \frac{1}{K_0} \sum_{k=\langle K_0 \rangle} |x[k]|^2 = \frac{1}{K_0} \sum_{k=k_1}^{k_1+K_0-1} |x[k]|^2 \qquad (3.11)$$

Where t_1 is an arbitrary real number and $k1$ is an arbitrary integer. The symbolsT_0 and K_0 are, respectively, the fundamental periods of the CT signal $x(t)$ and the DT signal $x[k]$. In Equ. (3.10), the duration of integration is one complete period over the range $[t_1, t_1+ T_0]$, where $t1$ can take any arbitrary value. In other words, the lower limit of integration can have any value provided that the upper limit is one fundamental period apart from the lower limit. To illustrate this mathematically, we introduce the notation $\int(T_0)$to imply that the integration is performed over a complete period T_0 and is independent of the lower limit.
Likewise, while computing the average power of a DT signal $x[k]$, the upper and lower limits of the summation in Equ. (3.11) can take any values as long as the duration of summation equals one fundamental period K_0.
A signal $x(t)$, or $x[k]$, is called an *energy signal* if the total energy Ex has a non-zero finite value, i.e. $0 < E_x < \infty$. On the other hand, a signal is called a *power signal* if it has non-zero finite power,

i.e. $0 < P_x < \infty$. Note that a signal cannot be both energy and a power signal simultaneously. The energy signals have zero average power whereas the power signals have infinite total energy. Some signals, however, can be classified as neither power signals nor as energy signals.
For example, the signal $e^{2t}u(t)$ is a growing exponential whose average power cannot be calculated. Such signals are generally of little interest to us. Most periodic signals are typically power signals.

Example 3.1
Consider the CT signals shown in (a) and (b) below. Calculate the instantaneous power, average power, and energy present in the two signals. Classify these signals as power or energy signals.
Solution

(a) The signal $x(t)$ can be expressed as follows:

$$x(t) = \begin{cases} 5 & -2 \le t \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

The instantaneous power, average power, and energy of the signal are calculated as follows:
Instantaneous power

$$P_x(t) = \begin{cases} 25 & -2 \le t \le 2 \\ 0 & \text{otherwise;} \end{cases}$$

Energy

$$E_x = \int_{-\infty}^{\infty} |x(t)|^2 \mathrm{d}t = \int_{-2}^{2} 25 \, \mathrm{d}t = 100;$$

Average power

$$P_x = \lim_{T \to \infty} \frac{1}{T} E_x = 0.$$

Because $x(t)$ has finite energy ($0 < Ex = 100 < \infty$) it is an energy signal.

(b) The signal $z(t)$ is a periodic signal with fundamental period 8 and is expressed as follows:

$$z(t) = \begin{cases} 5 & -2 \le t \le 2 \\ 0 & 2 < |t| \le 4, \end{cases}$$

With$z(t + 8) = z(t)$. The instantaneous power, average power, and energy of the signal are calculated as follows:

Instantaneous power

$$P_z(t) = \begin{cases} 25 & -2 \le t \le 2 \\ 0 & 2 < |t| \le 4 \end{cases}$$

and $Pz\ (t + 8) = Pz\ (t)$;

Average power

$$P_z = \frac{1}{8}\int_{-4}^{4} |z(t)|^2 \mathrm{d}t = \frac{1}{8}\int_{-2}^{2} 25 \,\mathrm{d}t = \frac{100}{8} = 12.5;$$

Energy

$$E_z = \int_{-\infty}^{\infty} |z(t)|^2 \mathrm{d}t = \infty.$$

Example 3.2

Consider the following DT sequence:

$$f[k] = \begin{cases} \mathrm{e}^{-0.5k} & k \geq 0 \\ 0 & k < 0. \end{cases}$$

Determine if the signal is a power or an energy signal.

Solution

The average energy present in the DT sequence is given by

$$E_f = \sum_{k=-\infty}^{\infty} |f[k]|^2 = \sum_{k=0}^{\infty} |\mathrm{e}^{-0.5k}|^2 = \sum_{k=0}^{\infty} (\mathrm{e}^{-1})^k = \frac{1}{1-\mathrm{e}^{-1}} \approx 1.582.$$

Because $E\,f$ is finite, the DT sequence $f\,[k]$ is an energy signal.

Example 3.3

Determine if the DT sequence $g[k] = 3\cos(\pi k/10)$ is a power or an energy signal.

Solution

The DT sequence $g[k] = 3\cos(\pi k/10)$ is a periodic signal with a fundamental period of 20. All periodic signals are power signals. Hence, the DT sequence $g[k]$ is a power signal.
Using Equ. (1.15), the average power of $g[k]$ is given by

$$P_g = \frac{1}{20}\sum_{k=0}^{19} 9\cos^2\left(\frac{\pi k}{10}\right) = \frac{9}{20}\sum_{k=0}^{19} \frac{1}{2}\left[1 + \cos\left(\frac{2\pi k}{10}\right)\right]$$
$$= \underbrace{\frac{9}{40}\sum_{k=0}^{19} 1}_{\text{term I}} + \underbrace{\frac{9}{40}\sum_{k=0}^{19} \cos\left(\frac{2\pi k}{10}\right)}_{\text{term II}}.$$

Clearly, the summation represented by term I equals 9(20)/40 = 4.5. To compute the summation in term II, we express the cosine as follows:

$$\text{term II} = \frac{9}{40}\sum_{k=0}^{19}\frac{1}{2}[\mathrm{e}^{\mathrm{j}\pi k/5} + \mathrm{e}^{-\mathrm{j}\pi k/5}] = \frac{9}{80}\sum_{k=0}^{19}(\mathrm{e}^{\mathrm{j}\pi/5})^k + \frac{9}{80}\sum_{k=0}^{19}(\mathrm{e}^{-\mathrm{j}\pi/5})^k.$$

Using the formulas for the geometric progression (GP) series yields

$$\sum_{k=0}^{19}(\mathrm{e}^{\mathrm{j}\pi/5})^k = \frac{1-(\mathrm{e}^{\mathrm{j}\pi/5})^{20}}{1-(\mathrm{e}^{\mathrm{j}\pi/5})} = \frac{1-\mathrm{e}^{\mathrm{j}\pi 4}}{1-(\mathrm{e}^{\mathrm{j}\pi/5})} = \frac{1-1}{1-(\mathrm{e}^{\mathrm{j}\pi/5})} = 0$$

and

$$\sum_{k=0}^{19}(\mathrm{e}^{-\mathrm{j}\pi/5})^k = \frac{1-(\mathrm{e}^{-\mathrm{j}\pi/5})^{20}}{1-(\mathrm{e}^{\mathrm{j}\pi/5})} = \frac{1-\mathrm{e}^{-\mathrm{j}\pi 4}}{1-(\mathrm{e}^{\mathrm{j}\pi/5})} = \frac{1-1}{1-(\mathrm{e}^{\mathrm{j}\pi/5})} = 0.$$

Term II, therefore, equals zero.

The average power of $g[k]$ is therefore given by

$Pg = 4.5 + 0 = 4.5.$

3.2.5. Deterministic and Random signals

Let us look at these two types of signal representation here. If the value of a signal can be predicted for all time (t or k) in advance without any error, it is referred to as a *deterministic signal*. Conversely, signals whose values cannot be predicted with complete accuracy for all time are known as *random signals*. Deterministic signals can generally be expressed in a mathematical, or graphical, form. Some examples of deterministic signals are as follows.

1. CT sinusoidal signal: $x_1(t) = 5\sin(20\pi t + 6)$;
2. CT exponentially decaying sinusoidal signal: $x_2(t) = 2e{-t}\sin(7t)$;
3. CTfinite duration complex exponential signal: $x_3(t) = = f(x) = \begin{cases} e^{j4\pi t}, |\ t| < 5 \\ 0, \quad elsewhere; \end{cases}$
4. DT real-valued exponential sequence: $x4[k] = 4e{-2k}$;
5. DT exponentially decaying sinusoidal sequence: $x5[k] = 3e{-2k}\times Sin(\frac{16\pi k}{5})$

Unlike deterministic signals, random signals cannot be modeled precisely. Random signals are generally characterized by statistical measures such as means, standard deviations, and mean squared values. In electrical engineering, most meaningful information-bearing signals are random signals. In a digital communication system, for example, data is generally transmitted using a sequence of zeros and ones. The binary signal is corrupted with interference from other channels and additive noise from the transmission media, resulting in a received signal that is random in nature. Another example of a random signal in electrical engineering is the thermal noise generated by a resistor. The intensity of the thermal noise depends on the movement of billions of electrons and cannot be predicted accurately.

The study of random signals is beyond the scope of this book. We therefore restrict our discussion to deterministic signals. However, most principles and techniques that we develop are generalizable to random signals. The readers are advised to consult more advanced books for analysis of random signals.

Consequently, a signal can be random or deterministic. We establish that a random signal is one in which there is some degree of uncertainty before it occurs. A non-random or deterministic signal is one in which there is uncertainty in its values. Signals may be classified as predictable or as unpredictable, as analog or discrete and of finite or infinite duration. Deterministic signals are defined exactly as a function of time. They can be periodic (such as a sine wave or square wave) or they can be an aperiodic "one shot" signal as explained in Section3.2.3.

Deterministic signals contain no information because their future is completely predictable by the receiver. They are easy to model and are useful since they can provide a reasonably accurate evaluation of communication system performance. Stochastic signals, on the other hand, are unpredictable and thus can communicate information. Although the time waveform of a stochastic signal is random, the signal power may be predictable. Examples of stochastic signals are thermal noise in electronic circuits, (i.e. "background hiss") and information signals such as voice or music as depicted in Figure 3.2.

3.2.6. Odd and even signals

Periodic functions or signals can be classified according to their symmetry. A function may be Odd, Even, or Neither Even nor Odd. All periodic functions can be classified in this way. Let us look at various signal appearances that could be observed in Figure 3.11.

i. EvenSignal

A CT signal $x_e(t)$ is said to be an even signal if

$x_e(t) = xe(-t)$.

In order words, functions are even if they are symmetrical about the y-axis.

$$f(x) = f(-x) \tag{3.12}$$

For instance, a cosine function is an even function.

ii. Odd Signal

A CT signal $x_o(t)$ is said to be an odd signal if

$x_o(t) = -x_o(-t)$.

As such, a function is odd if it is inversely symmetrical about the y-axis.

$f(x) = -f(-x)$

The Sine function is an odd function.

A CT signal $x_e[k]$ is said to be an even signal if

$$x_e[k] = x_e[-k]. \tag{3.13}$$

Conversely, a CT signal $x_o[k]$ is said to be an odd signal if

$$x_o[k] = -x_o[-k]. \tag{3.14}$$

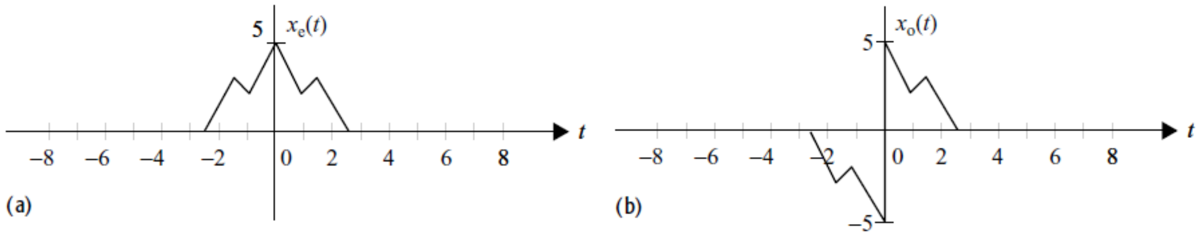


Figure 3.11. Example of (a) an even signal and (b) an odd signal.

Most practical signals are neither odd nor even as they do not exhibit any symmetry about the *y*-axis. We shall look at this category below

iii. Neither Even nor Odd Signal

Some functions are neither even nor odd. However, such functions can be written as a sum of even and odd functions. Neither odd nor even signals can be expressed as a sum of even and odd signals as follows:

$$x(t) = x_e(t) + x_o(t), \qquad (3.15)$$

Where the even component $x_e(t)$ is given by

$$x_e(t) = 1/2[x(t) + x(-t)], \qquad (3.16)$$

While the odd component $x_o(t)$ is given by

$$x_o(t) = 1/2[x(t) - x(-t)]. \qquad (3.18)$$

Example 3.4

Express the CT signal below as a combination of an even signal and an odd signal.

$$x(t) = \begin{cases} t & 0 \le t < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Solution

In order to calculate $x_e(t)$ and $x_o(t)$, we need to calculate the function $x(-t)$,which is expressed as follows:

$$x(-t) = \begin{cases} -t & 0 \le -t < 1 \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} -t & -1 < t \le 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Using Equ. (1.20), the even component $x_e(t)$ of $x(t)$ is given by

$$x_e(t) = \frac{1}{2}[x(t) + x(-t)] = \begin{cases} \frac{1}{2}t & 0 \le t < 1 \\ -\frac{1}{2}t & -1 \le t < 0 \\ 0 & \text{elsewhere} \end{cases}$$

while the odd component $x_o(t)$ is evaluated from Equ. (3.18) as follows:

$$x_o(t) = \frac{1}{2}[x(t) - x(-t)] = \begin{cases} \frac{1}{2}t & 0 \le t < 1 \\ \frac{1}{2}t & -1 \le t < 0 \\ 0 & \text{elsewhere.} \end{cases}$$

The waveforms for the CT signal $x(t)$ and its even and odd components are plotted in Figure. 3.11.

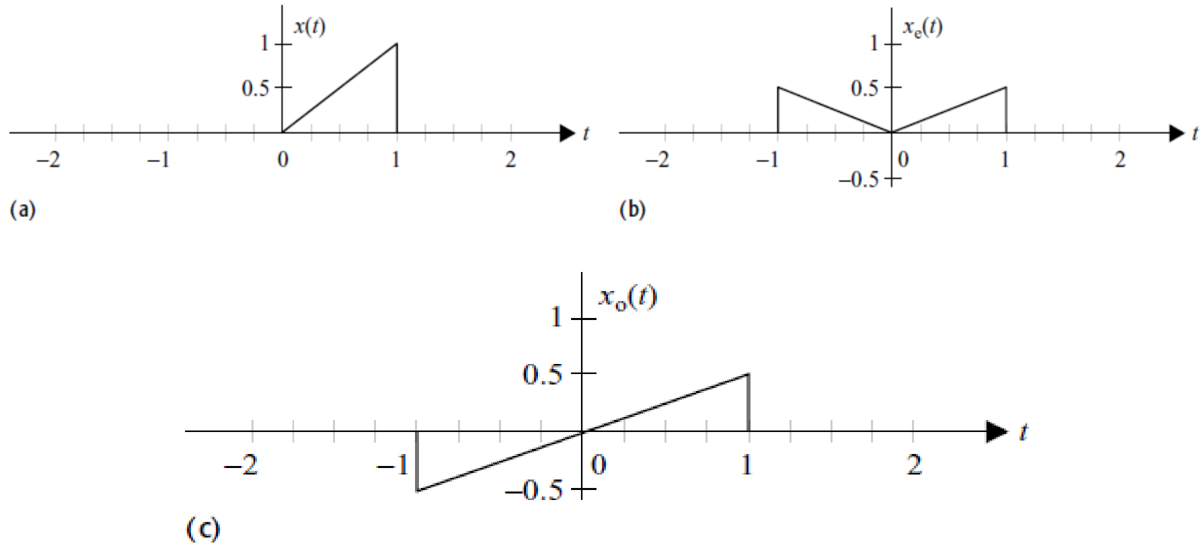


Figure 3.12(a) The CT signal *x* (*t*) for Even component of *x* (*t*) and Odd component of *x* (*t*).

So far, we have established that any function *f(x)* can be expressed as a sum of an odd function and an even function:

iv. Half-wave Symmetry

Another form of signal representation is shown in Figure 3.13. To determine if a signal with period 2*L* has half-wave symmetry, we need to examine a single period of the signal. If, when shifted by half the period, the signal is found to be the negative of the original signal, then the signal has half-wave symmetry. That is, the following property is satisfied:

$f(t - L) = -f(t)$

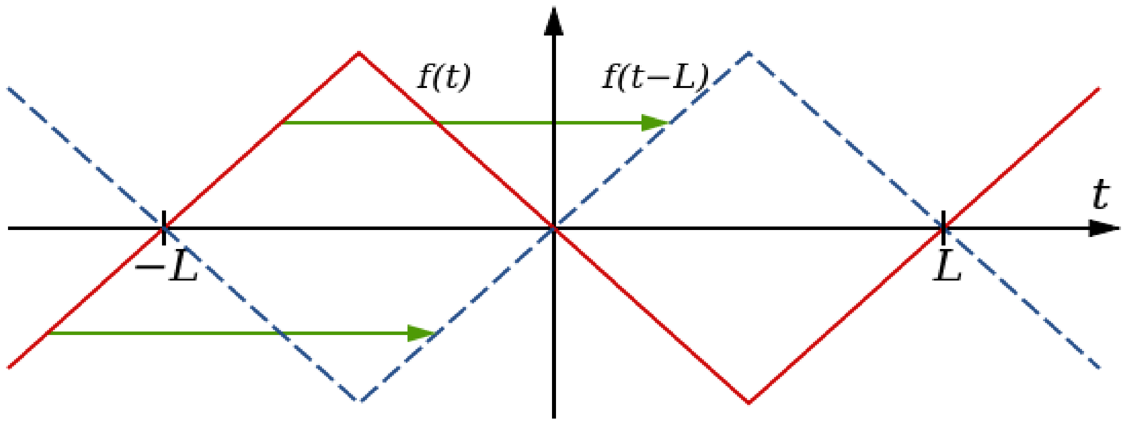


Figure 3.13. Half-Wave Symmetry representation.

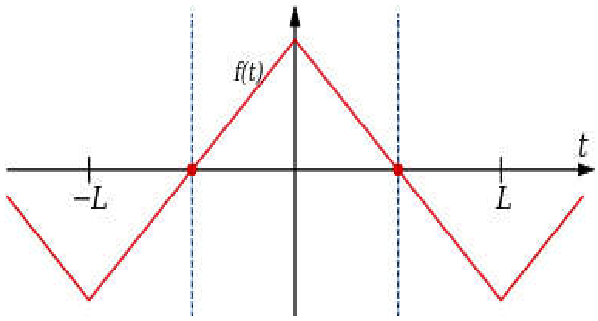
As shown in Figure 3.13, the half-wave symmetry implies that the second half of the wave is exactly opposite to the first half. A function with half-wave symmetry does not have to be even or odd, as this property requires only that the shifted signal is opposite, and this can occur for any temporal offset. However, it does require that the DC offset is zero, as one half must exactly cancel out the other. If the whole signal has a DC offset, this cannot occur, as when one half is added to the other, the offsets will add, not cancel.

Note that if a signal is symmetric about the half-period point, it is *not* necessarily half-wave symmetric. An example of this is the function t^3, periodic on (-1, 1), which has no DC offset and odd symmetry about t=0. However, when shifted by 1, the signal is *not* opposite to the original signal.Half wave symmetric signals doesn't have even "sine and cosine" harmonics

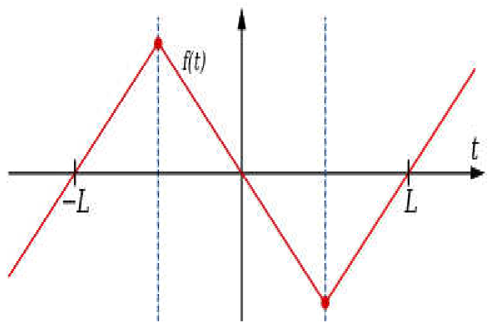
v. Quarter-Wave Symmetry

If a signal has the following properties, it is said to quarter-wave symmetric as shown in Figure 3.13:

- It is half-wave symmetric.
- It has symmetry (odd or even) about the quarter-period point (i.e. at a distance of L/2 from an end or the center).



a. Even Signal with Quarter-Wave Symmetry



b. Odd Signal with Quarter-Wave

Figure 3.13. Quarter-Wave Symmetry.

Any quarter-wave symmetric signal can be made even or odd by shifting it up or down the time axis. A signal does not have to be odd or even to be quarter-wave symmetric, but in order to find the quarter-period point, the signal will need to be shifted up or down to make it so. Figure 3.14 is an example of a quarter-wave symmetric signal (red) that does not show this property without first being shifted along the time axis (green, dashed)

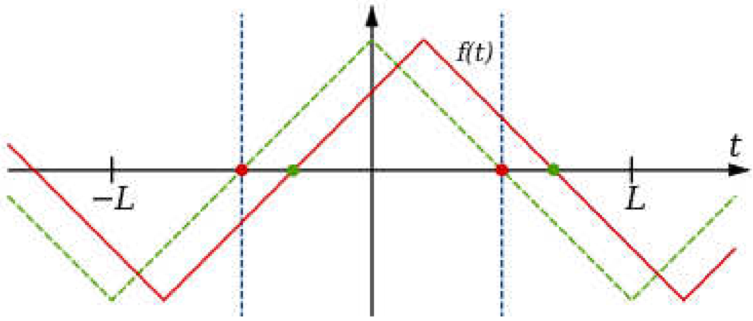


Figure 3.14. Asymmetric Signal with Quarter-Wave Symmetry

An equivalent operation is shifting the interval the function is defined in. This may be easier to reconcile with the formulae for Fourier series. In this case, the function would be redefined to be periodic on $(-L+\Delta, L+\Delta)$, where Δ is the shift distance.

vi. Discontinuities

In a graphical sense, a periodic signal as shown above has discontinuities whenever there is a vertical line connecting two adjacent values of the signal. In a more mathematical sense, a periodic signal has discontinuities anywhere that the function has an undefined (or an infinite) derivative. These are also places where the function does not have a limit, because the values of the limit from both directions are not equal.There are some common periodic signals that have their specific names. We will list those signals here, and discuss them below

i. Sinusoidal wave

The quintessential periodic waveform. These can be either Sine functions, or Cosine function.

$V_p\, Sin(2\pi f t)$.

ii. Square Wave

The square wave is exactly what it sounds like: a series of rectangular pulses spaced equidistant from each other, each with the same amplitude.

iii. Triangle Wave

The triangle wave is also exactly what it sounds like: a series of triangles. These triangles may touch each other, or there may be some space in between each wavelength.Example: Sinusoid, Square, and Triangle Waves

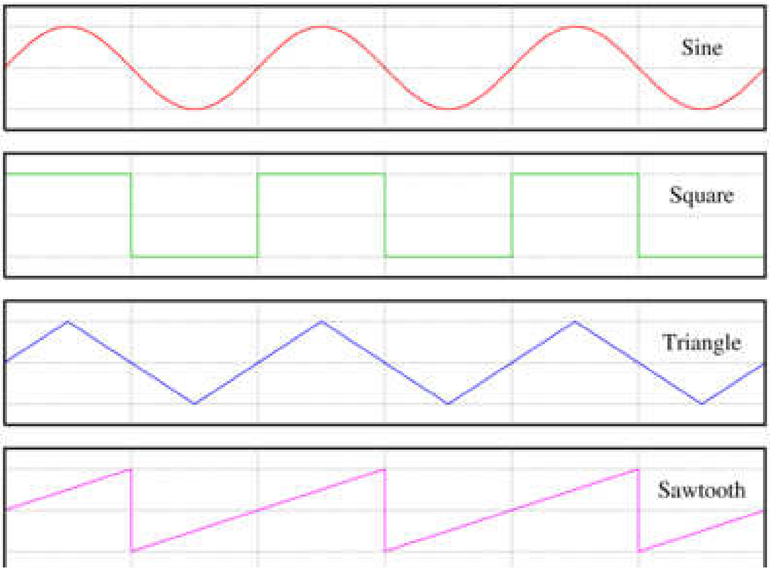


Figure 3.15. Common periodic Signalrepresentations

vii. Time discretization

We must know that one of the fundamental distinctions between different types of signals is between continuous and discrete time. In the mathematical abstraction, the domain of a continuous-time (CT) signal is the set of real numbers (or some interval thereof), whereas the domain of a discrete-time (DT) signal is the set of integers (or some interval). What these integers represent depends on the nature of the signal; most often it is time. As shown in Figure 3.16, if for a signal, the quantities are defined only on a discrete set of times, we call it a discrete-time signal.

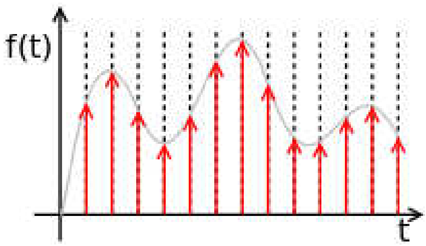


Figure 3.16. Discrete time Signal from a continuous signal by Sampling

A simple source for a discrete time signal is the sampling of a continuous signal, approximating the signal by a sequence of its values at particular time instants. A discrete-time real (or complex) signal can be seen as a function from (a subset of) the set of integers (the index labeling time instants) to the set of real (or complex) numbers (the function values at those instants).A continuous-time real (or complex) signal is any real-valued (or complex-valued) function which is defined at every time *t* in an interval, most commonly an infinite interval.

viii. Amplitude quantization

If a signal is to be represented as a sequence of numbers, it is impossible to maintain exact precision such that each number in the sequence must have a finite number of digits. As a result, the values of such a signal belong to a finite set; in other words, it is quantized. Quantization is the process of converting a continuous analog audio signal to a digital signal with discrete numerical values. See Figure 3.17.

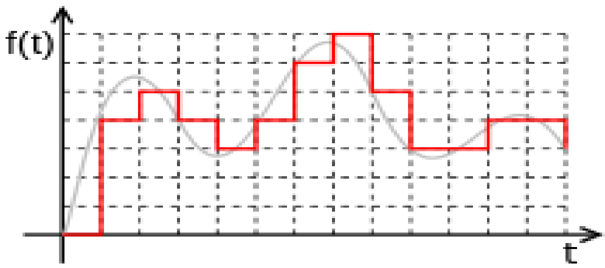


Figure 3.17. Digital signal from Approximation to Analogue signal.

Let us now look at some practical examples of signals.

- The motion of an object that can be monitored by various sensors to provide electrical signals. For example, radar can provide an electromagnetic signal for following aircraft motion. Orientation signals can be generated using a gyroscope.
- Sound vibration of a medium (such as air). A sound signal is converted to an electrical signal by a microphone, generating a voltage signal as an analog of the sound signal, making the sound signal available for further signal processing.
- Images of picture orimage which consists of a brightness or color signal as a function of a two-dimensional location. The object's appearance is presented as an emitted or reflected electromagnetic wave, one form of electronic signal. It can be converted to voltage or current waveforms using devices such as the charge-coupled device. A 2D image can have a continuous spatial domain, as in a traditional photograph or painting; or the image can be discretized in space, as in a raster scanned digital image. Color images are typically represented as a combination of images in three primary colors, so that the signal is vector-valued with dimension three.
- Videos. A video signal is a sequence of images. A point in a video is identified by its two-dimensional position and by the time at which it occurs, so a video signal has a three-dimensional domain. Analog video has one continuous domain dimension (across a scan line) and two discrete dimensions (frame and line).
- Biological membrane potentials. The value of the signal is an electric potential i.e. Voltage. However, this area is very complex in terms of analysis. Actually, some cells or organelles have the same membrane potential throughout; neurons generally have different potentials at different points. These signals have very low energies, but are enough to make nervous systems work; they can be measured in aggregate by the techniques of electrophysiology. Other examples of signals are the output of a thermocouple, which conveys temperature information, and the output of a pH meter which conveys acidity information.

3.3. Analysis of Periodic Signal Theories

A signal is a periodic signal if it completes a pattern within a measurable time frame, called a period and repeats that pattern over identical subsequent periods. The completion of a full pattern is called a cycle. A period is defined as the amount of time (expressed in seconds) required to complete one full cycle. The duration of a period represented by T, may be different for each signal but it is constant for any given periodic signal.

We will discuss here some of the common terminology that pertains to a periodic function. Let $g(t)$ be a periodic function satisfying $g(t + T) = g(t)$ for all t.

3.3.1. Period

The period is the smallest value of T satisfying $g(t + T) = g(t)$ for all t. The period is defined so because if $g(t + T) = g(t)$ for all t, it can be verified that $g(t + T') = g(t)$ for all t where $T' = 2T, 3T, 4T, \ldots$ In essence, it's the smallest amount of time it takes for the function to repeat itself. If the period of a function is finite, the function is called "periodic". Functions that never repeat themselves have an infinite period, and are known as "aperiodic functions".

The period of a periodic waveform will be denoted with a capital T. The period is measured in seconds.

3.3.2. Frequency

The frequency of a periodic function is the number of complete cycles that can occur per second. Frequency is denoted with a lower-case f. It is defined in terms of the period, as follows:

$$F = 1/T \qquad (3.19)$$

Frequency has units of *hertz* or cycle per second.

3.3.3. Radial Frequency

The radial frequency is the frequency in terms of radians. It is defined as follows:

$$\omega = 2\pi f \qquad (3.20)$$

3.3.4. Amplitude

The amplitude of a given wave is the value of the wave at that point. Amplitude is also known as the "Magnitude" of the wave at that particular point. There is no particular variable that is used with amplitude. The letters in capital A,M and $R$are most commonly used.The amplitude can be measured in different units, depending on the signal we are studying. In an electric signal the amplitude will typically be measured in volts. In a building or other such structure, the amplitude of a vibration could be measured in meters.

3.3.5. Continuous Signal

A continuous signal is a "smooth" signal, where the signal is defined over a certain range. For example, a sine function is a continuous signal, as is an exponential function or a constant function. A portion of a sine signal over a range of time 0 to 6 seconds is also continuous. Examples of functions that are not continuous would be any discrete signal, where the value of the signal is only defined at certain intervals.

3.3.6. DC Offset

A *DC Offset* is an amount by which the average value of the periodic function is not centered around the *x*-axis.A periodic signal has a DC offset component if it is not centered about the *x*-axis. In general, the DC value is the amount that must be subtracted from the signal to center it on the *x*-axis. Now, by definition:

$$A_0 = (1/T) * \int_{-T/2}^{T/2} f(x)dx \tag{3.21}$$

With A_0 being the DC offset. If $A_0 = 0$, the function is centered and has no offset.

3.4. Signal processing

In this subsection, it pertinent to note that signal processing deals with the analysis, synthesis, and modification of signals. These are broadly defined as functions passing information about the behavioral attributes of signals such as sound, images, and biological measurements. For instance, signal processing techniques are used to improve signal transmission fidelity, storage efficiency, and subjective quality, and to emphasize or detect components of interest in a measured signal.The various types of signal processing are primarily highlighted below.

3.4.1. Analog signal processing (ASP)

This ASP is for signals that have not been digitized, as in traditional radio, telephone, radar, and television systems. This involves linear electronic circuits as well as non-linear ones. The former are, for instance, passive filters, active filters, additive mixers, integrators and delay lines. Non-linear circuits include compandors, multiplicators (frequency mixers and voltage-controlled amplifiers), voltage-controlled filters, voltage-controlled oscillators and phase-locked loops.

3.4.2. Continuous-time signal processing (CTSP)

CTSP is essentially for signals that vary with the change of continuous domain(without considering some individual interrupted points).The methods of signal processing include time domain, frequency domain, and complex frequency domain. This technology mainly discusses the modeling of linear time-invariant continuous system, integral of the system's zero-state response, setting up system function and the continuous time filtering of deterministic signals.

3.4.3. Discrete-time signal processing (DTSP)

DTSP is meant for sampled signals, defined only at discrete points in time, and as such are quantized in time, but not in magnitude. In this regard, an analog discrete-time signal processing is a technology based on electronic devices such as sample and hold circuits, analog time-division multiplexers, analog delay lines and analog feedback shift registers. This technology was a predecessor of digital signal processing (see below), and is still used in advanced processing of gigahertz signals.The concept of discrete-time signal processing also refers to a theoretical discipline that establishes a mathematical basis for digital signal processing, without taking quantization error into consideration.

3.4.4. Digital signal processing (DSP)

This is a very import subject in electronic communication. DSP simply means the processing of digitized discrete-time sampled signals. This processing is realized by general-purpose computers or by digital circuits such as ASICs, field-programmable gate arrays or specialized digital signal processors (DSP chips). Typical arithmetical operations include fixed-point and floating-point, real-valued and complex-valued, multiplication and addition.

Other typical operations supported by the hardware are circular buffers and look-up tables. Examples of algorithms are the Fast Fourier transform (FFT), finite impulse response (FIR) filter, Infinite impulse response (IIR) filter, and adaptive filters such as the Wiener and Kalman filters.

3.4.5. Non-linear signal processing (NSP)

NSP involves the analysis and processing of signals produced from nonlinear systems and can be in the time, frequency, or spatio-temporal domains.Nonlinear systems can produce highly complex behaviors including bifurcations, chaos, harmonics, and sub-harmonics which cannot be produced or analyzed using linear methods.

Some devices applied in signalprocessing include

i. Filters – for example analog (passive or active) or digital (FIR, IIR, frequency domain or stochastic filters, etc.)
ii. Samplers and analog-to-digital converters for signal acquisition and reconstruction, which involves measuring a physical signal, storing or transferring it as digital signal, and possibly later rebuilding the original signal or an approximation thereof.
iii. Signal compressors

In communication systems, signal processing may occur at:

- OSI layer-1 of the 7layer OSI model, the physical Layer (modulation, equalization, multiplexing, etc.);
- OSI layer 2, the Data Link Layer (Forward Error Correction);
- OSI layer 6, the Presentation Layer (source coding, including analog-to-digital conversion and signal compression).

3.5. Mathematical methods of Signal processing

In electronic communication systems, there are several mathematical approaches used for signal analysis. These include: differential equations, recurrence relation, transform theory, time-frequency analysis (for processing non-stationary signals), spectral estimation (for determining the spectral content (i.e., the distribution of power over frequency) of a time series), statistical signal processing (analyzing and extracting information from signals and noise based on their stochastic properties), linear time-invariant system theory, and transform theory, system identification and classification, calculus, vector spaces and linear algebra, functional analysis, probability and stochastic processes, detection theory, estimation theory, optimization, numerical methods, time series, data mining (for statistical analysis of relations between large quantities of variables (in this context representing many physical signals), to extract previously unknown interesting patterns). Each of these mathematical methods can be further studied from the references given in the reference section. We shall only lay emphasis on Fourier analysis owing to its over-riding relevance in communications systems.

3.5.1. Fourier analysis

In this course, we assert that the process of decomposing a function or signal into oscillatory components is referred to as Fourier analysis, while the operation of regenerating the function from these pieces is referred to as Fourier synthesis.

Take for example, in determining what component frequencies are present in a musical note, this would involve computing the Fourier transform of a sampled musical note. One could then re-synthesize the same sound by including the frequency components as revealed in the Fourier analysis. In mathematical contexts, Fourier analysis simply refers to the study of both operations mention above.The decomposition process itself is called a Fourier transformation as shown in Figure 3.18. In this case, the signalon the left can be viewed as a typical noise signal (A) while the signal function processing technique called Fourier transformation(B) reveals that it contains 5 well defined frequency components.

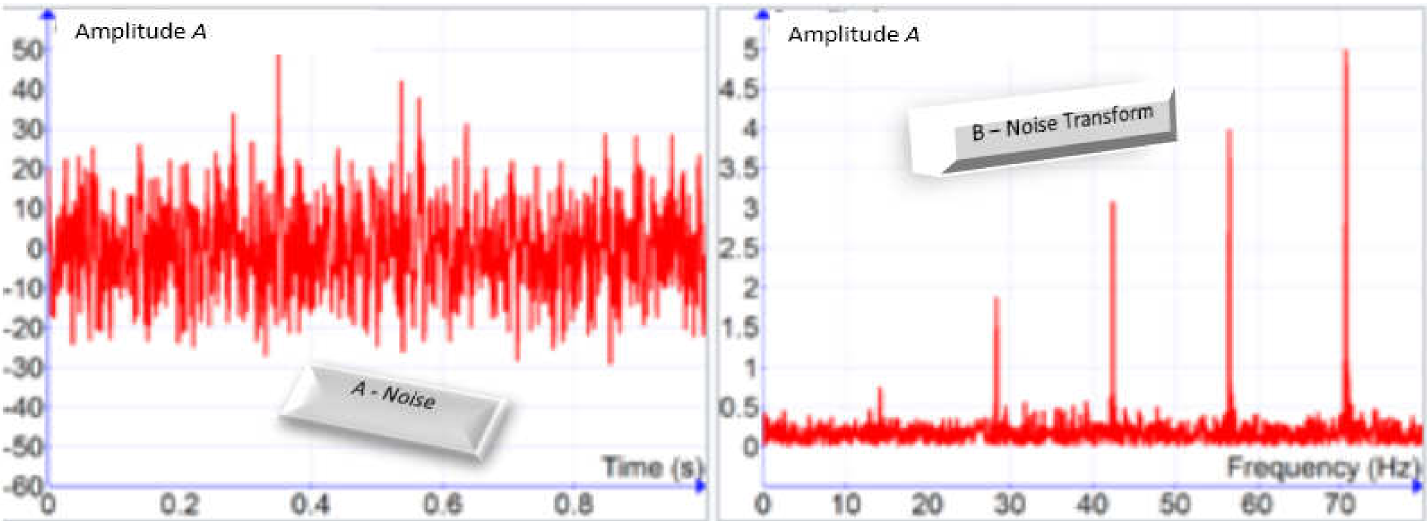


Figure 3.18. Noise and its Fourier Transformation

Figure 3.19 shows a Fourier transform and 3 distinct variations causeby periodic (at interval *T*) and /or periodic summation (at interval P) of the underlying time-domain function. The relative computational ease of DFT sequence and the insight it gives into *S (f)* make it a very widely accepted analysis tool.

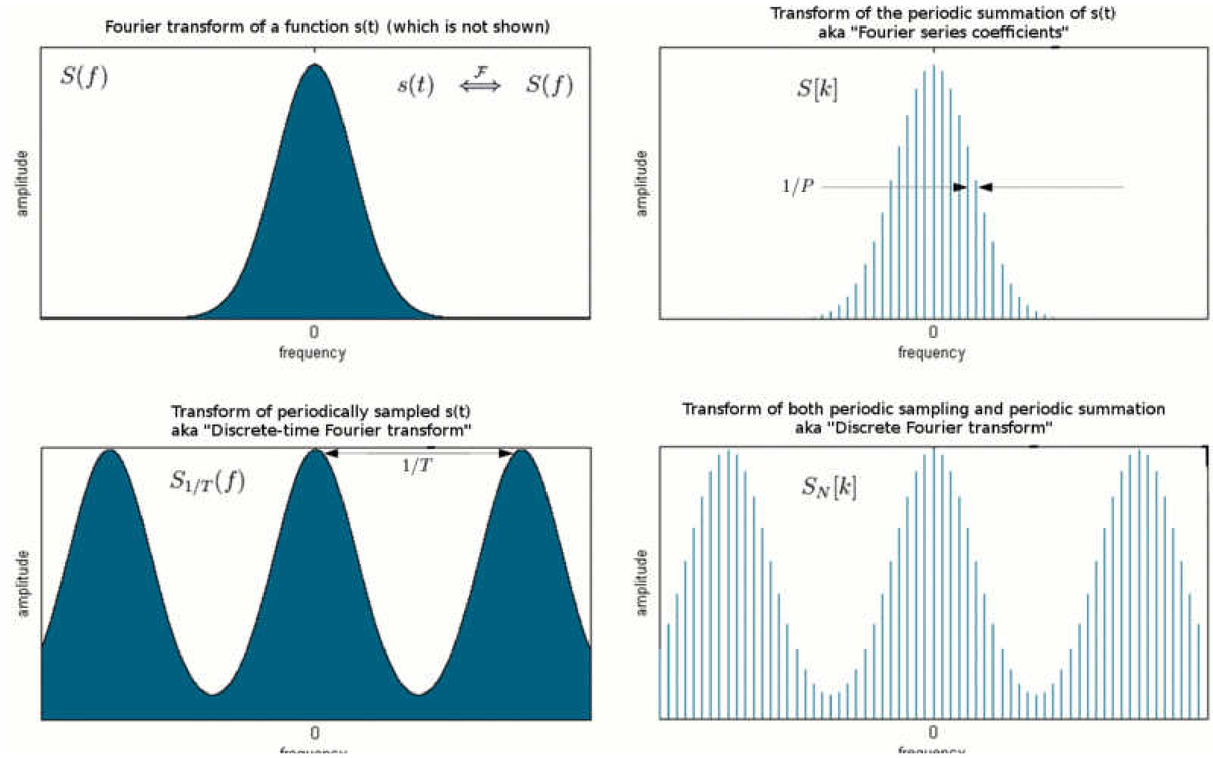


Figure 3.19. Fourier transform by Periodic Summation.

Readers must understand that Fourier analysis has many scientific applications but we shall focus on its application in signal processing.

3.5.2. Fourier Applicationin Signal Processing

When processing signals, such as audio, radio waves, light waves, seismic waves, and even images, Fourier analysis can isolate narrowband components of a compound waveform, concentrating them for easier detection or removal. Numerous signal processing techniques consist of Fourier-transform of signal, and the manipulation of Fourier-transformed data in a simple way, and reversing the transformation.Typical application scenario includes:

- Equalization of audio recordings with a series of bandpass filters;
- Passive sonar used to classify targets based on machinery noise
- Digital radio reception without a superheterodyne circuit, as in a modern cell phone or radio scanner;
- Image processing to remove periodic or anisotropic components from interlaced video, strip artifacts from strip aerial photography, or wave patterns from radio frequency interference in a digital camera;
- Cross correlation of similarity images for co-alignment;
- Generation of sound spectrograms used to analyze sounds;
- X-ray crystallography to reconstruct a crystal structure from its diffraction pattern;
- Fourier transform ion cyclotron resonance mass spectrometry to determine the mass of ions from the frequency of cyclotron motion in a magnetic field;
- Many other forms of spectroscopy, including infrared and nuclear magnetic resonance spectroscopies;

3.5.3. Fourier Analysis Methods

There are different Fourier methods available for analyzing continuous-time and discrete-time signals. We shall however, focus on four major Fourier analysis methods namely:

i. Continuous-time Fourier series (CTFS)
ii. Discrete-time Fourier series (DTFS)
iii. Continuous-time Fourier transform (CTFT)
iv. Discrete-time Fourier transform (DTFT)

3.5.3.1. Continuous-Time Fourier Series (CTFS)

In the CTFS method, for a continuous-time signal *x(t)*, the Fourier series representation of a signal over a representation time $t_0 < t < t_0 + T_F$ is defined as

$$x(t) = \sum_{k=-\infty}^{\infty} X[k] e^{j2\pi k f_F t} \tag{3.22}$$

Where *X[k]* is the harmonic function, *k* is the harmonic number and $f_F = 1/T_F$. The harmonic function is computed as

$$X[k] = \frac{1}{T_F} \int_{t_0}^{t_0 + T_F} x(t) e^{-j2\pi k f_F t} dt \tag{3.23}$$

The signal and its harmonic function form a pair which is indicated by the notation

$$x(t) \overset{FS}{\Leftrightarrow} X[k]$$

The left hand side of the above relation represents the signal in time domain while the right hand side represents the transformation of the signal to a "harmonic-number" domain.

3.5.3.2. Discrete-time Fourier series (DTFS)

For a discrete-time signal *x(n)*, the Fourier series representation over a representation time $n_0 \leqslant n < n_0 + N_F$ is defined as

$$x[n] = \sum_{k=\langle N_F \rangle} X[k] e^{j2\pi k n / N_F} \tag{3.24}$$

Where *X[k]* is the harmonic function and the harmonic function can be found as

$$X[k] = \frac{1}{N_F} \sum_{n=n_0}^{n_0 + N_F - 1} x[n] e^{-j2\pi k n / N_F} \tag{3.25}$$

The discrete-time signal and its harmonic function form a pair which is indicated by the notation below

$$x(n) \overset{FS}{\Leftrightarrow} X[k] \tag{3.26}$$

Where 'N_F' is the representation time and the notation $\sum_{k=\langle N_F \rangle}$ means summation over any range of consecutive '*k*' exactly N_F in length.

3.5.3.3. Continuous-time Fourier transform (CTFT)

In this method, the continuous-time Fourier transform (CTFT) is defined as

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \quad (3.27)$$

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df$$

The forward and inverse transforms are almost same, only the sign of the exponent and the variable of integration change.The continuous-time signal *x(t)* and its Fourier transform *X(f)* pairs are indicated by the notation below.

$$x(t) \overset{FT}{\Leftrightarrow} X(f) \quad (3.28)$$

3.5.3.4. Discrete-time Fourier transform (DTFT)

In DTFT, the discrete-time Fourier transform is defined by

$$x[n] = \int_{1} X(F) e^{j2\pi Fn} dF$$ and

$$X(F) = \sum_{n=-\infty}^{\infty} x[n] e^{-j2\pi Fn} \quad (3.29)$$

The signal and its harmonic function form a pair which is indicated by the notation below

$$x[n] \overset{FT}{\Leftrightarrow} X(F)$$

There still some salient properties of Fourier analysis methods which areexhibited by the continuous-time as well as discrete-time Fourier transform/Fourier series. This is outside the scope of this module. Besides, there are many similarities among different Fourier analysis methods and their relationshipprovides the basis of processing of the analog signal in digital domain.

When one wants to compute an arbitrary number (N) of discrete samples of one cycle of a continuous DTFT, $S_{1/T}(f)$ it can be done by computing the relatively simple DFT of $s_N[n]$, as defined above. In most cases, *N* is chosen equal to the length of non-zero portion of s[n]. Increasing N, known as zero-padding or interpolation, results in more closely spaced samples of one cycle of $S_{1/T}(f)$ Decreasing N, causes overlap (adding) in the time-domain (analogous to aliasing), which corresponds to decimation in the frequency domain. (Sampling the DTFT).

The DFT can be computed using a fast Fourier transform (FFT) algorithm, which makes it a practical and important transformation on computers.

4.0 Conclusion

We have shown that signals plays a very prominent role in conveys information in electronic communication. These signals could be audio, video, speech, image, communication, geophysical, sonar, radar, medical and musical signals. By understanding the complex behavior of signals, we can minimize noise effects in communication systems by using estimation theory as probabilistic approach to suppressing random disturbances. This conclusive unit has dealt with these issues.

5.0 Summary

In this unit, we have discussed the concept of Signalanalysiswith broad perspectives. We showed that signal analysis is very essential especially in cases where noise as an undesirable random disturbancesignals conflicts the desired signal. Its prevention deals with signal recovery techniques discussed in this unit. We have discussed the concept of signals in communication systems. A good classification of signals, its application and attributes was presented. The different types of signal Fourier analysis as well as the mathematical formulations were explored. The next module will focus on signal application in the context of modulation and demodulation systems.

6.0 Tutor-Marked Assignment (TMA)

1. What is Fourier analysis and outline four types of Fourier analysis methods namely:
2. Discuss the various OSI layers where signal processing may occur in communication systems
3. What are the 6 major classifications of signals and discuss them briefly?
4. List five examples of signals and outline their attributes.
5. What is a Sisyphean signal? Distinguish between Periodic and aperiodic signals.

7.0 References/Further Readings

1. Roland Priemer (1991). Introductory Signal Processing. World Scientific. p. 1. ISBN 9971509199.
2. "Aims and scope". IEEE Transactions on Signal Processing. IEEE.
3. T. H. Wilmshurst (1990). Signal Recovery from Noise in Electronic Instrumentation (2nd ed.). CRC Press. pp. 11 ff. ISBN 0750300582.
4. Proakis, John G.; Manolakis, Dimitris G. (2007-01-01). Digital Signal Processing. Pearson Prentice Hall. ISBN 9780131873742.
5. Paul Horowitz; Winfield Hill (2015). The Art of Electronics. Cambridge University Press. ISBN 9780521809269.
6. Vinod Kumar Khanna, Digital Signal Processing, 2009
7. In Bao-Liang Lu; Liqing Zhang & James Kwok. Neural Information Processing: 18th International Conference, Iconip 2011, Shanghai, China, November 13-17, 2011. Springer. pp. 506 ff. ISBN 3642249647.
8. M. N. Armenise; Caterina Ciminelli; Francesco Dell'Olio; Vittorio Passaro (2010). "§4.3 Optical gyros based on a fiber ring laser". Advances in Gyroscope Technologies. Springer. p. 47. ISBN 364215493X.
9. The optical reading process is described by Mark L. Chambers (2004). CD & DVD Recording for Dummies (2nd ed.). John Wiley & Sons. p. 13. ISBN 0764559567.
10. David McMahon (2007). Signals & Systems Demystified. New York: McGraw Hill. ISBN 978-0-07-147578-5.
11. M.J. Roberts (2011). Signals and Systems: Analysis Using Transform Methods & MATLAB. New York: McGraw Hill. ISBN 978-0073380681.
12. Hsu, P. H. Schaum's Theory and Problems: Signals and Systems, McGraw-Hill 1995, ISBN 0-07-030641-9
13. Lathi, B.P., Signal Processing & Linear Systems, Berkeley-Cambridge Press, 1998, ISBN 0-941413-35-7

14. Shannon, C. E., 2005 [1948], "A Mathematical Theory of Communication," (corrected reprint), accessed Dec. 15, 2005. Orig. 1948, Bell System Technical Journal, vol. 27, pp. 379–423, 623-656.
15. https://en.wikipedia.org/wiki/Signal
16. https://en.wikipedia.org/wiki/Signal_processing
17. Roland Priemer (1991). Introductory Signal Processing. World Scientific. p. 1. ISBN 9971509199.
18. Sengupta, Nandini; Sahidullah, Md; Saha, Goutam (August 2016). "Lung sound classification using cepstral-based statistical features". Computers in Biology and Medicine. **75** (1): 118–129. doi:10.1016/j.compbiomed.2016.05.013.
19. Alan V. Oppenheimer and Ronald W. Schafer (1989). Discrete-Time Signal Processing. Prentice Hall. p. 1. ISBN 0-13-216771-9.
20. Oppenheim, Alan V.; Schafer, Ronald W. (1975). Digital Signal Processing. Prentice Hall. p. 5. ISBN 0-13-214635-5.
21. Anastassiou, D. (2001). Genomic signal processing. IEEE.
22. Boashash, Boualem, ed. (2003). Time frequency signal analysis and processing a comprehensive reference (1st ed.). Amsterdam: Elsevier. ISBN 0-08-044335-4.
23. Stoica, Petre; Moses, Randolph (2005). Spectral Analysis of Signals (PDF). NJ: Prentice Hall.
24. Billings, S. A. (2013). Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains. Wiley. ISBN 1119943590.
25. iitg.vlab.co.in,. (2011). Exp-3 Fourier analysis of signals. Retrieved 25 August 2017, from iitg.vlab.co.in/?sub=59&brch=166&sim=799&cnt=1
26. http://web.mst.edu/~kosbar/test/ff/fourier/periodic.html
27. CUUK852-Mandal & Asif, (2006). Introduction to signals , 9:12, November 17.

MODULE 2 MODULATION FUNDAMENTALS

UNIT 1 AMPLITUDE MODULATION CONCEPTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
3.1 Expression for Amplitude Modulation Voltage
3.2 Waveform of Amplitude Modulated Voltage
3.3 Sidebands Produced in Amplitude Modulation
3.4 AM Power and Current Derivations
3.4.1. Power Computations in Amplitude Modulated Wave
3.4.2. Amplitude Modulated Current Computation
3.4.3. Modulation by Sine Wave Cascades
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0References/Further Readings

1.0 INTRODUCTION

The concept of modulation is very important in electronic communications systems. In Section 3.3, a good definition of modulation is given as the process by which some characteristics, usually amplitude, frequency or phase of a voltage (usually sinusoidal voltage) is varied in accordance with the instantaneous value of some other voltage, called the modulating voltage. We explained that the term carrier is applied to the voltage whose characteristic is varied and the term modulating voltage (or signal) is used for the voltage in accordance with which the variation is made. The carrier frequency is the frequency of the carrier voltage being modulated.Usually the modulation frequency is considerably lower than the carrier frequency. But this is not inherent in the definition and in exceptional cases, the carrier frequency may be lower the modulation frequency as in carbon button transmitter or a vacuum tube amplifier.Here, you will learn about AM mathematical expression, the power equations, and the different types of AM systems, their applications, and topologies. We shall also examine various concepts on amplitude modulation systems. Towards the end of this unit, samples tutorials will be discussed.

2.0 OBJECTIVES

After going through this unit, you should be able to:

- ✓ understand the concept of amplitude modulation systems
- ✓ differentiate between different types of amplitude modulation systems
- ✓ understand the different application of amplitude modulation systems
- ✓ compare the different amplitude modulation systems
- ✓ Highlight the importance of amplitude modulation systems
- ✓ know the importance of using amplitude modulation systems
- ✓ understand the mathematical models of amplitude modulation by sine wave cascading.

3.0 MAIN CONTENT

3.1. Expression for Amplitude Modulation Voltage

The modulation process usually involves varying one of the following three quantities: (i) amplitude f_c (ii) frequency ω_c and (iii) phase angle θ of the carrier voltage, in accordance with some function of the instantaneous value of the modulating voltage.

Accordingly modulation process could be amplitude, frequency or phase modulations depending upon whether the amplitude $V_{C,}$ frequency ω_c or the phase angle θ of the carrier voltage is varied. In most cases, the frequency modulation and phase modulation are classified together under angular modulation. It is possible to produce simultaneously, amplitude, frequency and phase modulation of the carrier voltage by varying all the three parameters, $V_{c,}\omega_c$ and θ simultaneously but is commercial radio transmitters, care is taken to produce only one type of modulation with the exclusion of the other two.

As we have seen in Equ 1.1, let the carrier voltage be represented by the expression.

$v_c = v_c \cos[\omega_{c^t} + \theta(t)]$

Where $t = time$ (3.1)

ω_c =Angular frequency of the voltage v_c in radians/second

And θ =Phase angle (3.2)

$\omega_c = 2\pi f_c$

Where f_c is the frequency of the carrier voltage in Hz

In amplitude modulation, the amplitude of the carrier voltage varies in accordance with the instantaneous value of the modulating voltage or the signal be given by the expression,

$$v_m = V_m \cos \omega_{m^t} \quad (3.3)$$

Where ω_m is the angular frequency of the signal and V_m is the amplitude.

Let the carrier voltage be given by the expression,

$$v_c = V_c \cos \omega_{c^t} \quad (3.4)$$

In Equ. (3.4), for convenience in calculation, the phase angle θ has been taken as zero, since it does not play any part in the modulation process. This however, does not in any way reduce the generality of the expression.On amplitude modulation, amplitude of the carrier no longer remains but varies with time as given by the following expression:

$$V(t) = V_c + K_{a.}V_m \cos \omega_{m^t} \quad (3.5)$$

Where $k_{a.}V_m\ cos\omega_{m^t}$represents the change in carrier amplitude.

The instantaneous value of modulated carrier voltage is then given by,

$$v = V(t)cos\omega_{m^t} \quad (3.6)$$

or $v = [V_c + K_a V_m \cos \omega_{m^t}] \cos \omega_{c^t}$ (3.7)

Equ. (3.7) may be written as,

$$v = V_c \left[1 + \frac{K_a V_m}{V_c}\ cos\omega_{m^t}\right] \cos \omega_{c^t}$$

$$v = V_c[1 + m_a\ cos\omega_{m^t}]cos\omega_{c^t} \quad (3.8)$$

Where $m_a$is the modulation index or modulation factor or depth of modulation and is given by,

$$m_a = \frac{K.V_m}{V_c} \quad (3.9)$$

Now, by establishing that $100\ \times m_a$, this then gives the percentage modulation.

3.2. Waveform of Amplitude Modulated Voltage

Figure 3.1 (a) shows the waveform unmodulated carrier voltage. Figure. 3.1 (b) gives the waveform of sinusoidal modulating voltage. Figure. 3.1 (c) gives the waveform of amplitude modulated carrier voltage.

Now, from Figure 3.1, we can see that frequency of carrier remains unaltered but its amplitude varies in accordance with the variation of the modulating voltage $v_{m.}$

It may further be seen that,

$$m_a = \frac{V_c \max - V_c}{V_c} \tag{3.10}$$

Also, $m_a = V_c - \frac{V_c \min}{V_c}$ (3.11)

Hence, $$m_a = \frac{V_c \max - V_c \min}{2V_c} = \frac{V_c \max - V_c \min}{V_c \max + V_c \min} \tag{3.12}$$

With the help of Equ. (3.12), the modulation m_a may be experimentally determined by applying the amplitude modulated voltage to Y-Y deflection plates of a cathode ray oscilloscope and applying a suitable sweep voltage (time-base voltage) to the X-X deflection plates. The pattern shown in Figure 3.1, then appears on the screen of the CRO from which values of V_{amax} *and* V_{cmin} may be observed. The modulation index may then be calculated.

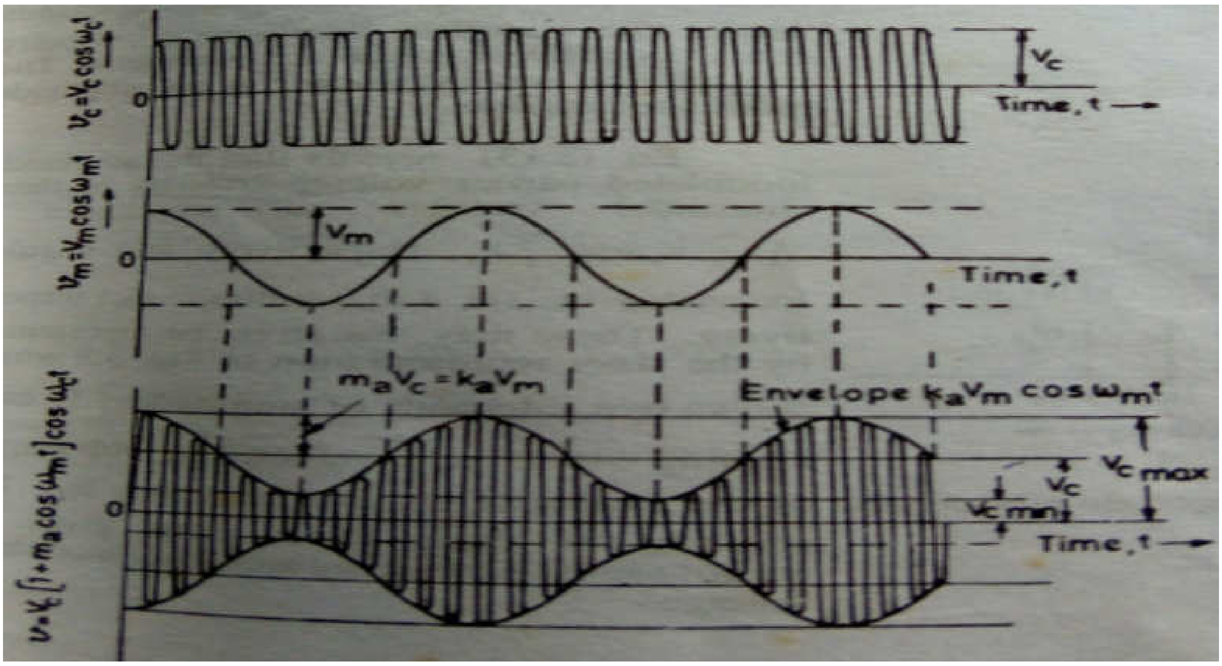


Figure 3.1. Waveform of amplitude modulated carrier voltage.

3.3. Sidebands Produced in Amplitude Modulation

The lower sideband term and the upper sideband term are located in frequency spectrum on either side of the carrier at a frequency interval of ω_m as shown in Figure 3.2a, the magnitude of both the upper and lower sidebands is m_a /2 times the carrier amplitude $V_{C.}$ If modulation index

m_a is equal to unity, then each sideband has amplitude half of the carrier amplitude. Figure 3.2a, shows a plot of the frequency spectrum of the amplitude modulated voltage.

Amplitude modulation thus shifts the intelligence from audio frequency level to the level of carrier frequency $\omega_{c.}$ Also, the intelligence appears in the form of two sideband symmetrically placed relative to the carrier frequencyω_c. Each of these sideband carries the complete intelligence original as contained in the signal before modulation. The intelligence occurs twice in amplitude Modulation carrier.

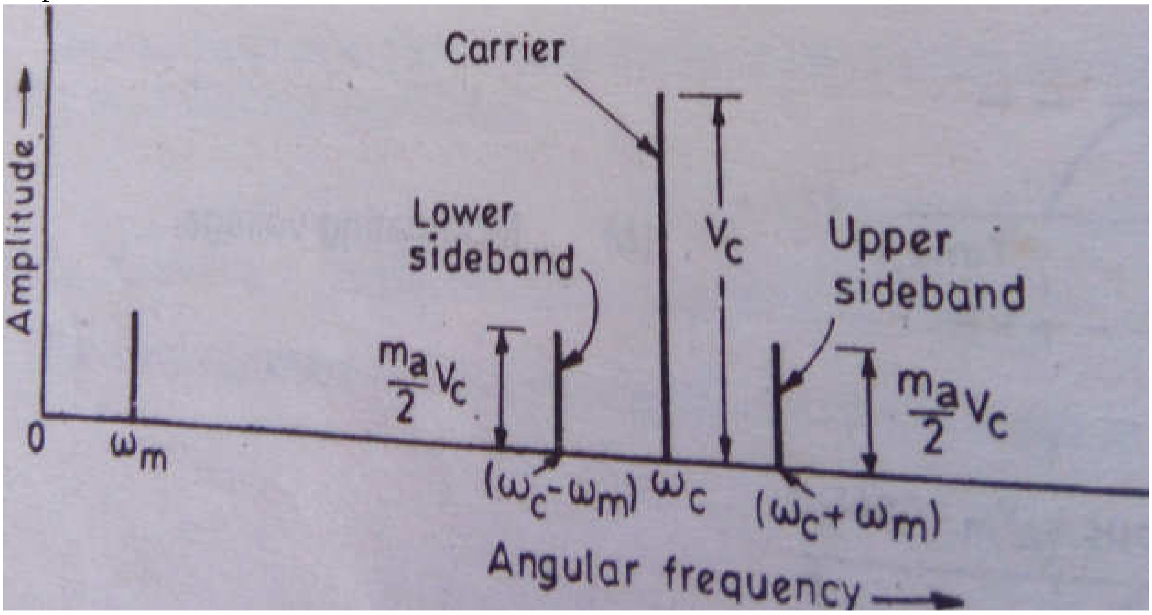


Figure 3.2a. Plot of frequency spectrum of amplitude modulated voltage

Equ. (3.8) for amplitude modulated carrier may be expanded in the following form

$v = V_c \cos \omega_{ct} + \dfrac{m_a V_c}{2} [2 \cos \omega_{ct} \cos \omega_{mt}]$

or $\quad v = V_c \cos \omega_{ct} + \frac{m_a}{2} V_c \cos\ (\omega_c + \omega_m)^t + \frac{m_a}{2} V_c \cos\ (\omega_c - \omega_m)t$(3.13)

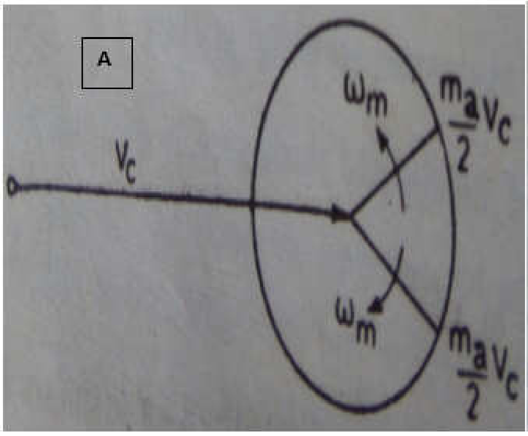
Equ.(3.13) reveals that the sinusoidal carrier voltage on being amplitude modulated by a single sinusoidal modulating voltage consists of the following frequency terms:

(i) Original carrier voltage $V_c \cos \omega_{ct}$ of angular frequency $\omega_{c.}$

(ii) Upper sideband term $\frac{m_a}{2} V_c\ (\omega_c + \omega_m)^t$of angular frequency$(\omega_c + \omega_m)$.

(iii) lower sideband term$\frac{m_a}{2} V_c \cos\ (\omega_c - \omega_m)^t$of angular frequency$(\omega_c - \omega_m)$.

Equ. (3.13) reveals that the amplitude Modulated carrier voltage consists of the Sinusoidal voltages of amplitudes$V_{c,} \frac{m_a}{2} V_c and \frac{m_a}{2} V_{c,}$respectively and angular Frequency $\omega_{c,}\ (\omega_c + \omega_{m)} and\ (\omega_c - \omega_m)$ respectively. These may, therefore, be represented by the phasor representation of Figure. 3.3 which shows two sidebands of amplitude$\frac{m_a}{2} V_c$ each rotating with angular velocity $\omega_{c.}$ with respectively to the carrier of annular frequency $\omega_{c.}$ the resultant of these three phasor gives the amplitude of the resultant amplitude modulation voltage.



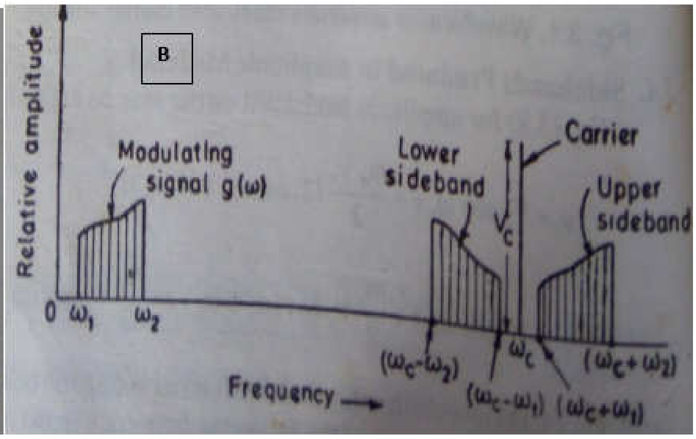


Figure. 3.3a). Phasor representation of modulating voltage and amplitude modulated voltage. Figure.3.4b)Frequency spectrum of a complex amplitude modulating voltage.

So far we have assumed the modulating signal to be a single frequency tone. In practice, however, the modulating voltage has a complex waveform. If the modulating voltage is periodic, it may be represented Fourier series. On the other hand, if the modulating voltage is non-periodic, it may be represented by Fourier integral. In any case, the modulating signal consists of a band of frequency of different amplitudes and phases as shown Figure 3.4 by(ω). Each frequency term in this modulating signal produces on modulation a pair of sideband terms. The entire modulating signals then produces two sidebands symmetrically dispose about the carrier as shown in Figure 3.4. Here $\omega_1 and \omega_2$ are the lowest and the highest frequencies in the modulating signal.

3.4. AM Power and Current Derivations

3.4.1. Power Computations in Amplitude Modulated Wave

We have seen that the carrier component of the amplitude modulated wave has the same amplitude as the unmodulated carrier. However, the amplitude wave contains two sideband components as well. Evidently, therefore, the modulated wave contains more power than the carrier before modulation. Further since the amplitude of the sideband depends on the modulation index, the total power in the modulated wave will also depend on the modulation index. The total power in the modulated wave may be expressed as Equ. 3.14.

$$P_t = P_c + P_{LSB} + P_{USB} \tag{3.14}$$

$$P_t = \frac{V_{carr}}{R} + \frac{V_{LSB}^2}{R} + \frac{V_{USB}^2}{R} \tag{3.15}$$

Where all the three voltage are *r.m.s.* values and R is resistance in which the power is dissipated the first term on the RHS in Equ. (3.14) is the unmodulated carrier power and is given by

$$P_c = \frac{V_{carr}^2}{R} = \frac{(V_c\sqrt{2})^2}{R} = \frac{V_c^2}{2R} \tag{3.16}$$

Similarly, $P_{LSB} = P_{USB} = \frac{V_{LSB}^2}{R}$

$$= (\frac{m_a \frac{V_c}{2}}{\sqrt{2}})^2 * \frac{1}{R} = \frac{m_a^2 V_c^2}{8R} = \frac{m_a^2}{4}\frac{V_c^2}{2R} \tag{3.17}$$

Substituting Equ. (3.16) and (3.13) into Equ. (3.14), we get

$$P_t = \frac{V_c^2}{2R} + \frac{m_a^2}{4}\frac{V_c^2}{2R} + \frac{m_a^2}{4}\frac{V_c{}^2}{2R} = P_c + \frac{m_a^2}{4}P_c + \frac{m_a^2}{4}P_c \tag{3.18}$$

$$P_t = P_c(1 + \frac{m_a^2}{2}) \tag{3.19}$$

Equ. (3.19) relates the total power P_t in the amplitude modulated wave to the unmodulated carrier power. The equation reveals that the maximum power in the AM wave is P_t =1.5P_c when m=1.

3.4.2. Amplitude Modulated Current Computation

Often it is relatively easy to measure the modulated and unmodulated carrier current. Such as a situation arise when the antenna current of a transmitter is metered. We may the compute the modulation index m_a from these two currents.

Let I_c be the unmodulated current and let I_t be the total or the modulated current of an AM transmitter both currents being RMS values. Let R be the resistance into which these current flow.Then,

$$\frac{P_t}{P_c} = \frac{I_t^2 R}{I_c^2 R} = (\frac{I_t}{I_c})^2 = 1 + \frac{m_a^2}{2}$$

Hence, $$\frac{I_t}{I_c} = \sqrt{1 + \frac{m_a^2}{2}} \tag{3.20}$$

Or $$I_t = I_C\sqrt{1 + \frac{m_a^2}{2}} \tag{3.21}$$

3.4.3. Modulation by Sine Wave Cascades

In several sine waves scenario in practice as shown in Figure 3.5, the carrier is usually modulated by several sine wave simultaneously. It is of interest to calculate the resulting power condition.

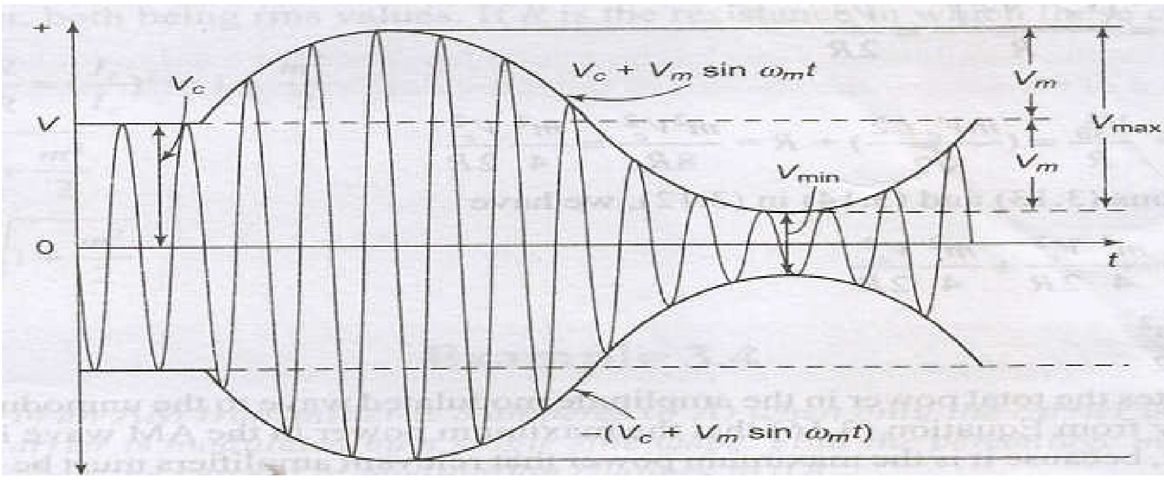


Figure 3.5. Plot of Time domain representation of the AM wave.

The procedure includes calculating the total modulation index and the substituting it in Equ. (3.19) to compute the total power P_r. There are two methods of computing the total modulation index $m_{1.}$ these are given below.

MethodI.
Let $V_{1,}V_{2,}V_{3,}$ etc. be the simultaneous modulating voltages. Then the total modulating voltage V_t equals the square root of the sum of the squares of the individual stages.

Thus, $V_t = \sqrt{V_1^{2+} V_2^{2+} V_3^2 + \cdots}$
On dividing both sides by $V_{c,}$ we get

$$\frac{V_t}{V_c} = \sqrt{\frac{V_1^2 + V_2^2 + V_3^2 + \ldots.}{V_c}} = \sqrt{\frac{V_1^2}{V_c^2} + \frac{V_2^2}{V_c^2} + \frac{V_3^2}{V_c^2} + \cdots}$$

Hence, $m_t = \sqrt{m_t^2 + m_2^2 + m_3^2 + \cdots}$ (3.22)

Method II.
Equ. (3.19) may be written as,

$$P_t = P_c + P_c \frac{m_a^2}{2} = P_c + P_{SB} \tag{3.23}$$

Where $P_{SB}$is the total sideband power and is given by,
$$P_{SB} = P_C \frac{m_a^2}{2} \tag{3.24}$$

When several sine simultaneously modulate the carrier, the carrier power P_C remains unaltered but the total waves sideband power P_{SBT} is now the sum of the individual sideband powers. Thus, we get

$P_{SBT} = P_{SB1} + P_{SB2} + P_{SB3} + \cdots$ (3.25)

On substituting Equ. (3.24) into Equ. (3.25), we get

$$\frac{P_C m_t^2}{2} = \frac{P_C m_1^2 +}{2} + \frac{P_{C m_2^2}}{2} + \frac{P_C m_3^2}{2} + \cdots$$

Hence, $m_t^2 = m_1^2 + m_2^2 + m_3^2 + \cdots$ (3.26)

It may be noted that total modulation index m_t must not exceed unity or else distortion will result.

4.0 Conclusion

So far, we have provided a fundamental understandingto the concept of amplitude modulation systems. In order to reinforce our knowledge, we differentiated between different types of amplitude modulation systems. A good understanding of the different application of amplitude modulation systems has been presented. We have compared between the different amplitude modulation systems while highlighting the importance of amplitude modulation systems.

5.0 Summary

- We have seen that the concept of modulation is very important in electronic communications systems.
- The modulation process usually involves varying one of the following three quantities: (i) amplitude f_c (ii) frequency ω_c and (iii) phase angle θ of the carrier voltage, in accordance with some function of the instantaneous value of the modulating voltage.
- In amplitude modulation, the amplitude of the carrier voltage varies in accordance with the instantaneous value of the modulating voltage or the signal.
- Amplitude modulation thus shifts the intelligence from audio frequency level to the level of carrier frequency $\omega_{c.}$
- the carrier component of the amplitude modulated wave has the same amplitude as the unmodulated carrier.
- Modulation by Sine Wave Cascades can enable the calculation of the calculating the total modulation index
- The total modulating voltage V_t equals the square root of the sum of the squares of the individual stages.

5.0 Tutor-Marked Assignment (TMA)

We shall present some Self-Tutorials for study assimilation. Please pay attention to the steps.

Example 1.

A sinusoidal carrier of frequency 1 MHz and amplitude 100 volts is modulated by sinusoidal voltages of frequency 5 kHz producing 50% modulation. Calculate the frequency and amplitude of upper and lower sidebands considering Figure 3.6.

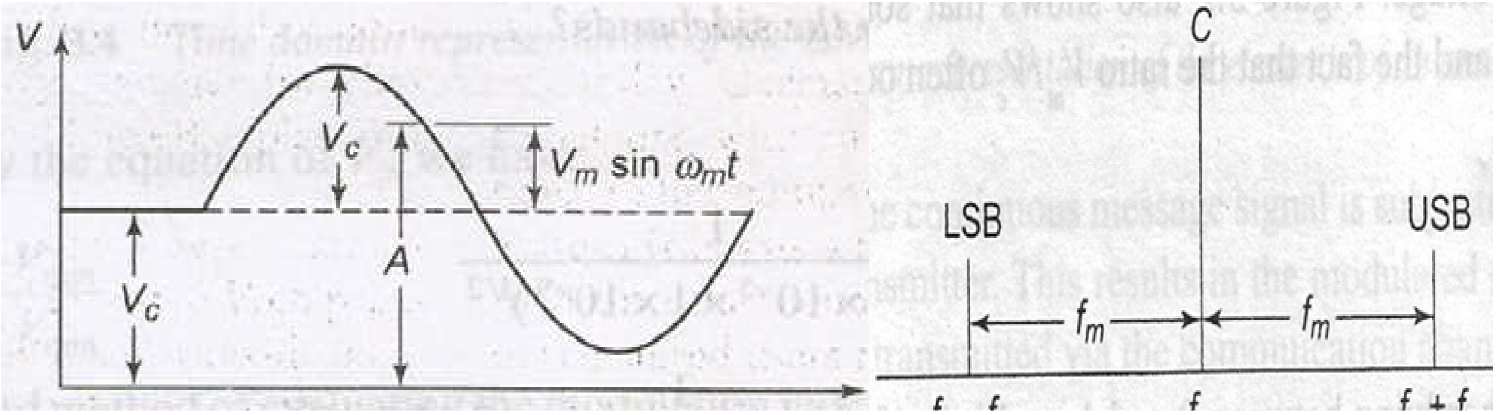


Figure 3.6.a) Amplitude of an AM wave b) Frequency spectrum of an AM wave.

Solution.
Frequency of upper sideband =1 MHz +5 kHz = 1005 kHz.
Frequency of lower sideband = 1 MHz – 5 kHz = 995 kHz.
Amplitude of each sideband term$\frac{m_a}{2} V_c = \frac{0.5}{2} \times 100 = volts.\boldsymbol{Ans}.$

Example 2

The tuned circuit of the oscillator in a simple AM transmitter uses a coil of 40 UH and shunts capacitor of value 1nf. If the oscillator output is modulated by audio frequency up to 5 kHz, what is the frequency range occupied by the sidebands?

Solution.

$$fc = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{40 \times 10^{-6} \times\ 1 \times 10^{-9}}} H = 7.96\ \times 10^{5\ Hz.}$$

= 796 kHz.

Since the highest modulating frequency is 5 kHz, the frequency range will extend from 5 kHz below to 5 kHz above the carrier frequency, i.e. from 791 to 801 kHz.

Example 3

A sinusoidal carrier voltage of amplitude 100 volts is amplitude modulated by a sinusoidal voltage of frequency 10 kHz resulting in maximum modulated carrier amplitude of 120 volts. Compute the modulation index.

Solution.

$$m_a = \frac{V_{Cmax} - V}{V} = \frac{120 - 100}{100} = 0.2.\boldsymbol{Ans}.$$

Example 4.

A sinusoidal carrier voltage of frequency 1200 kHz is amplitude modulated by a sinusoidal voltage of frequency 20 kHz resulting in maximum and minimum. Modulated carrier amplitude of 110 volts and 90 volts respectively. Calculate (a) the frequency of lower and upper sidebands (b) the unmodulated carrier amplitude (c) modulation index and amplitude of each sideband.

Solution.

(a) Lower sidebands frequency= (1200-20) kHz = 1180 kHz.
Upper sideband frequency= (1200 +20) kHz = 1220 Khz.

(b) Unmodulated carrier amplitude$V_c = \frac{V_{cmax}+V_c max}{2}$

$= \frac{(110+90)}{2} = 100\ volts.$

(c) Modulation index, $m_a = \frac{V_{cmax^-} - V_{cmin}}{V_{cmax}+V_{cmin}} = \frac{(110-90)}{110+90} = 0.1$

(d) Amplitude of each sideband $= \frac{m_a}{2} V_c = 0.05\ \times 100 = 5\ volts. \boldsymbol{Ans}$

Example 5.

The *rms* value of a carrier voltage is 100 volts. Compute its *rms* value when it has been amplitude modulated by a sinusoidal audio voltage to a depth of (a) 20% and (b) 40%.

Solution.

$P_t = P_c[1 + \frac{m_a^2}{2}]$

or$\frac{V_{rms}^2}{2} = \frac{V_{carr}^2}{2}\ [1 + \frac{m_a^2}{2}]$

Hence , $V_{rms}^2 = V_{carr}^2[1 + \frac{m_a^2}{2}]$

Or $V_{rms} = V_{carr}\sqrt{+\frac{m_a^2}{2}}$

i. $V_{rms} = 100\sqrt{1 + \frac{(0.2)^2}{2}} = 101\ volts$

ii. $V_{rms} = 100\ \sqrt{1 + \frac{(0.4)^2}{2}} = 103.9\ volts$

Example 6.

The *rms* of a radio frequency voltage is 80 volts. After amplitude modulation by a sinusoidal audio voltage. The *rms* value of the RF voltage becomes 88volts. Compute the modulation index.

Solution.

$V_{rms} = \ V_{carr}\ [1 + \frac{m_a^2}{2}]$

Hence, $m_a = \ \sqrt{2[\left(\frac{V_{rms}}{V_{car}}\right)^2 - 1]} = \sqrt{2[(\frac{88}{80})^2 - 1]} = 0.6481.$

Example 7.

Unmodulated RFcarrier power of 10kw sends a current of 10 amp rms through an antenna. On amplitude modulation by another sinusoidal voltage, the antenna current increases to 11.6 amps calculate (a) the modulation index and (b) carrier power after modulation.

Solution

Let *rms* antenna current before and after modulation be$I_\circ$ and I respectively.

Then $m_a = \sqrt{2[(\frac{I}{I_\circ})^2 - 1]} = \ \sqrt{2[(\frac{11.6}{10})^2 - 1]} = 0.8314$

Power of carrier after modulation is given by,

$$P = P_c\left[1+\frac{m_a^2}{2}\right] = 10\ \left[1+\frac{(0.8314)^2}{2}\right] = 13.456\ kw.$$

Example 8.

The *rms* value of a carrier voltage after amplitude modulation to a depth of 40% by an sinusoidal modulating voltage is 55 volts. Calculate the *rms* value of carrier voltage when amplitude modulated to a depth of 80%.

Solution.

$$V_{carr} = \frac{V_{rms}}{\sqrt{1+(m_a^2/2)}} = \frac{55}{\sqrt{1+(0.4)^2/2}} = 52.93\ volts.$$

At 80% modulation, the modulated carrier rms voltage is given by,

$$V_{rms} = V_{carr}\sqrt{1+m_a^2} = 52.93\sqrt{1+\frac{(0.8)^2}{2}} = \quad 60.81\ volts.$$

Example 9.

An lkW carrier is modulated to a depth of 60%. Calculate the total power in the modulated wave.

Solution.

$$P_t = P_C\left[1+\frac{m_a^2}{2}\right] = 1\left[1+\frac{(0.6)^2}{2}\right] = 1.18\ kW.$$

Example 10.

A broadcast radio transmitter radiates 8 kW when the modulation percentage is 50. What is the unmodulated carrier power?

Solution.

$$P_C = \frac{P_t}{1+m_a^2/2} = \frac{8}{1+(0.5)^2/2} = 7.111\ kW.$$

Example 11.

A broadcast transmitter radiates 4.72 kW when the modulation percentage is 60. Calculate the total power when the modulation has been reduced to 40 per cent.

Solution.

$$P_C = \frac{P_t}{1+m_{a\ /2}^2} = \frac{4.72\ kW}{1+(0.6)^2/2} = 4kW$$

With 40 per cent modulation,

$$P_t = 4\ kW\left[1+\frac{(0.4)^2}{2}\right] = 4\ [1+0.08] = 4.32\ kW.$$

Example 12.

A radio telephone transmitter using amplitude modulation has unmodulated carrier output power of 20 kW and can be modulated to maximum depth of 80% by a sinusoidal modulating voltage without causing overloading. Find the value to which unmodulated carrier power may be increased without resulting in overloading if the maximum permitted modulation index is restricted to 60%.

Solution.

Modulated carrier power

$$P_t = P_C\left(1+\frac{m_a^2}{2}\right) = 20\ [1+\frac{(0.8)^2}{2}] = 26.4$$

This is the maximum power which can be handled by the transmitter without causing overloading. The increased unmodulated carrier power is then is given by,

$$26.4P_0[1+\frac{(0.6)^2}{2}] = 1.18P_0$$

Hence,

$$P_0' = \frac{26.4}{1.18} = 22.38\ kW.$$

Example 13.

A radio transmitter radiates 10 Kw with the carrier unmodulated and 11.25 kW when the carrier is modulated by a sinusoidal voltage. Calculate the modulation index. Another sine wave is capable of producing 30% modulation. If both the sine waves simultaneously modulate the carrier determine the total radiated power.

Solution.

$$\frac{m_1^2}{2} = \frac{P_t}{P_\circ} - 1 = \frac{11.25}{10} - 1 = 0.125$$

Hence

$$m_1 = \sqrt{2 \times 0.125} = 0.5$$

With simultaneous modulation by the two sine waves, this now gives

$$m_t = \sqrt{m_1^2 + m_2^2} = \sqrt{(0.5)^2 + (0.3)^2} = 0.58$$

Hence,

$$P_t = P_0\left(1+\frac{m_t^2}{2}\right) = 10\left[\frac{1+(0.58)^2}{2}\right] = 11.7\ kW.$$

Example 14.

The tuned circuit of the oscillator in a simple AM transmitter employs a 50-microheary (50-µH) coil and t 1-nanofarad (1-nF) capacitor. If the oscillator output is modulated by audio frequencies up to 10 kHz, what is the frequency range occupied by the sidebands?

Solution

$$f_c = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi(5\times10^{-5}\times1\times10^{-9})^{1/2}}$$

$$\frac{1}{2\pi(5\times10^{-14})^{1/2}} = \frac{1}{2\pi\sqrt{5\times10^{-7}}} = 7.12 \text{ x } 10^5$$

Since the highest modulating frequency is 10 kHz, the frequency range occupied by the sidebands will range 'from 10 kHz above to 10 kHz below the carrier, extending from 722 to 702 kHz.

Example 15.

A 400-watt (400-W) carrier is modulated to a depth of 75 percent. Calculate the total power in the modulated wave.

Solution

$P_{AM} = P_c\left[1+\frac{m^2}{2}\right] = 400\ [1 + \frac{0.75^2}{2} = 400\ \times 1.281$

= 512.5W.

Example 16

A broadcast radio transmitter radiates 10 kilowatts (10 kW) when the modulation percentage is 60. How much of this is carrier power?

Solution

$P_c = \frac{P_t}{1+m^2/2} = \frac{10}{1+0.62/2} = \frac{10}{1.18} = 8.47KW$

Example 17

The antenna current of an AM transmitter is 8 amperes (8A) when only the carrier is sent, but it increases to 8.93A when the carrier is modulated by a single sine wave. Find the percentage modulation. Determine the antenna current when the percent of modulation changes to 0.8.

Solution

$$\left[\frac{I_t}{I_c}\right]^2\ 1+\frac{m^2}{2}$$

$$\frac{m^2}{2} = \left[\frac{I_t}{I_c}\right]^2 - 1$$

$$m = \sqrt{2\left[\left(\frac{I_t}{I_c}\right)^2 - 1\right]}$$

Here,

$$m = \sqrt{2\left[\left(\frac{8.93}{8}\right)^2 - 1\right]} = \sqrt{2[(1.116)^2 - 1]}$$

$$\sqrt{2(1.246-1)} = \sqrt{0.492} = 0.701 = 70.1\%.$$

For the second part, we have I

$$I_t = I_c\sqrt{1+\frac{m^2}{2}} = 8\sqrt{1+\frac{0.8^2}{2}} = \sqrt{1+\frac{0.64}{2}}$$

$$= 8\sqrt{1.32} = 8 \times 1.149 = 9.19A$$

Example 18

A certain transmitter radiates 9 kW with the carrier unmodulated, and 10.125 kW when the carrier is sinusoidal modulated. Calculate the modulation index. If another sine wave is simultaneously transmitted with modulation index 0.4, determine the total radiated power.

Solution

$$\frac{m^2}{2} = \frac{P_t}{P_c} - 1 = \frac{10.125}{9} - 1 = 1.125 - 1 = 0.125$$
$$m^2 = 0.125 \times 2 = 0.250$$
m = $\sqrt{0.25} = 0.50$

For the second part, the total modulation index will be

$$m_t = \sqrt{m_1^2 + m_2^2} = \sqrt{0.5^2 + 0.4^2} = \sqrt{0.25 + 0.16} = \sqrt{0.41} = 0.64$$
$$P_{AM} = P_c\left(1 + \frac{m_t^2}{2}\right) = 9\left(1 + \frac{0.64^2}{2}\right) = 9(1 + 0.205) = 10.84KW$$

Example 19

The antenna current of an AM broadcast transmitter, modulated to a depth of 40 percent by an audio sine wave, is 11A. It increases to 12 A as a result of simultaneous modulation by another audio sine wave. What is the modulation index due to this second wave?

Solution

$$I_c = \frac{I_t}{\sqrt{1 + m^2/2}} = \frac{11}{\sqrt{1 + 0.4^2/2}} = \frac{11}{\sqrt{1 + 0.08}} = 10.58$$

From the modulation index Equation, here the modulation index is the total modulation index m_t, we obtain

$$m_t = \sqrt{2\left[\left(\frac{I_T}{I_C}\right)^2 - 1\right]} = \sqrt{2\left[\left(\frac{12}{10.58}\right)^2 - 1\right]} = \sqrt{2(1.286 - 1)}$$

= $\sqrt{2 \times 0.286} = 0.757$.

We obtain

$m_2 = \sqrt{m_t^2 - m_1^2} = \sqrt{0.757^2 - 0.4^2} = \sqrt{0.573 - 0.16} = \sqrt{0.413}$= 0.643.

7.0 References/Further Readings

1. Roland Priemer (1991). Introductory Signal Processing. World Scientific. p. 1. ISBN 9971509199.
2. "Aims and scope". IEEE Transactions on Signal Processing. IEEE.
3. T. H. Wilmshurst (1990). Signal Recovery from Noise in Electronic Instrumentation (2nd ed.). CRC Press. pp. 11 ff. ISBN 0750300582.
4. Proakis, John G.; Manolakis, Dimitris G. (2007-01-01). Digital Signal Processing. Pearson Prentice Hall. ISBN 9780131873742.
5. Paul Horowitz; Winfield Hill (2015). The Art of Electronics. Cambridge University Press. ISBN 9780521809269.
6. Vinod Kumar Khanna, Digital Signal Processing, 2009
7. In Bao-Liang Lu; Liqing Zhang & James Kwok. Neural Information Processing: 18th International Conference, Iconip 2011, Shanghai, China, November 13-17, 2011. Springer. pp. 506 ff. ISBN 3642249647.
8. M. N. Armenise; Caterina Ciminelli; Francesco Dell'Olio; Vittorio Passaro (2010). "§4.3 Optical gyros based on a fiber ring laser". Advances in Gyroscope Technologies. Springer. p. 47. ISBN 364215493X.

9. The optical reading process is described by Mark L. Chambers (2004). CD & DVD Recording for Dummies (2nd ed.). John Wiley & Sons. p. 13. ISBN 0764559567.
10. David McMahon (2007). Signals & Systems Demystified. New York: McGraw Hill. ISBN 978-0-07-147578-5.
11. M.J. Roberts (2011). Signals and Systems: Analysis Using Transform Methods & MATLAB. New York: McGraw Hill. ISBN 978-0073380681.
12. Hsu, P. H. Schaum's Theory and Problems: Signals and Systems, McGraw-Hill 1995, ISBN 0-07-030641-9
13. Lathi, B.P., Signal Processing & Linear Systems, Berkeley-Cambridge Press, 1998, ISBN 0-941413-35-7
14. Shannon, C. E., 2005 [1948], "A Mathematical Theory of Communication," (corrected reprint), accessed Dec. 15, 2005. Orig. 1948, Bell System Technical Journal, vol. 27, pp. 379–423, 623-656.

UNIT 2 TYPES AMPLITUDE MODULATION (AM) TECHNIQUES

1.0 Introduction
2.0 Objectives
3.0 Main Content
3.1. Double Sideband Suppressed Carrier (DSB-SC)/(DSB-WC) Technique
3.1.1 Frequency Spectrum of the DSBSC Wave
3.1.2Time Domain Representation of the DSBSC Wave
3.1.3. Power Computations in the DSBSC Wave
3.1.4. Demodulation of DSBSC Wave
3.2.Single Sideband (SSB or SSB-AM) Technique
3.2.1. Frequency Spectrum of the SSB Wave
3.2.2. Time Domain Representation of the SSB Wave
3.2.3. Power Computations in the SSB Wave
3.3. Vestigial Sideband (VSB) Modulation Technique (VSBMT)
3.3.1. VSB Instantaneous Voltage
3.3.2. VSB Wave Power Computations
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0 References/Further Readings

1.0 INTRODUCTION

In the last unit, we looked at various AM concepts. Now, generally, Amplitude modulation (AM) is a modulation technique used in electronic communication, most commonly for transmitting information via a radiocarrier wave. Usually, in AM, the amplitude (signal strength) of the carrier wave is varied in proportion to the waveform being transmitted. That waveform may, for instance, correspond to the sounds to be reproduced by a loudspeaker, or the light intensity of television pixels. This technique contrasts with frequency modulation, in which the frequency of the carrier signal is varied, and phase modulation, in which its phase is varied. AM was the earliest modulation method used to transmit voice by radio. Today, it is actively in use in many forms of communication e.g. used in portable two-way radios, VHF aircraft radio, and in computer modems (in the form of QAM). AM is often used to refer to mediumwaveAM radio broadcasting. In the context of analog modulation in which the modulation is applied continuously in response to the analog information signal, this unit shall discuss common analog modulation techniques under amplitude modulationincluding double sideband suppressed carrier (DSB-SC)/(DSB-WC) technique, single sideband (SSB or SSB-AM) technique and vestigial sideband (VSB) modulation technique (VSBMT)

6.0 Objective

After going through this unit, you should be able to:

- ✓ explain the concept of double-sideband modulation with carrier (DSB-WC) (used on the AM radio broadcasting band).
- ✓ understand the concept of double-sideband suppressed-carrier transmission (DSB-SC)
- ✓ understand the concept of Single-sideband modulation (SSB, or SSB-AM)
- Understand the concept of vestigial sideband modulation (VSB, or VSB-AM).

3.0. Main Content

3.1. Double Sideband Suppressed Carrier (DSB-SC)/(DSB-WC)Technique

Double-sideband suppressed-carrier transmission (DSB-SC) or wave is transmission in which frequencies produced by amplitude modulation (AM) are symmetrically spaced above and below the carrier frequency and the carrier level is reduced to the lowest practical level, ideally being completely suppressed.

In the DSB-SC modulation, unlike in AM, the wave carrier is not transmitted; thus, much of the power is distributed between the sidebands, which imply an increase of the cover in DSB-SC, compared to AM, for the same power used.DSB-SC transmission is a special case of double-sideband reduced carrier transmission. It is used for radio data systems.

In terms of spectrum shown in Figure 2.1, DSB-SC is basically an AM wave without the carrier, therefore reducing power waste, giving it a 50% efficiency. This is an increase compared to normal AM transmission (DSB), which has a maximum efficiency of 33.333%, since 2/3 of the power is in the carrier which carries no intelligence, and each sideband carries the same information. Single Side Band (SSB) Suppressed Carrier is 100% efficient. DSB-SC is generated by a mixer. This consists of a message signal multiplied by a carrier signal.

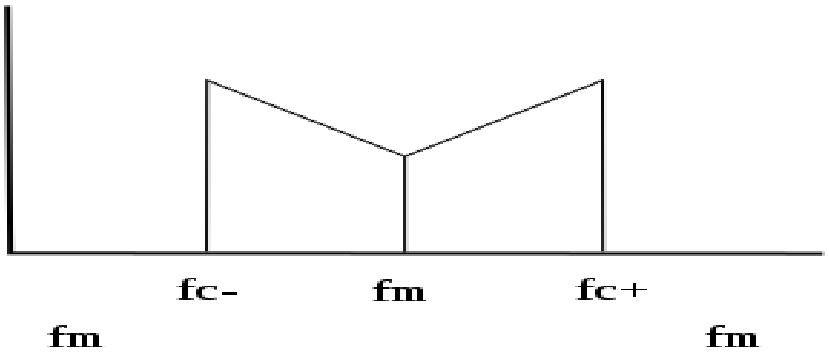


Figure 2.1. Spectrum plot of a DSB-SC signal.

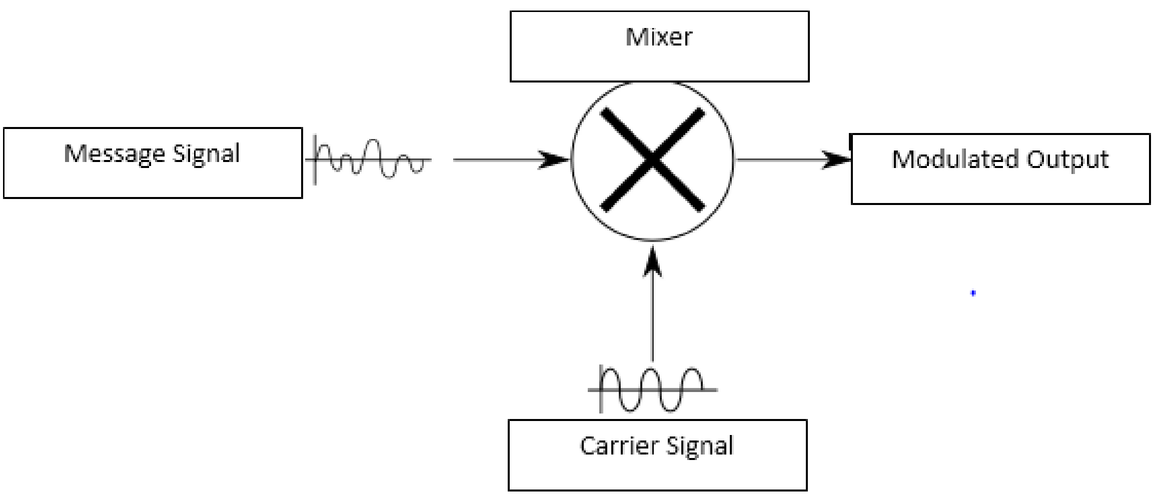


Figure 2.2. DSB-SC generation.

Now, its demodulation is done by multiplying the DSB-SC signal with the carrier signal just like the modulation process. This resultant signal is then passed through a low pass filter to produce a scaled version of original message signal. DSB-SC can be demodulated by a simple envelope detector, like AM, if the modulation index is less than unity. Full depth modulation requires carrier re-insertion.

Recall that in the previous section, we established that the AM signal as derived in the previous section is given by

$$v_{AM} = V_c sin\omega_c t + \frac{mV_c}{2}\cos(\omega_c - \omega_m)\, t - \frac{mV_c}{2}\cos(\omega_c + \omega_t)\, t \qquad (2.1)$$

Thus, the AM signal has three components, namely, unmodulated carrier, LSB and USB. The message be transmitted is present only in LSB and USB. Further, if we consider the power relation given by

$$P_{AM} = P_c\left(1 + \frac{m^2}{2}\right) \qquad (2.2)$$

Therefore, the power required for the carrier component is given by

$$P_c = \frac{P_{AM}}{\left(1+\frac{m^2}{2}\right)} \qquad (2.3)$$

Let the modulation index be unity, i.e., m = I. I, hence, we now have

$$P_c = \frac{2}{3}P_{AM} \qquad (2.4)$$

From Equ.3.30, two-third of total AM power is utilized for the transmission of carrier component, which does not bear any message. This is quiet unacceptable, but a significant saving in power requirement can be achieved by suppressing the carrier before transmission. This is the major reason behind the AM DSBSC technique. Basically, the instantaneous voltage of DSBSC may be related to that of AM as

$$V_{DSBSC} = V_{AM} - V_c sin\omega_c t \qquad (2.5)$$

Substituting for V_{AM} from Equation (2.1), we get

$$V_{AM} = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\, t - \frac{mV_c}{2}\cos(\omega_c + \omega_m)\, t \qquad (2.6a)$$

From Figure 2.2, the DSB-SC is generated by a mixer. This consists of a message signal multiplied by a carrier signal. The mathematical representation of this process is shown in Equ. (2.6a) where the product-to-sum trigonometric identity is used

$$V_m Cos(\omega_m t\,) * V_c Cos(\omega_c t) = \frac{V_m V_c}{2}[\cos\big((\omega_m + \omega_c)t\big) + Cos\big((\omega_m - \omega_c)t\big)] (2.6b)$$

*Message *Carrier = Modulated Signal*

The next question will therefore be why AM is still in use? The significant power saving in case of DSBSC does not come without price. DSBSC technique accordingly adds complexity at the receiving point to recover the message. Thus depending on the application, we can go either for AM or DSBSC. Suppose your application requirement will make the cost of receiver needs to be significantly low, then, AM is preferred, as in the case of AM broadcasting. Alternatively, if the application is meant for point-to-point service, then DSBSC is preferable.

3.1.1. Frequency Spectrum of the DSBSC Wave

Using the AM frequency spectrum of the DSBC, the situation of instantaneous value of DSBSC wave is illustrated in Figure. 2.3. This plot shows how the maximum amplitude of the DSBSC modulated voltage is made to vary with modulating voltage changes. It could be observed that when there is no modulation, the instantaneous value is zero and is expected, since there is no carrier component in this case. From Figure. 2.3, it is possible to write an equation for the peak amplitude of the DSBSC modulated voltage. In this case, we have

$$A = v_m = V_m sin\omega_m t = mV sin\omega_m t \qquad (2.7)$$

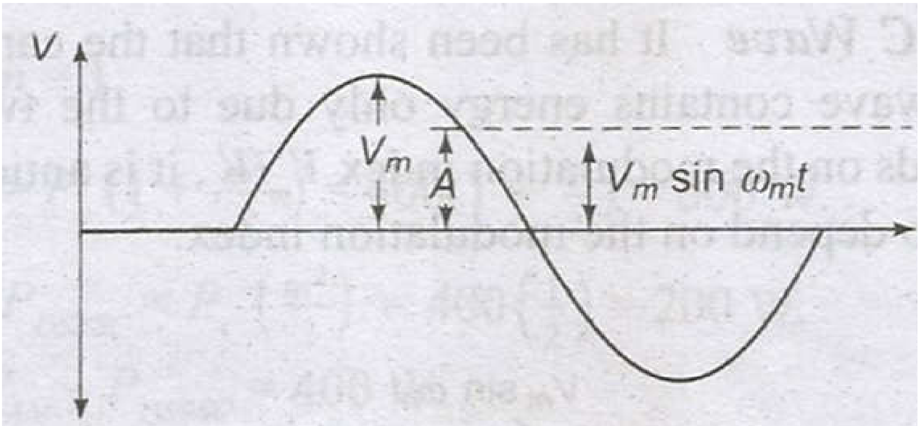


Figure. 2.3. Amplitude of a DSBSC wave.

The instantaneous voltage of the resulting amplitude modulated wave is given by

$$v_{DSBSC} = A sin\theta = A sin\omega_m t sin\omega_c t \qquad (2.8)$$

This equation may be expanded to give

$$v_{DSBSC} = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\, t - \frac{mV_c}{2}\cos(\omega_c + \omega_m)\, t \qquad (2.9)$$

Thus, the equation of DSBSC wave contains two terms, namely, LSB and USB, as discussed earlier. The bandwidth required for DSBSC is twice the frequency of the modulating signal, as in the case of AM. That is,

$$B_{DSBSC} = (f_c + f_m) - (f_c - f_m) = 2f_m \qquad (2.10)$$

The frequency spectrum of DSBSC wave is shown in Figure 2.4 as obtained using the Equ. (2.9). As illustrated, DSBSC consists of two discrete frequencies separated by f/2 and having equal amplitudes.

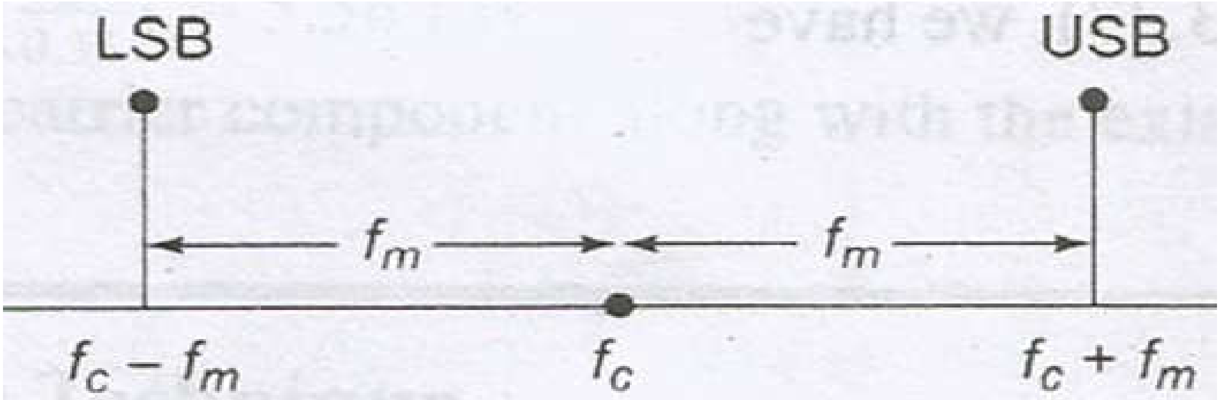


Figure.2.4. Frequency spectrum of the DSBSC wave.

3.1.2. Time Domain Representation of the DSBSC Wave

As you can see, the appearance of the DSBSC wave explains the difficulty in recovering message from it as shown in Figure 2.5for one cycle of the modulating sine wave.

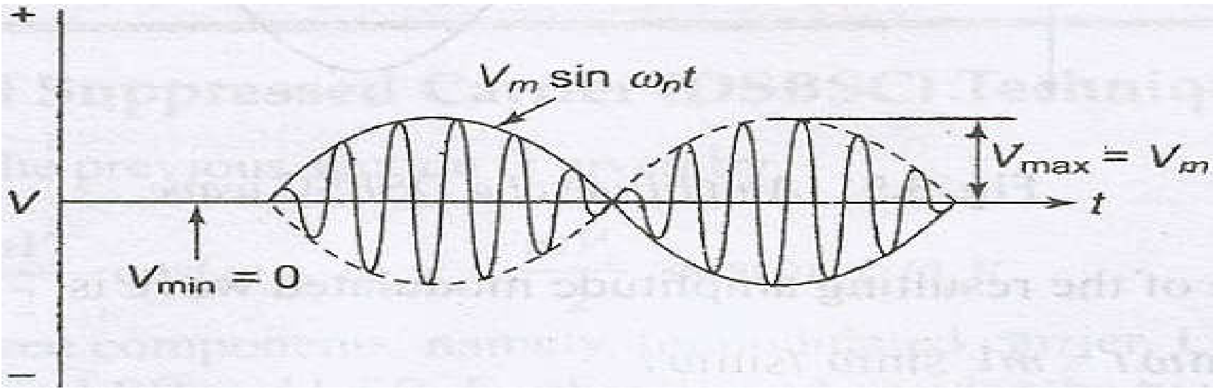


Figure 2.5. Time domain representation of the DSBSC wave.

It shows the amplitude called the top envelope of the DSBSC wave given by the relation $A = V_m Sin\omega_m t$

The maximum negative amplitude or bottom envelope is given by$-A = -V_m Sin\omega_m t$

The modulated wave extends between these two limiting envelopes and has a repetition rate equal to the unmodulated carrier frequency. For better distinction, the bottom envelope is shown as dotted line. The top envelope crosses below the zero reference amplitude value and similarly the bottom envelope crosses above the zero reference amplitude value.

However, in case of AM wave, this will never happen. At the most, the top envelope can touch the zero reference, but cannot cross it. Same thing is true with respect of bottom envelope also. Thus, the information from AM can be recovered uniquely either from top or bottom envelope by a simple envelope detector circuit (assume, as diode rectifier for time being). But this is not the case In case of DSBSC. This is the price we pay by suppressing the carrier. There are ways to overcome this problem for recovering message.

3.1.3. Power Computations in the DSBSC Wave

We have established that the carrier component is suppressed in DSBSC wave. Now, the modulated wave contains energy only due to the two sideband components. Since, the amplitude of the sidebands depends on the modulation index$\frac{V_m}{V_c}$, it is anticipated that the total power in theDSBSC modulated wave will also depend on the modulation index. The total power in the DSBSC modulated wave will be

$$P_{DSBSC} = \frac{V_{LSB}^2}{R} + \frac{V_{USB}^2}{R} \tag{2.11}$$

Where all the voltages are *rms* values and *R* is the resistance in which the power is dissipated.

$$P_{LSB} = P_{USB} = \frac{V_{SB}^2}{R} = \left(\frac{\frac{mV_C}{2}}{\sqrt{2}}\right) \div R = \frac{m^2V_c^2}{8R} = \frac{m^2}{4}\frac{V_c^2}{2R} \tag{2.12}$$

Substituting Equ.(2.11) in (2.12), we have

$$P_{DSBSC} = \frac{m^2}{4}\frac{V_c^2}{2R} + \frac{m^2}{4}\frac{V_c^2}{2R} \tag{2.13}$$

$$P_{DSBSC} = P_c\left(\frac{m^2}{2}\right) \tag{2.14}$$

Equ(2.14) relates the total power in the DSBSC modulated wave to the unmodulated carrier power. It is interesting to know from Equ (3.40) that the maximum power in the DSBSC wave is $P_{DSBSC} = P_c/2$ when *m* = 1. Thus we need only maximum of 50% of unmodulated carrier power for the transmission of DSBSC wave. This is correct also, because, in case of AM wave, two-third of total power is utilized by the carrier component alone and rest one-third by both the sidebands. This one-third constitutes 50% of unmodulated carrier power.

3.1.4. Demodulation of DSBSC Wave

Demodulation is done by multiplying the DSB-SC signal with the carrier signal just like the modulation process. This resultant signal is then passed through a low pass filter to produce a scaled version of original message signal. DSB-SC can be demodulated by a simple envelope detector, like AM, if the modulation index is less than unity. Full depth modulation requires carrier re-insertion.

Modulated Signal ∗ carrier = Original message

$$\frac{V_mV_c}{2}\left[\cos\left((\omega_m + \omega_c)t\right) + Cos\left((\omega_m - \omega_c)t\right)\right] * V_c{}'Cos(\omega_m t\,)$$
$$= \left(\tfrac{1}{2}V_cV_c{}'\right)V_mCos(\omega_m t\,) + \tfrac{1}{4}V_cV_c{}'V_m[Cos\left((\omega_m + 2\omega_c)t\right) + \cos\left(\left(\left((\omega_m - 2\omega_c)t\right)\right]\right. \tag{2.15}$$

Equ. 2.15/ 3.41 shows that by multiplying the modulated signal by the carrier signal, the result is a scaled version of the original message signal plus a second term. Since $\omega_c \geq \omega_m$,this second term is much higher in frequency than the original message. Once this signal passes through a low pass filter, the higher frequency component is removed, leaving just the original message.

For demodulation, the demodulation oscillator's frequency and phase must be exactly the same as modulation oscillators, otherwise, distortion and/or attenuation will occur.To see this effect, take the following conditions:

- Message signal to be transmitted:f_t
- Modulation (carrier) signal: $Cos(\omega_c t\ )$
- Demodulation signal (with small frequency and phase deviations from the modulation signal): $V_c{}'Cos[(\omega_c + \Delta\omega)t +\ \theta]$

The resultant signal can then be given by

$$f(t) * V_c Cos(\omega_c t) * V_c{}^{'} Cos[(\omega_c + \Delta\omega)t +\ \theta] = \left(\frac{1}{2} V_c V_c{}'\right) V_c V_c{}' f(t)[Cos\big((\Delta\omega . t +\ \theta)\big) + \left(\frac{1}{2} V_c V_c{}'\right) V_c V_c{}' f(t) \cos\big[\,((2\omega_c + \Delta\omega)t + \theta)\big] \rightarrow \left(\frac{1}{2} V_c V_c{}'\right) f(t) Cos[(\Delta\omega + \theta)t \qquad (2.16)$$

From Equ (2.16), the $Cos[(\Delta\omega + \theta)t]$ terms results in distortion and attenuation of the original message signal.

Example 20.

A 400W carrier is amplitude modulated to a depth of 100%, Calculate the total power in case of AM and DSBSC techniques. How much power saving (in W) is achieved for DSBSC? If the depth of modulation is changed to 75%, then how much power (in W) is required for transmitting the DSBSC wave? Compare the powers required for DSBSC in both the cases and comment on the reason for change in the power levels.

Solution
Case 1:
Given, P_c = 400 W and m = 1.
Total power in AM,$P_{AM} = P_c\left(1 + \frac{m^2}{2}\right) = 400\left(1 + \frac{1}{2}\right) = 600\ W.$
Total power in DSBSC, *P* DSBSC $P_{DSBSC} = P_c\left(\frac{m^2}{2}\right) = 400\left(\frac{1}{2}\right) = 200\ W.$
Power saving (in W) = $P_{AM} - P_{DSBSC} = 400\ W$
Thus we require only 200 W in case of DSBSC which is one-third of total AM power. This is the gain we achieve using DSBSC.

Case 2:
Given, $P_c = 400\ W$ and m = 0.75
Total power in DSBSC, $P_{DSBSC} = P_c\left(\frac{m^2}{2}\right) = 400\left(\frac{(0.75)^2}{2}\right) = 112.5\ W.$
The power required in this case is lower than m = 1 case. This infers that the total power in DSBSC also depends on 'the depth of modulation. It will be maximum, that is, one-third of total AM power when m = 1 and less for *m < 1*.

Example 22

A DSBSC transmitter radiates 1 kW when the modulation percentage is 60%. How much of carrier power (in kW) is required if we want to transmit the same message by an AM transmitter?

Solution

Given$P_{DSBSC} =$ 1KW and m = 0.6.

Carrier power, $P_c = P_{DSBSC}\left(\frac{2}{M^2}\right) = 1\left(\frac{2}{0.36}\right) = 5.56\ KW.$

We require 5.56 kW to transmit the carrier component along with the existing 1 kW for the sidebands when *m*=0.6.

3.2. Single Sideband (SSB or SSB-AM) Technique

In radio communications, single-sideband modulation (SSB) or single-sideband suppressed-carrier modulation (SSB-SC) is a type of modulation, used to transmit information, an audio signal, by radio waves. It is a refinement of amplitude modulation which uses transmitter power and bandwidth more efficiently. Amplitude modulation produces an output signal that has twice the bandwidth of the original baseband signal. Single-sideband modulation avoids this bandwidth doubling, and the power wasted on a carrier, at the cost of increased device complexity and more difficult tuning at the receiver.

We must understand that the basic version of AM is modified by suppressing the carrier component to yield DSBSC technique. The bandwidth requirement of DSBSC is still same as that of AM. Both the sidebands, namely, LSB and USB carry the same information. Hence, saving in bandwidth can be achieved by suppressing one of the sidebands. This thought process led to the development of another variant of AM, on top of DSBSC termed as single sideband suppressed carrier (SSBSC) technique.
This is generally referred to as SSB. In this Unit, both SSB can be used interchangeably with SSBSC. Since only one of the sidebands is selected for transmission, SSB needs a bandwidth equal to that of message. That is,

$$B_{SSB} = f_m \tag{2.17}$$

Where $f_m$is maximum frequency component in the message.
The DSBSC signal is given by

$$v_{DSBSC} = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t - \frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \tag{2.18}$$

If LSB is chosen for transmission in case of SSB, then

$$v_{SSB} = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t \tag{2.19}$$

Alternatively, if USB is chosen for transmission, then

$$v_{SSB} = -\frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \tag{2.20}$$

Compared to AM and DSBSC, SSB significantly saves power, since carrier and one sideband are suppressed and saves bandwidth, since only one sideband is chosen for transmission. Then the next question is why not use only SSB? The answer is same as in the case of existence of AM, even after the development of DSBSC technique. The SSB technique further complicates the receiver structure to recover message. As will be explained later, an equally important limitation of SSB is the practical difficulty in suppressing the unwanted sideband, since it lies close to the wanted sideband. Therefore still all the three versions of AM, namely, AM, DSBSC and SSB coexist in the analog communication field.

3.2.1. Frequency Spectrum of the SSB Wave

One way of viewing SSB is DSBSC followed by bandpass filtering, as illustrated in Figure2.6. The mathematical treatment here follows this assumption. The situation of instantaneous value of SSB wave is same as in DSB, illustrated in Figure 2.5, which shows how the DSBSC modulated voltage is made to vary with modulating voltage changes. From Figure 2.5. It is possible to write an equation for the amplitude of the DSBSC modulated voltage.

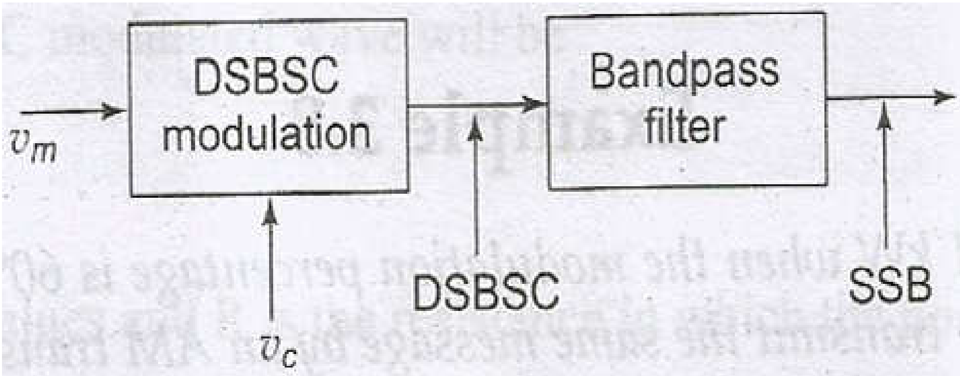


Figure 2.6. Block diagram representation of SSB generation by band pass filtering

$$v_{DSBSC} = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t - \frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \qquad (2.21)$$

Now, for generating the SSB, the DSBSC is passed through the bandpass filter. Depending on the cut-off frequencies, either LSB or USB comes out of the bandpass filter if the cut-off frequencies are (f_c-f_m) and f_c, then LSB is chosen for transmission and instantaneous voltage of SSB signal is given by

$$v_{SSB} = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t \qquad (2.22)$$

Alternatively, if the cut-off frequencies are *f.* and if. + “1", the instantaneous voltage of the USB chosen for transmission is given by

$$v_{SSB} = \frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \qquad (2.23)$$

It has thus been shown that the equation of SSB wave contains one term, that is, either LSB or USB. The bandwidth required for SSB is the frequency of the modulating signal. That is,

$$B_{SSB} = (f_c\text{-}f_m) - f_c = f_c - (f_c\text{-}f_m) = f_m \qquad (2.24)$$

The frequency spectrum of SSB wave is shown in Figure. 2.7. Using the equations of SSB as illustrated, SSB, consists of one discrete frequency either at f_c-f_m or $f_c + f_m$.

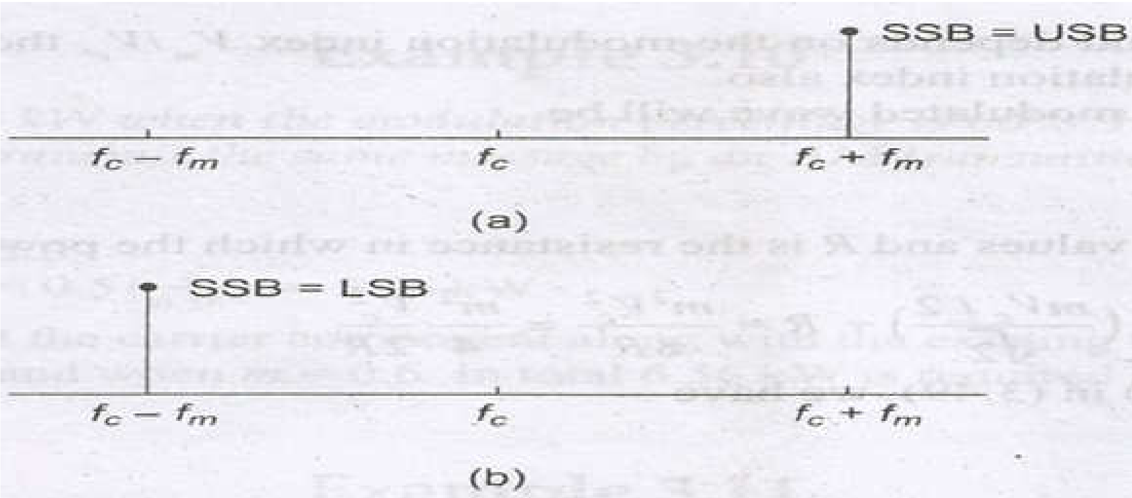


Figure. 2.7. Frequency spectrum of the SSB wave. Spectrum for (a) SSB = USB, and (b) SSB = LSB.

3.2.2. Time Domain Representation of the SSB Wave

Figure 2.8 shows the time domain representation of SSB wave for one cycle of message signal. The modulated wave will have only one sine wave. The only wave to distinguish is to compare with carrier signal. Its frequency will be either lower or more than carrier frequency by an amount of modulating signal frequency. The envelope of SSB does not contain message and hence a simple envelope detector circuit is not useful for recovering the message. This is the price we pay by suppressing the carrier and one of the sidebands. Of course, here also, there are ways to overcome this problem to recover message.

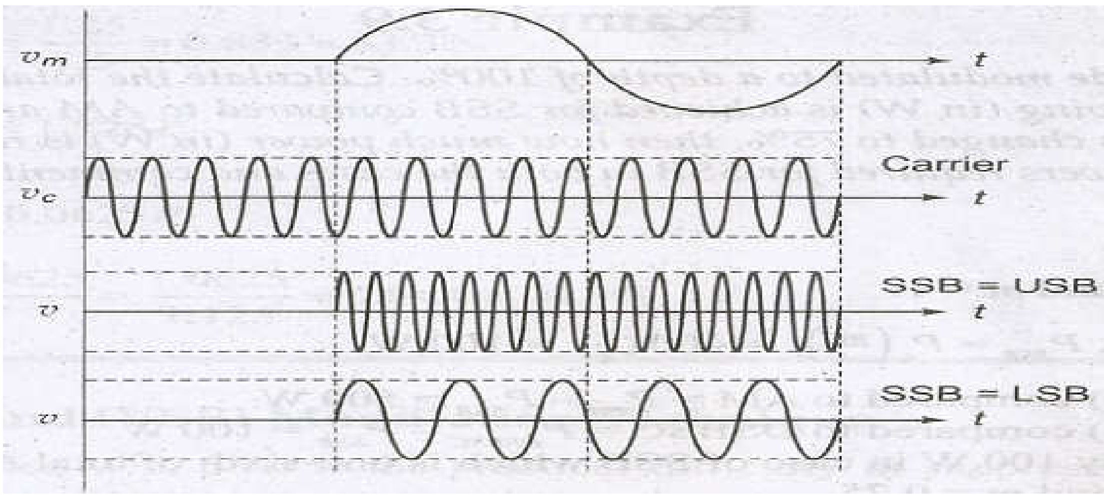


Figure 2.8. Time domain representation of the SSB wave.

3.2.3. Power Computations in the SSB Wave

Form the discussions above, It has been shown that the carrier component and one sideband are suppressed in the SSB wave. The modulated wave contains energy only due to one sideband component. Since amplitude of the sideband depends on the modulation index $\frac{V_m}{V_c}$the total power in the modulated wave will depend on the modulation index also. The total power in the SSB modulated wave will be

$$P_{SSB} = \frac{V_{LSB}^2}{R} = \frac{V_{USB}^2}{R} \tag{2.25}$$

Where all the voltages are *rms* values and *R* is the resistance in which the power is dissipated.

$$P_{LSB} = P_{USB} = \frac{V_{SB}^2}{R} = \left(\frac{\frac{mV_C}{2}}{\sqrt{2}}\right) \div R = \frac{m^2}{4}\frac{V_c^2}{2R} \tag{2.26}$$

Substituting Equ. (2.25) in (2.26), we have

$$P_{SSB} = \frac{m^2}{4}\frac{V_c^2}{2R} \tag{2.27}$$

$$P_{SSB} = \frac{m^2}{4} \tag{2.28}$$

Equ. (2.28) relates the total power in the SSB modulated wave to the unmodulated carrier power. It is interesting to know from Equ.(2.28) that the maximum power in the SSB wave is $P_{SSB} = \frac{\text{Pc}}{4}$when *m = 1*. Thus, we need only maximum of 25% of unmodulated carrier power for the transmission of SSB wave. This is correct also, because, in case of SSB wave, one-sixth of the total power is utilized by the sideband and this constitutes 25% of unmodulated carrier power.

Example 23.

A 400 W carrier is amplitude modulated to a depth of 100%. Calculate the total power in case of SSB technique. How much power saving (in W) is achieved for SSB compared to AM and DSBSC of Example 20? If the depth of modulation is changed to 75%, then how much power (in W) is required for transmitting the SSB wave? Compare the powers required for SSB in both the cases and comment on the reason for change in the power levels.

Solution

Case 1 Given, P_c = 400 W and m = 1.

Total power in SSB, P_{SSB}= $P_c\left(\frac{m^2}{4}\right) = 400\left(\frac{1}{2}\right) = 100\,W$.

Power saving (in W) compared to AM = P_{AM}- P_{SSB} = 500 W.

Power saving (in W) compared to DSBSC = P_{DSBSC} - P_{SSB} = 100 W.

Thus, we require only 100 W in case of SSB which is one-sixth of total AM power.

Case 2 Given, P_c = 400 W and m = 0.75

Total power in SSB, P_{SSB}= $P_c\left(\frac{m^2}{4}\right) = 400\left(\frac{(0.75)^2}{4}\right) = 56.25\ W$

The power required in this case is lower than $m = 1$ *case*. This infers that the total power in SSB also depends on the depth of modulation. It will be maximum, that is, one-sixth of total AM power when m = 1 and less for $m<1$.

Example 24

A SSB transmitter radiates 0.5 kW when the modulation percentage is 60%. How much of carrier power (in kW) is required if we want to transmit the same message by an AM transmitter?

Solution

Given, P_{SSB} = 0.5 kW and m = 0.6.

Carrier power,$P_c = P_{SSB}\left(\frac{4}{m^2}\right) = 0.5\left(\frac{4}{0.36}\right) = 5.56\ KW$..

We require 5.56 kW to transmit the carrier component along with the existing 0.5 kW for one sideband and 0.5 kW more for another sideband when m = 0.6. In total 6.56 kW is required by the AM transmitter.

Example 25

Calculate the percentage power saving when the carrier and one of the sidebands are suppressed in an AM wave modulated to a depth of (a) 100 percent, and (b) 50 percent.

Solution

(a) $P_{AM} = P_c\left(1+\frac{m^2}{2}\right) = P_c\left(1+\frac{m^2}{2}\right) = 1.5P_c$

$P_{SSB} = P_c\left(\frac{m^2}{4}\right) = P_c\left(\frac{1^2}{4}\right) = 0.25P_c$

$Saving = \frac{1.5-0.25}{1.5} = \frac{1.25}{1.5} = 0.833 = 83.3\%$

(b) $P_{AM} = \left(1+\frac{0.5^2}{2}\right) = 1.125P_c$

$P_{SSB} = P_c\left(\frac{0.5^2}{4}\right) = 0.0625P_c$

$Saving = \frac{1.125-0.0625}{1.125} = \frac{1.0625}{1.125} = 0.944 = 94.4\%$

3.3. Vestigial Sideband (VSB) Modulation Technique (VSBMT)

Vestigial sideband (VSB) is a type of amplitude modulation technique that encodes data by varying the amplitude of a single carrier frequency. Portions of one of the redundant sidebands are removed to form a vestigial sideband signal - so-called because a vestige of the sideband remains.In AM, the carrier itself does not fluctuate in amplitude. Instead, the modulating data appears in the form of signal components at frequencies slightly higher and lower than that of the carrier. These components are called sidebands.

The lower sideband (LSB) appears at frequencies below the carrier frequency; the upper sideband (USB) appears at frequencies above the carrier frequency.

The actual information is transmitted in the sidebands, rather than the carrier; both sidebands carry the same information. Because LSB and USB are essentially mirror images of each other, one can be discarded or used for a second channel or for diagnostic purposes.

Again, the VSB transmission is similar to single-sideband (SSB) transmission, in which one of the sidebands is completely removed. In VSB transmission, however, the second sideband is not completely removed, but is filtered to remove all but the desired range of frequencies.The main limitation associated with SSB is the practical difficulty in suppressing the unwanted sideband frequency components. It was observed in practice that such a process results in eliminating even some portion of the wanted sideband.

This is because, in many cases the message has information starting from zero frequency and spreads up to a maximum of f_m Hz. In such a scenario the first wanted and unwanted frequency components lie very close to each at the carrier frequency f_c. Therefore an attempt to attenuate unwanted component will in turn leads to attenuation of wanted component. One way to compensate for this loss is to allow a vestige or trace or fraction of unwanted sideband along with the wanted sideband. This thought process lead to the development of AM variant known as vestigial sideband suppressed carrier (VSBSC) technique. VSBSC is more commonly termed as VSB representing vestigial sideband and suppressed carrier as implied.

3.3.1. VSB Instantaneous Voltage Expressions

Now, the DSBSC signal is given by Equ 2.29.

$$v_{DSBSC} = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t - \frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \tag{2.29}$$

If LSB is wanted sideband in case of VSB, the instantaneous voltage of the VSB signal may be expressed as

$$v_{VSB} = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t + F\left(-\frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t\right) \tag{2.30}$$

Alternatively, if USB is wanted sideband, the instantaneous voltage of VSB may be given by

$$v_{VSB} = -\frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t + F\left(\frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t\right) \tag{2.31}$$

Where$f$ represents the fraction. The power and bandwidth requirements in case of VSB will be slightly m than SSB, but less than DSB. The frequencyspectrum andtime domain representation of the VSB Wave have their mathematical relationships which could be directly derived from DSBSC. Let us quickly look at the power computations in VSB wave.

3.3.2VSB Wave PowerComputations

In the VSB wave, it contains one sideband completely and a vestige of other sideband. The modulated wave contains energy due to these two components. Since amplitude of the sidebands depends on the modulation index$\frac{V_m}{V_c}$the total power in the modulated wave will depend on the modulation index also. The total power in the DSBSC modulated wave will be

$$P_{DSBSC} = \frac{V_{LSB}^2}{2} + \frac{V_{USB}^2}{2} \tag{2.32}$$

Where all the voltages are *rms* values and *R* is the resistance in which the power is dissipated.

$$P_{LSB} = P_{USB} = \frac{V_{SB}^2}{R} = \left(\frac{\frac{mV_C}{2}}{\sqrt{2}}\right)/R = \frac{m^2V_c^2}{8R} = \frac{m^2}{4}\frac{V_c^2}{2R} \tag{2.33}$$

Substituting these equations in the total power equation, we have

$$P_{DSBSC} = \frac{m^2}{4}\frac{V_c^2}{2R} + \frac{m^2}{4}\frac{V_c^2}{2R} \tag{2.34}$$

If LSB is wanted sideband in VSB, then

$$P_{VSB} = \frac{m^2}{4}P_c + F\left(\frac{m^2}{4}P_c\right) \tag{2.35}$$

Alternatively, if USB is wanted sideband in VSB, then

$$P_{VSB} = F\left(\frac{m^2}{4}P_c\right) + \frac{m^2}{4}P_c \tag{2.36}$$

Equ. (2.36) relates the total power in the VSB modulated wave to the unmodulated carrier power, It is interesting to know from this that the maximum power in the VSB wave is$P_{VSB} = {}^{P_c}/_4 + f({}^{P_c}/_4)$when *m =1*. Thus, we need only maximum of25% to 50% of unmodulated carrier power for the transmission of VSB wave. This is correct also, because, in case of VSB wave, one-sixth of total power is utilized by one sideband and a fraction of one-sixth for the transmission of the vestige.

Always remember that the limitation of single-sideband modulation being used for voice signals and not available for video/TV signals is what led to the usage of vestigial sideband. A vestigial sideband (inradio communication) is a sideband that has been only partly cut off or suppressed. Television broadcasts (in analog video formats) use this method if the video is transmitted in AM, due to the large bandwidth used. It may also be used in digital transmission, such as the ATSC standardized8VSB. The Milgo 4400/48 modem (around 1967) used vestigial sideband and phase-shift keying to provide 4800 bit/s transmission over a 1600 Hz channel.The broadcast or transport channel for TV in countries that use NTSC or ATSC has a bandwidth of 6 MHz to conserve bandwidth, SSB would be desirable, but the video signal has significant low-frequency content (average brightness) and has rectangular synchronizing pulses. The engineering compromise is vestigial-sideband transmission. In vestigial sideband, the full upper sideband of bandwidth W_2 = 4.75 MHz is transmitted, but only W_1 = 1.25 MHz of the lower sideband is transmitted, along with a carrier. This effectively makes the system AM at low modulation frequencies and SSB at high modulation frequencies. The absence of the lower sideband components at high frequencies must be compensated for, and this is done by the RF and IF filters

Example 26

A VSB transmitter that transmits 25% of the other sideband along with wanted sideband radiates 0.625 kW when the modulation percentage is 60%. How much of carrier power (in kW) is required if we want to transmit the same message by an AM transmitter?

Solution
Given, P_{VSB} = 0.625 kW and m = 0.6.
Carrier power, $P_c = P_{VSB}\left(\frac{4}{1.25^{*}m^2}\right) = 0.625\left(\frac{4}{1.25^{*}0.36}\right) = 5.56\ kW.$
We require 5.56 kW to transmit the carrier component along with the existing 0.625 kW for one side band and 0.375 kW more for rest of the other sideband when m = 0.6. In total 6.56 kW is required by the AM transmitter.

4.0. Conclusion
This unit has explained the concept of double-sideband modulation with carrier (DSB-WC) while presenting a simplified perspective on the concept of double-sideband suppressed-carrier transmission (DSB-SC). We have equally gained insight in the concept of Single-sideband modulation (SSB, or SSB-AM) and have understood vestigial sideband modulation.

5.0. Summary
- Double-sideband suppressed-carrier transmission (DSB-SC) or wave is transmission in which frequencies produced by amplitude modulation (AM) are symmetrically spaced above and below the carrier frequency and the carrier level is reduced to the lowest practical level.
- DSBSC wave contains two terms, namely, LSB and USB.
- Demodulation is done by multiplying the DSB-SC signal with the carrier signal just like the modulation process.
- For demodulation, the demodulation oscillator's frequency and phase must be exactly the same as modulation oscillators; otherwise, distortion and/or attenuation will occur.
- Vestigial sideband (VSB) is a type of amplitude modulation technique that encodes data by varying the amplitude of a single carrier frequency.
- Compared to AM and DSBSC, SSB significantly saves power, since carrier and one sideband are suppressed and saves bandwidth, since only one sideband is chosen for transmission.

6.0 Tutor marked Assignments
1. How do you distinguish between analog and digital communication?
2. Define amplitude modulation?
3. Write the expression for the instantaneous voltage of AM wave?
4. Define modulation index of amplitude modulation?
5. Derive the expression for the total modulation index in case of modulation by several sine waves?
6. A 360-W carrier is simultaneously, modulated by two audio waves with modulation percentages of 55 and 65, respectively. What is the total sideband power radiated?

7. A transistor class C amplifier has maximum permissible collector dissipation of 20 Wand a collector efficiency of 75 percent. It is to be collector-modulated to a depth of 90 percent, (a) Calculate (i) the maximum unmodulated carrier power and (ii) the sideband power generated. (b) If the maximum depth of modulation is now restricted to 70 percent, calculate the new maximum sideband power generated.
8. Derive the expression for the instantaneous voltage of DSBSC wave?
9. What is the difference between SSB and DSBSC wave?
10. Derive the expression for the instantaneous voltage of SSB wave?
11. Derive the expression for the total power in case of SSB wave?

7.0. Reference/Further Readings

1. Roland Priemer (1991). Introductory Signal Processing. World Scientific. p. 1. ISBN 9971509199.

2. T. H. Wilmshurst (1990). Signal Recovery from Noise in Electronic Instrumentation (2nd ed.). CRC Press. pp. 11 ff. ISBN 0750300582.

3. Proakis, John G.; Manolakis, Dimitris G. (2007-01-01). Digital Signal Processing. Pearson Prentice Hall. ISBN 9780131873742.

4. Paul Horowitz; Winfield Hill (2015). The Art of Electronics. Cambridge University Press. ISBN 9780521809269.

UNIT 3 **AMPLITUDE MODULATED SIGNAL GENERATION**

1.0 Introduction
2.0 Objectives
3.0 Main Content
3.1. Generation of AM Signal
3.1.1. AM Generation Using Analog Multiplier
3.1.2. AM Generation Using a Nonlinear Resistance Device
3.2. Generation of DSBSC Signal Using a Balanced Modulator
3.3. Generation of SSB Signal
3.3.1. Analog multiplier
3.3.2. Filter Method
3.3.3. Phase Shift Method
3.3.4. Weaver Phase Shift Method
3.4 Generation of VSB Signal
3.4.1. Using Analog Multiplier
3.4.2. Using the Filter Method
4.0. Conclusion
5.0. Summary
6.0 Tutor-Marked Assignment (TMA)
7.0. References/Further Readings

1.0 INTRODUCTION

In this Section, we shall discuss the principles of amplitude modulation generation.

3.0 Objectives

4.0 After going through this unit, you should be able to

- ✓ explain the principle of AM signal generation using various techniques
- ✓ distinguish between the various methods of AM signal generation
- ✓ explain DSBSC using balanced modulator.
- ✓ analyze VSB using filter and analog multiplier methods

2.0 MAIN CONTENT

3.1. Generation of AM Signal

The following are the key methods of AM signal generation

3.1.1. AM Generation Using Analog Multiplier

Figure 3.1 illustrates a conceptual way to realize the generation of AM signal with the help of an analog multiplier and a summer connected together. It shows a block diagram representing the generation ofAM signal using analog multiplier.

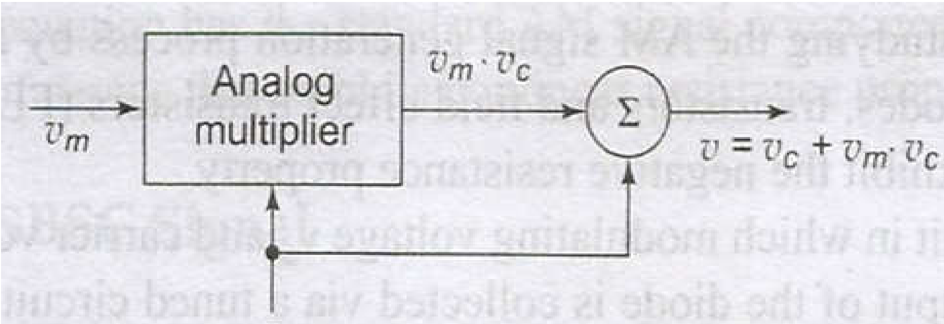


Figure 3.1. Block diagram representation of generation of AM signal using analog multiplier.

The output of the analog multiplier is given by

$$v = v_m v_c = V_m sin\omega_m t V_c sin\omega_c t = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\, t - \frac{mV_c}{2}\cos(\omega_c + \omega_m)\, t \qquad (3.1)$$

Thus, at the output of the analog multiplier, we have two sidebands.

By adding the unmodulated carrier component to this, we now get the requisite AM signal and is given by

$$v = v_c + v_m v_c = V_c sin\omega_c t = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\, t - \frac{mV_c}{2}\cos(\omega_c + \omega_m)\, t.$$

3.1.2. AM Generation Using a Nonlinear Resistance Device

The relationship between voltage and current in a linear resistance is given by

$$i = bv \qquad (3.2)$$

Where$b$ is some constant of proportionality. If the above equation refers to a resistor, then b is obviously its conductance.In a nonlinear resistance, the current is still to a certain extent proportional to the applied voltage, but no longer directly as before. If the curve of current versus voltage is plotted, as in Figure 3.2, it is found that there is now some curvature in it. The previous linear relation seems to apply to certain point, after which current increases more (or less) rapidly with voltage. Whether the increase is more or less rapid depends on whether the device begins to saturate, or else some sort of avalanche current multiplication takes place. Current now becomes proportional not only to voltage but also to the square, cube and higher powers of voltage.This nonlinear relation is most conveniently expressed as

$$\mathrm{i = a + bv + cV^2 + dV^3 + higher\ powers} \qquad (3.3)$$

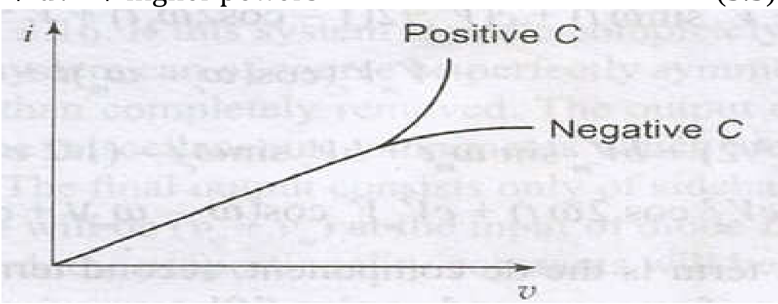


Figure.3.2. Nonlinear resistance characteristics.

The reason that the initial portion of the graph is linear is simply that the coefficient c is much smaller than b. A typical numerical equation might well be something like $i = 5 + 15v + 0.2v^{2,}$ in which case curvature is insignificant until v equals at least 3. Therefore, c in practical nonlinear resistances is much greater than d, which is in turn larger than the constants preceding the higher-power terms. The devices like diodes, transistors and field effect transistors (FET) can be biased with suitable' voltage to constrain them to exhibit the negative resistance property. Figure 3.3shows the circuit in which modulating voltage V_m and carrier voltage V_{mc} are applied in series at the input of the diode. The output of the diode is collected via a tuned circuit tuned to the carrier frequency with bandwidth of twice the message bandwidth.

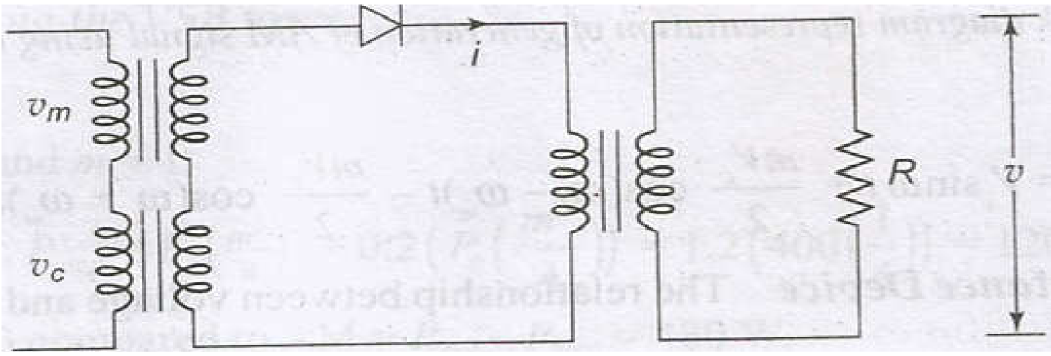


Figure 3.3. Generation of AM signal using nonlinear resistance characteristics of diode.

The diode is biased such that it exhibits the negative resistance property. Under this condition, its output current is given by Equ 3.4

$$i = \alpha + b(v_m + v_c) + c(v_m + v_c)^2 = \alpha + b(v_m + v_c) + c(v^2_m + v^2_c + 2v_mv_c) \qquad (3.4)$$

Substituting for $v_m = V_m sin\omega_m t$ and $v_c = V_c sin\omega_m t$,we get,

$$i = \alpha + b(V_m sin\omega_m t + V_c sin\omega_c t) + c(V_m^2 sin^2\omega_m t + V^2_c\, sin^2\omega_c t + 2V_m v_c sin\omega_m t sin\omega_c t) \qquad (3.5)$$

Using the trigonometric expressions,

Sinx Siny = 1/2 [Cos (x - y) - Cos(x + y)] and $sin^2 x = 1/2(1 - cos2x)$.

We get,

$$i = \alpha + b(V_m sin\omega_m t + V_c sin\omega_c t) + c(V_m^2/2(1 - cos2\omega_m t) + V_c^2/2(1 - cos2\omega_c t) + V_m Vc(cos(\omega_c - \omega_m)t + cos(\omega_c + \omega_m))t \qquad (3.6)$$

$$i = (\alpha + cV_m^2/2 + cV_c^2/2) + bV_m sin\omega_m t + bV_c sin\omega_c t - (1/2cV^2_m cos2\omega_m t + 1/2CV_c^2 cos2\omega_c t) + cV_m V_c\, cos(\omega_c - \omega_m)t + cV_m V_c\, cos(\omega_c + \omega_m)t \qquad (3.7)$$

In the above equation the first term is the D.C component, second term is message, third term is carrier, fourth term contains the harmonics of message and carrier, fifth term represents the lower sideband and sixth term represents the upper sideband. The requisite AM components can be selected by using the tuning circuit that resonates at the carrier frequency with a bandwidth equal to twice the message bandwidth. At the output of the tuning circuit the current will be

$i = bV_c sin\omega_c t + cV_m V_c\, cos(\omega_c - \omega_m)t - cV_m V_c\, cos(\omega_c + \omega_m)t$ (3.8)

If R is the load resistance, then the amplitude modulated voltage is given by

$$v = iR = V_c sin\omega_c t + cRV_c \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t - cRV_c \frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \quad (3.9)$$

$$v = iR = V_c sin\omega_c t + c'\frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t - c'\frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \quad (3.10)$$

Where$c' = cRV_c$.

The above equation has the standard AM signal components. In this way, we can generate me AM signal with the help of device that exhibits nonlinear resistance property.

3.2. Generation of DSBSC Signal Using a Balanced Modulator

From our previous studies, we already know that a non-linear resistance or non-linear device may be used to produce Amplitude Modulation i.e. one carrier and two sidebands. However, a DSB-SC signal contains only two sidebands. Thus, if two non-linear devices such as diodes, transistors etc. are connected in a balanced mode so that they suppress the carriers of each other, then only sidebands are left and a DSB-SC signal is generated. Therefore, a balanced modulator may be defined as a circuit in which two non-linear devices are connected in a balanced mode to produce a DSB-SC signal.

This section shall discuss a balanced modulator circuit using diodes. Figure 3.4a shows the balanced modulator using diodes as non-linear device. The modulating signal *x(t)* is applied equally with 180^o phase reversal at the inputs of both the diodes through the input center tapped transformer.The carrier is applied to the center tap of the secondary.

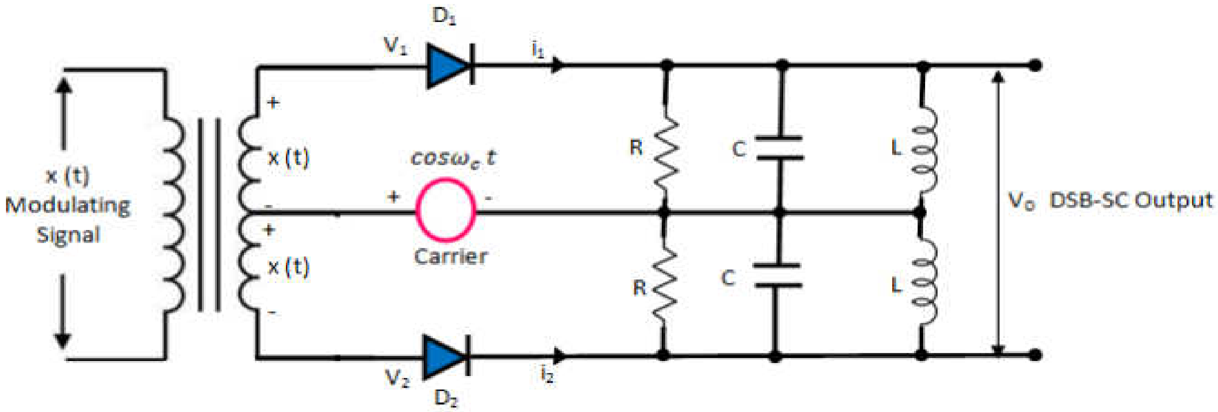


Figure3.4a. DSB-SC signal balance modulator.

From Figure 3.4a, a balanced modulator can be constructed using the non-linear devices like diodes and transistors. The balanced modulator using the diodes is given in Figure 3.4a. The diodes use its nonlinear resistance property for generating modulated signals. Both the diodes receive the carrier voltage in phase; whereas the modulating voltage appears 1800 out of phase at the input of diodes, since they are at the opposite ends of a center-tapped transformer. The modulated output currents of the two diodes are combined in the center-tapped primary of the output transformer. They therefore subtract, as indicated by the direction of the arrows in the Figure3.4a. If this system is made completely symmetrical, the carrier frequency will be completely canceled. No system can of course be perfectly symmetrical in practice, so that the carrier will be heavily suppressed rather than completely removed. The output of the balanced modulator contains the two sidebands and some of the miscellaneous components which are taken care of by tuning the output transformer's secondary winding. The final output consists only of sidebands.

As indicated, the input voltage will be$(V_c + V_m)$ at the input of diode D_i and$(V_c + V_m)$ at the input of diodeD_2. If perfect symmetry is assumed, the proportionality constants will be the same for both diodes and may be called (1, b, and c) as before.

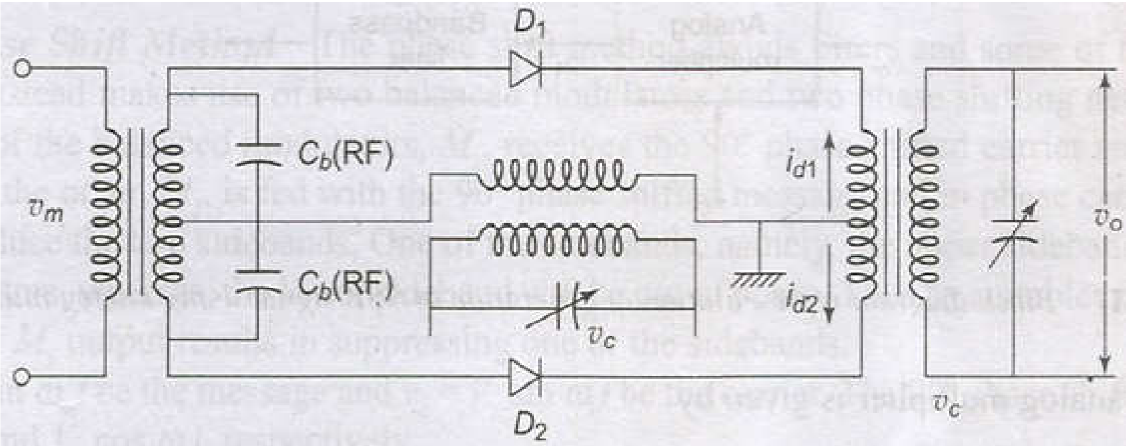


Figure 3.4b. Generation of DSBSC signal using balanced modulator based on nonlinear resistance characteristics of diode.

From Figure 3.4b, the two diode output currents will be

$i_{d1} = \alpha + b(v_c + v_m) + c(v_c + v_m)^2$ (3.11)

$i_{d1} = \alpha + bv_c + bv_m + cv_c^2 + cv_m^2 + 2cv_m v_c$ (3.12)

$i_{d2} = \alpha + b(v_c - v_m) + c(v_c - v_m)^2$ (3.13)

$i_{d2} = \alpha + bv_c - bv_m + cv_c^2 + cv_m^2 - 2cv_m v_c$ (3.14)

As previously indicated, the primary current is given by the difference between the individual diode output currents. Thus

$i_i = i_{d1} - i_{d2} = 2bv_m + 4cv_m v_c$ (3.15)

Substituting for $v_m$and v_c and simplifying we get

$i_i = 2bV_m sin\omega_m t + 4c\frac{mV_c}{2}\cos(\omega_c - \omega_m)\, t - 4c\frac{mV_c}{2}\cos(\omega_c + \omega_m)\, t$ (3.16)

The output voltage $V_0$is proportional to this primary current.

Let the constant of proportionality beα then

$v_0 = \alpha i_1 = 2b\alpha V_m sin\omega_m t + 4\alpha c\frac{mV_c}{2}\cos(\omega_c - \omega_m)\, t - 4\alpha c\frac{mV_c}{2}\cos(\omega_c + \omega_m)\, t$ (3.17)

Let, $P2\alpha bV_m$and $Q = 2\alpha c\frac{mV_c}{2}$, Then

$v_0 = Psin\omega_m t + 2Qcos(\omega_c - \omega_m)t - 2Qcos(\omega_c + \omega_m)t$ (3.18)

This equation shows that the carrier has been canceled out, leaving only the two sidebands and the modulating frequencies. The tuning of the output transformer will remove the modulating frequencies from the output.

$$v_0 = 2Qcos(\omega_c - \omega_m)t - 2Qcos(\omega_c + \omega_m)t \tag{3.19}$$

3.3. Generation of SSB Signal

3.3.1. Analog multiplier

In this case, using analog multiplier, the conceptual way to realize the generation of SSB signal is with the help of an analog multiplier followed by a bandpass filter as shown in Figure 3.5. The block diagram represents the generation of SSB signal using analog multiplier.

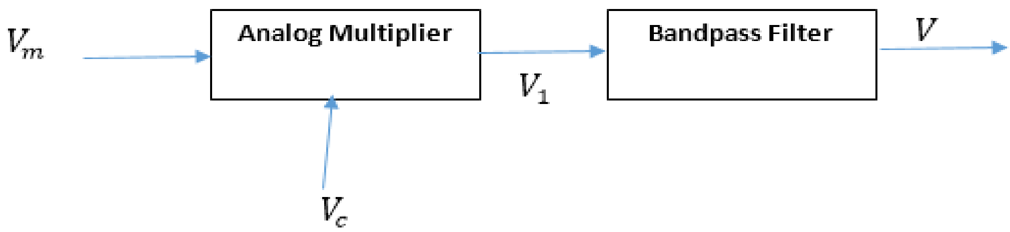


Figure 3.5. Generation of SSB signal using analog multiplier.

The output of the analog multiplier is given by

$$v_1' = v_m v_c = V_m sin\omega_m t V_c sin\omega_c t = \frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \tag{3.20}$$

Thus, at the output of the analog multiplier, we have the DSBSC signal. This signal is passed through a bandpass filter which, depending on the cut-off frequencies, will attenuate one sideband and allows the other to pass through. If the lower sideband is passed out then the output of the bandpass filter will be

$$v = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t \tag{3.21}$$

Alternatively, if upper sideband is passed out, then the output of the bandpass filter will be

$$v = -\frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \tag{3.22}$$

Equ. (3.22 (3.89)results in the generation of SSB signal.

3.3.2. Filter Method

We must understand that the justification for the filter method is that after the balanced modulator, the unwanted sideband is removed by a filter. The block diagram for the filter method of SSB generation is given in Figure 3.6. The balanced modulator generates the DSBSC signal and the sideband suppression filter suppresses the unwanted sideband and allows the wanted sideband. As derived in the previous section, the output of the balanced modulator isgiven by

$$v_1' = 2\alpha c v_m v_c(\cos(\omega_c - \omega_m)\,t - \cos(\omega_c + \omega_m)\,t) \qquad (3.23)$$

The sideband suppression filter is basically a bandpass filter that has a flat bandpass and extremely high attenuation outside the bandpass. Depending on the cut-off frequency values we can represent the output of the filter as

$$v = 2\alpha c v_m v_c \cos(\omega_c - \omega_m)\,t \qquad (3.24)$$

$$v = -2\alpha c v_m v_c \cos(\omega_c + \omega_m)\,t \qquad (3.25)$$

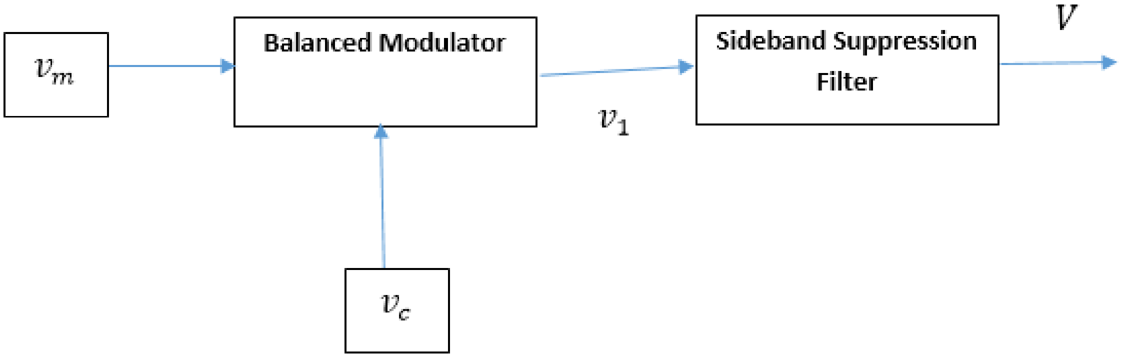


Figure. 3.6.Generation of SSB signal using filter method.

Using Figure. 3.6, the SSB is generated vis-a-vis the filter method.

3.3.3. Phase Shift Method

This approach eschews or avoids filters and some of their inherent disadvantages. It however makes use of two balanced modulators and two phase shifting networks, as shown in Figure. 3.7.This shows a representation of generation of SSB signal using phase shift method.

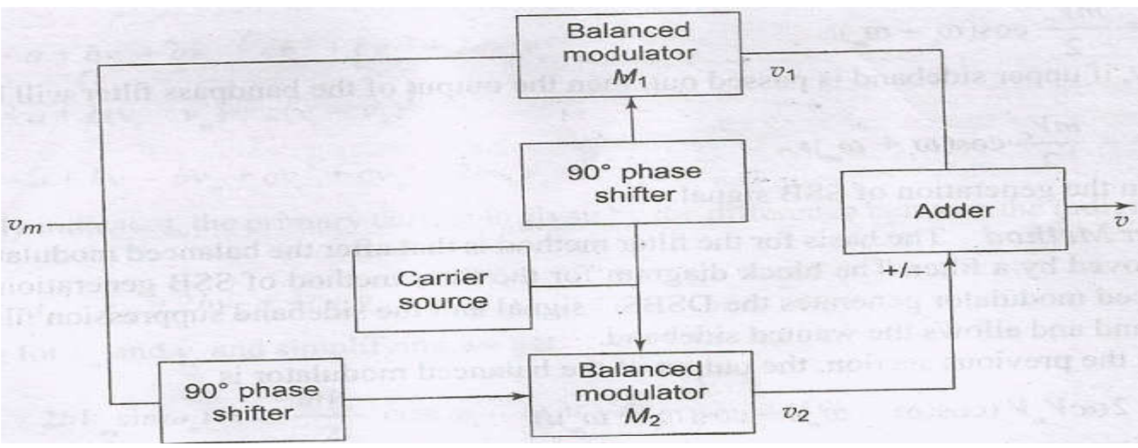


Figure.3.7. Generation of SSB signal using phase shift method.

One of the balanced modulators M_1, receives the 90° phase shifted carrier and in phase message signal, whereas the other, M_2 is fed with the 90° phase shifted message and in phase carrier

signal. Both the modulators produce the two -sidebands. One of the sidebands, namely, the upper sideband will be in phase in both the modulators, whereas, the lower sideband will be out of phase. Thus by suitable polarity for M_2 output and adding with M_1 output results in suppressing one of the sidebands.
Let $v_m = V_m sin\omega_m t$be the message and $v_c = V_c sin\omega_c t$be the carrier. The 90° phase shifted versions of them are $V_m sin\omega_m t$ and $V_c sin\omega_c t$respectively.

The output of the balanced modulator M_1 is given by

$$v_1 = v_m v_c sin\omega_m t cos\omega_c t = \frac{V_m V_c}{2}(\sin(\omega_c + \omega_m)\, t + \sin(\omega_c - \omega_m)\, t) \tag{3.26}$$

The output of the balanced modulator $M_2$is given by

$$v_2 = v_m v_c cos\omega_m t sin\omega_c t = \frac{V_m V_c}{2}(\sin(\omega_c + \omega_m)\, t - \sin(\omega_c - \omega_m)\, t) \tag{3.27}$$

The output of the adder is

$$v = V_1 \pm V_2 \tag{3.28}$$

In one case, we have

$$v = V_m V_c \sin(\omega_c + \omega_m)\, t \tag{3.29}$$

In the other case, we have

$$v = V_m V_c \sin(\omega_c - \omega_m)\, t \tag{3.30}$$

Thus resulting in the generation of SSB signal.

3.3.4. Weaver Phase Shift Method

Another approach of generating SSBis the Weaver phase shift method which retains the advantages of the phase shift method, such as its ability to generate SSB at any frequency and use of low audio frequencies without the associated disadvantage of an audio frequency phase shift network required to operate over a large range of audio frequencies.The block diagram of this third method is shown in Figure3.8.

It could be seen that the later part of this circuit is identical to that of the phase shift method, but the way in which appropriate voltages are fed to the last two' balanced modulators (M_3 and M_4) has been changed. Instead of trying to phase shift the whole range of audio frequencies, this method combines them with an audio frequency carrier ω_0, which is a fixed frequency in the middle of audio frequency band. A phase shift is then applied to this frequency only, and after the resulting voltages have been applied to the first pair of balanced modulators (M_1 and M_2), the low pass filters whose cut-off frequency is ω_0 ensure that the input to the last pair of balanced modulators, results in proper eventual sideband suppression.

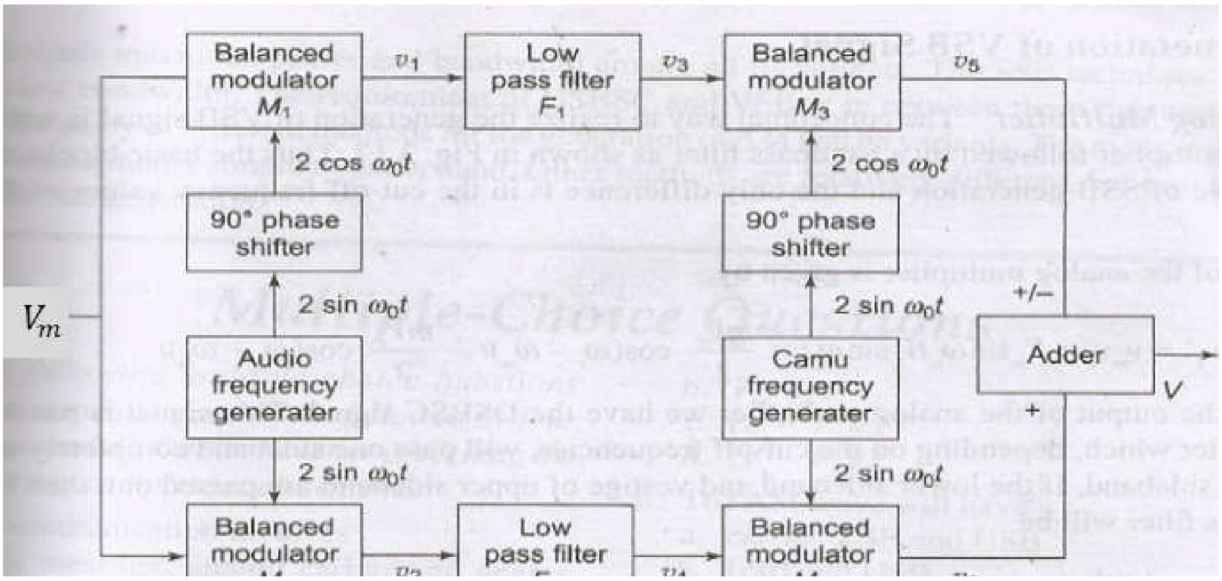


Figure.3.8.Generation of SSB signal using Weaver Phase Shift Method

The output of M_1 is

$v_1 = 2sin\omega_m tcos\omega_c t = \cos(\omega_m + \omega_0)\, t) + \sin(\omega_m + \omega_0)\, t$ (3.31)

The output of M_2 is

$v_2 = 2sin\omega_m tsin\omega_0 t = \cos(\omega_m + \omega_0)\, t) - \cos(\omega_m + \omega_0)\, t$ (3.32)

The output of the low pass filter F_1 is

$v_3 = \sin(\omega_m - \omega_0)\, t$ (3.33)

The output of the low pass filter F_2 is

$v_4 = \cos(\omega_m - \omega_0)\, t$ (3.34)

The output of M_3 is

$v_5 = 2cos\omega_c t \sin(\omega_m - \omega_0)\, t = \sin(\omega_c + (\omega_m - \omega_0))\, t - \sin(\omega_c - (\omega_m - \omega_0))t$ (3.35)

The output of M_4 is

$v_6 = 2sin\omega_c t \cos(\omega_m - \omega_0)\, t = \sin(\omega_c + (\omega_m - \omega_0))\, t + \sin(\omega_c - (\omega_m - \omega_0))t$ (3.36)

The output of the adder is

$v = v_6 \pm v_5$ (3.37)

In one case we have

$v = \sin(\omega_c + (\omega_m - \omega_0))t$ (3.38)

In the other case we have

$v = \sin(\omega_c - (\omega_m - \omega_0))t$ (3.39)

Thus resulting in the generation of SSB signal by the third method.

3.4 Generation of VSB Signal

3.4.1. Using Analog Multiplier

The conceptual way to realize the generation of VSB signal is with the help of an analog multiplier followed by a bandpass filter as shown in Figure. 3.5. Thus, the basic blocks remain same as in the case of SSB generation and the only difference is in the cut-off frequency values of the bandpass filter. The output of the analog multiplier is given by

$$v_1^{'} = v_m v_c = V_m sin\omega_m t V_c sin\omega_c t = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t - \frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t \quad (3.40)$$

Thus at the output of the analog multiplier we have the DSBSC signal. This signal is passed through a bandpass filter which, depending on the cut-off frequencies, will pass one sideband completely and a vestige of the other sideband. If the lower sideband and vestige of upper sideband are passed out, then the output of the bandpass filter will be

$$v = \frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t - F\left(\frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t\right) \quad (3.41)$$

Alternatively, if upper sideband is passed out, then the output of the bandpass filter wilt be

$$v = -\frac{mV_c}{2}\cos(\omega_c + \omega_m)\,t + F\left(\frac{mV_c}{2}\cos(\omega_c - \omega_m)\,t\right) \quad (3.42)$$

This results in the generation of VSB signal.

3.4.2. Using the Filter Method

The basis for the filter method is simple. Now, after the balanced modulator, the unwanted sideband is removed by a filter. The block diagram for the filter method of VSB generation will also remain same as that of SSB case given in Figure3.6. The balanced modulator generates the DSBSC signal and sideband suppression filter suppresses most of the unwanted sideband and allow a vestige of it along with other sideband. As derived in the previous section, the output of the balanced modulator is

$$v_1^{'} = 2\alpha c v_m v_c(\cos(\omega_c - \omega_m)\,t - \cos(\omega_c - \omega_m)\,t). \quad (3.43)$$

The sideband suppression filter is basically a bandpass filter that has a flat bandpass and-extremely high attenuation outside the bandpass. Depending on the cut-off frequency values, we can represent the output of the filter as

$$v = 2\alpha c V_m V_c \cos(\omega_c - \omega_m)\,t - F(2\alpha c V_m V_c \cos(\omega_c + \omega_m)\,t) \quad (3.44)$$

or

$$v = -2\alpha c V_m V_c \cos(\omega_c + \omega_m)\,t + F(2\alpha c V_m V_c \cos(\omega_c - \omega_m)\,t) \quad (3.45)$$

In this way VSB is generated in case of filter method.

4.0. Conclusion

This unit began with the definition of analog and digital communication. The block diagram description of analog communication system was described next to illustrate the fact that the signal at all stage with analog in nature. The theory of basic amplitude modulation and its variants together DSBSC, SSB and was presented next. We understood that the study of all the amplitude modulation techniques gives a better understanding about their nature in time and frequency domains, and power and bandwidth requirements. The basicAM technique needs maximum power and bandwidth among all its variants. The SSB technique needs minimum power and bandwidth. The requirement of DSBSC and VSB is in between these two cases. This was followed by the study of different methods for the generation of AM and its variants. The method using analog multiplier is conceptually simple to understand. Other methods are relatively different, but provide practical approaches for the generation.

5.0. Summary

- The major methods of generating AM signals include:
  - Using Analog Multiplier
  - Using a Nonlinear Resistance Device
- DSBSC Signal can be generated using a Balanced Modulator.
- The relationship between voltage and current in a linear resistance is given by $i = bv$where b is some constant of proportionality.
- Generation of SSB Signal = Analog multiplier + bandpass filter
- In generating the SSB Signal, the balanced modulator generates the DSBSC signal and the sideband suppression filter suppresses the unwanted sideband and allows the wanted sideband.
- Phase Shift Method avoids filters and some of their inherent disadvantages while. making use of two balanced modulators and two phase shifting networks.
- Weaver phase shift method is another approach of generating SSB which retains the advantages of the phase shift method.
- The VSB Signal can be generated using Analog Multiplier and the Filter Method

6.0 Tutor-Marked Assignment (TMA)

1. Describe the AM Wave Generation process Using Analog Multiplier?
2. Analyze the AM Wave Generation process Using Diode As Nonlinear Resistor?
3. Explain DSBSC Wave Generation Process Using Analog Multiplier?
4. Describe The DSBSC Wave Generation Process Using Balanced Modulator?

7.0. References/Further Readings

1. Green D.C (1990); Radio Systems Technology, Longman Scientific and Technical Publishers, Essex, England.
2. G.N.Onoh (2001); Principles of Telecommunications, 2nd Edition, ISBN 978-036-029-8, Immaculate Business Support Services, Enugu, Nigeria.
3. Kennedy George (1977); Electronic Communication Systems, McGraw-Hill, Publishers, London.
4. MacMillan Charles W, (1968); Communication Theory Principles, Macmillan Publishers, New York.
5. Young-Chai Ko (2011); Vestigial Sideband Modulation, KEEE343 Communication Theory Lecture #11, April 7,
6. Ronald L. Allen and Duncan W Mills (2004), Signal Analysis, Time, Frequency, Scale, and Structure, IEEE, Press, Willy –Inter-Science, John Wily &Sons, Inc. publication

UNIT 4 AMPLITUDE MODULATION TECHNIQUES

1.0 Introduction
2.0 Objectives
3.0 Main Content
3.1. Classification of AM Methods
3.2. Classification of Linear Modulation Methods
3.2.1. Linear Shunt Plate Modulation/Anode Choke Modulation/Heising Modulation
3.2.2. Linear Series Plate Modulation
3.2.3 Analysis of Linear Series Plate Modulation
3.2.4. Linear series plate modulation with Class *A* modulating amplifier.
3.2.5. Linear series plate modulation with Class *B* modulating amplifier
3.2.5.1. Advantages of Linear series plate modulation
3.2.5.2. Disadvantages of Linear series plate modulation
3.2.6. Grid Bias Modulation
3.2.6.1. Analysis of Grid Bias Modulation System.
3.2.6.2. Performance comparison between Linear Series Plate Modulation and Grid Bias Modulation
3.2.7. Cathode Modulation
3.2.8. Suppressor Grid Modulation
3.2.8.1. Advantages of Suppressor Grid Modulation
3.2.8.2. Limitations of Suppressor Grid Modulation
3.2.9. Screen Grid Modulation
3.2.10. Collector Modulation
3.3. Classification of Square Law Modulation Methods
3.3.1. Square low Diode Modulation
3.3.2. Van der Biji Modulation
3.4. Suppressed Carrier Balance Modulation
3.5. Balanced modulator
3.6. Ring Modulation
3.7. Switching Modulator
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0 References/Further Readings

1.0 INTRODUCTION

In the previous unit, you have learnt about amplitude modulated signal generation using analog multiplier, and nonlinear Resistance device. The unit also explained the generation of DSBSC Signal using a Balanced Modulator. We explained the generation of SSB signal using analog multiplier, filter method, phase shift method and weaver phase shift method. Finally, the unit discussed the generation of VSB signal using analog multiplier and filter method. As you can see, the AM signal generation are well understood in terms of the elementary concepts on the principle of electronic communication. With these background, it is now important to understand the different classifications of amplitude modulation techniques so as to put the module into context. You will learn that about the different classification of amplitude modulation techniques and then make your choice in any application context.

2.0 OBJECTIVES

After going through this unit, you should be able to

- ✓ explain the different classifications of amplitude modulation techniques
- ✓ distinguish linear modulation methods from grid bias modulation
- ✓ analyze the square law diode modulation, Van der Biji modulation, suppressed carrier balanced modulation, balanced modulation ring modulation, and switching modulator with their various block diagrams/schematics
- ✓ relate spatial coherence with the visibility of fringe pattern, and
- ✓ derive mathematical models for problems based on coherence.

3.0 MAIN CONTENT

3.1. Classification of Amplitude Modulation Methods

Now, you may ask: What are the different types of AM generation methods? Well, generally there are two basic AM generation methods. These methods include:

- Linear Modulation Methods: Theses utilize the linear region of the current-voltage characteristics of the amplifying device, i.e., transistor (or electron tube).
- Square Law Modulation Methods: These methods utilize the square-law region of some current voltage characteristics of a diode or transistor or electron tube. This will be discussed later on in this unit while highlighting the other AM modulation techniques.

3.2. Classification of Linear Modulation Methods

A large number of linear modulation methods have been devised and have used to varying degree. However, the more important amongst these methods include: Linear shunt plate modulation or anode choke modulation or heising modulation, linear series plate modulation, grid-bias modulation, cathode modulation, suppressor modulation, screen grid modulation, and collector modulation.It is very necessary that we discussthis classification for proper understanding.

3.2.1. Linear Shunt Plate Modulation/Anode Choke Modulation/Heising Modulation

A good question is what is linear shunt plat modulation? This is the method of modulation previously used for AM but owing to its inefficiency, the linear series plate modulation for high-level modulation system is preferred. It may be used only for low level modulation systems. Figure (4.1) shows the basic circuit. Here, the carrier voltage to be modulated, drives a tuned class C amplifier using tubeT_2. Modulating voltagev_m is applied at the input of an audio class. Amplifier using tube T_1 and associated circuit. The plate current of both tube T_1 and T_2flow through the audio choke. Looking from the plate supply source side, the two plate current find parallel or shunt paths through tube T_1 and T_2 and hence the name shunt plate modulation is given to this system of modulation.

As the modulation voltage v_m at the input of modulation tube T_1 varies, the plate current of tubeT_1 through the audio choke varies and this changes the plate potentials of tube T_2. Thus, in the modulation amplifier, the carrier voltage v_c appears in the grid circuit. Anode-choke modulation is conventionally used because anode modulation of carrier takes place through the action of the choke.

In general the frequency ω_m of the modulation voltage is very smell as compared with the carrier frequency ω_c so that any variation in plate voltage of modulated class C amplifier caused due to modulating voltage may be considered to be nothing but slow variations in the plate supply voltage. Thus the instantaneous total plate supply voltage v_{bb} consist of a steady component V_{bb} pulse the time varying component V_{mm} cos ω_m^t caused by modulation voltage.Hence,v_{bb} may be expressed by relation,

$$V_{bb} + V_{bb} \cos \omega_m t \tag{4.1}$$

Where V_{bb} is the plate supply voltage of the of the modulated amplifier tube T_2 in the absence of the modulating voltage. $V_{mm}$is the maximum variation of plate voltage as cause by the modulating voltage.

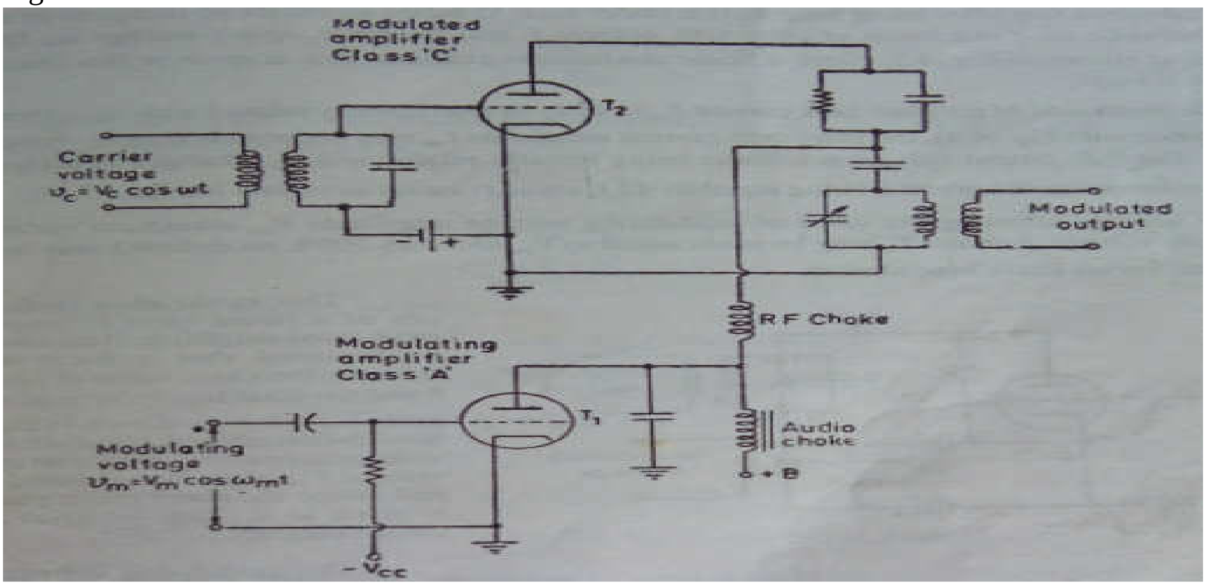


Figure 4.1. Basic shunt plate modulation/heisting modulation

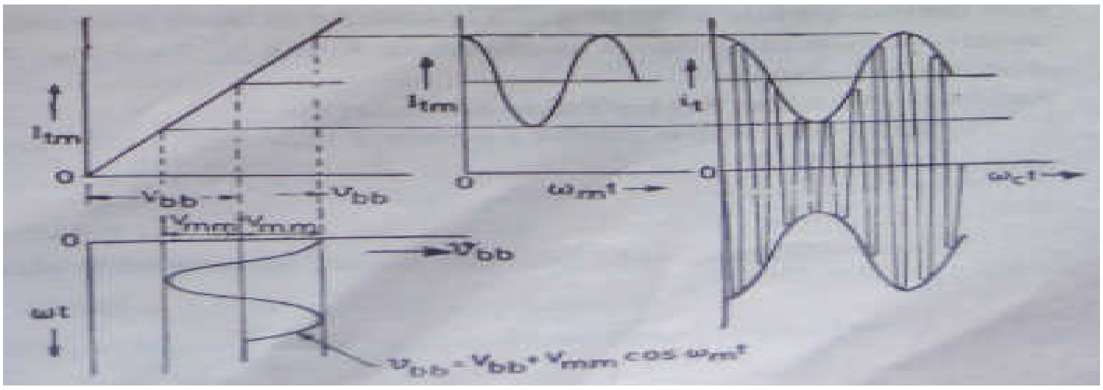


Figure. 4.2. Variation of plate tank current in linear plate modulation.

The amplitude of the plate tank current $I_{tm} = \sqrt{2} I_b$is also linearly related with v_{bb}. Hence as v_{bb} varies in accordance with Equ. (4.1), plate tank current amplitude I_{tm} also undergoes similar changes as shown in Figure. 4.2. The R.F output voltage amplitude being linearly related with I_{tm} changes in accordance with v_{bb}. The instantaneous tank current i_t being equal to $\sqrt{2}I_t$ sin ω_b^t varies as shown in Figure 4.2.By suitable selecting the value of modulating voltage amplitude V_{ms} maximum variation V_{mm} in the plate supply voltage for tube T_2 may be made equal to v_{bb} and thus 100% modulation may be obtained.

3.2.2. Linear Series Plate Modulation

In this section, we shall discuss linear series plate modulation**.** This modulation method also makes use of linear class C tuned amplifier shown in Figure 4.3.

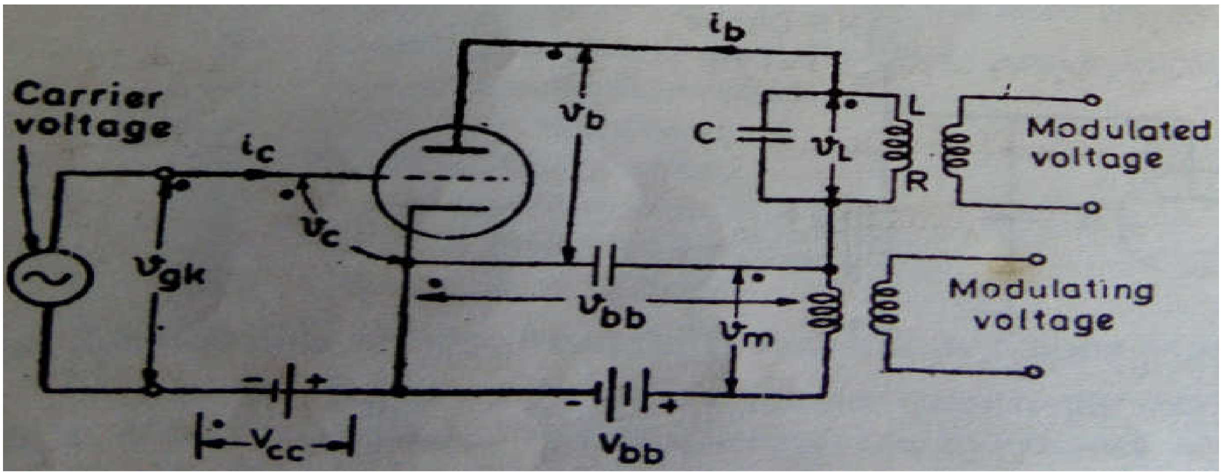


Figure. 4.3. Basic circuit of linear series modulated class C amplifier

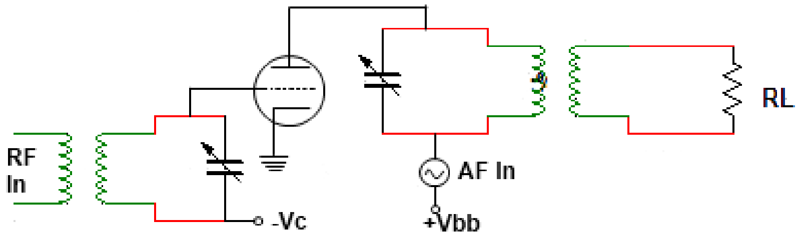


Figure 4.4. Plate Modulator

The linear series plate modulation derives its names because, the message signal AF is superimposed on +Vsb and -Vsb and the applied to the plate of a triode tube. A brief explanation of how the modulation occurs in Figure 4.4 is described. Now, the audio voltage (AF) is paced in series with the plate supply voltage +Vbb of a class C amplifier. In actual circuitsas shown in Figure 4.4, this condition is obtained by applying AF signal to the AF driver transformer which varies the grid bias of both triodes in accordance to the message signal. Due to which the plate currents of both triodes vary with respect to the frequency of AF signal hence the voltage +Vbb applied to the pate of class C amplifier vary in accordance to the amplitude of the AF signal

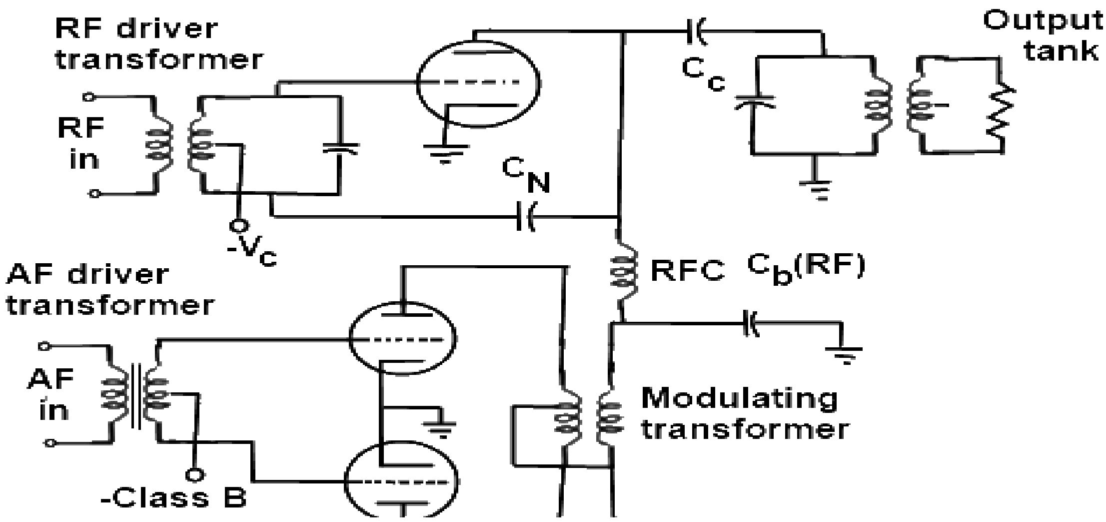


Figure 4.5. Class C Amplifier of plate modulator

3.2.3. Analysis of Linear Series Plate Modulation

From Figure 4.3, It is assumed that carrier frequency$\omega_c \gg \omega_m$. At the grid of class C modulation amplifier we have

$v_{gk} = V_{gm} \cos \omega_c t$ (4.2)

and $v_c = V_{cc} + v_{gk} = V_{cc} + V_{gm} cos\omega_c t$ (4.3)

The modulation voltage in the modulated amplifier plate circuit is given by,

$v_m = V_{mm} cos\omega_m t$ (4.4)

Instantaneous plate supply voltage in modulated amplifiers is given by,

$v_{bb} = V_{bb} + v_m = V_{bb} + V_{mm} cos\omega_m t$ (4.5)

$= V_{bb}[1 + m_a cos\omega_m t$ (4.6)

Where modulation index $m_a$is given by,

$m_a = \frac{V_{mm}}{V_{bb}}$ (4.7)

But from Figure. 4.3, we see that linear relation exists between (i) *r.m.s.* tank current I_b and v_{bb}. This may be as,

$I_t = k_t . v_{bb}$ (4.8)

$I_b = k_b . v_{bb}$ (4.9)

$I_{bb} = k_b . v_{bb}$ (4.10)

Combining Eqns. (4.6) to (4.10), we get

$I_t = k_t . v_{bb}(1 + m_a cos\omega_m t)$ (4.11)

$I_b = k_b . v_{bb}(1 + m_a cos\omega_m t)$ (4.12)

Or$I_b = I_{bb}(1 + m_a cos\omega_m t)$ (4.13)

The instantaneous tank current is given by,

$i_t = \sqrt{2} I_t \sin\omega_c t$ (4.14)

Or$i_t = \sqrt{2} I_t V_{bb}(1 + m_a cos\omega_m t) sin\omega_c t$ (4.15)

The *r.m.s* value of our potential across the tank circuit is given by,

$V_t = jXI_t = XI_t \angle 90^o$ (4.16)

Where X= reactance of inductance or condenser in the tuned circuit.
Hence instantaneous output voltage is given by,

$v_t = \sqrt{2}XI_t \sin(\omega_c t + 90^o)$

$= \sqrt{2}Xk_tV_{bb}(1 + m_a cos\omega_m t)cos\omega_c t$ (4.17)

Plate-to-cathode voltage of modulated amplifier is given by,

$v_b = v_{bb} - v_t$

Or $v_b = v_{bb}(1 + m_a cos\omega_c t) - \sqrt{2}xk_tV_{bb}(1 + m_a cos\omega_c t)cos\omega_c t$

Or $v_b = v_{bb}(1 + m_a cos\omega_c t)[1 - \sqrt{2}xk_t cos\omega_c t$ (4.18)

The *r.m.s* value of plate current at resonance frequency is given by,

$I_{p1} = \frac{V_t}{R_o} \approx j\frac{I_t}{Q} = \frac{I_t}{Q}\angle 90^O$ (4.19)

Hence, instantaneous plate current is given by,

$i_{p1} = \sqrt{2}I_{p1} \sin\omega_c t = \sqrt{2}\frac{I_t}{Q}\sin(\omega_m t + 90^o) = \frac{\sqrt{2}}{Q}k_tV_{bb}(1 + m_a cos\omega_c t)cos\omega_m t$

(4.20)

Average power supplied by d.c plate power source is given by,

$P_{bb} = \frac{1}{T_m}\int_o^{T_m} V_{bb}I_b dt$ (4.21)

Where$T_m$is the periodic time of the modulation cycle

Or $P_{bb} = \int_o^{T_m} V_{bb}I_{bb}(1 + m_a \cos\omega_c t)dt$ (4.22)

$= V_{bb}I_{bb} = k_bV_{bb}^2$ (4.23)

Average power supplied by the modulating voltage source is,

$P_m = \frac{1}{T_m}\int_o^{T_m} v_m I_b dt$ (4.24)

Substituting the value of $v_m$and I_b from Eqs.(4.4) and (4.13) respectively into Equ. (4.24)

$P_m = \frac{1}{T_m}\int_o^{T_m} v_m cos\omega_c t I_{bb}(1 + m_a cos\omega_m t)dt$ (4.25)

$= V_{mm}I_{bb}\frac{m_a}{2} = V_{bb}I_{bb}\frac{m_a^2}{2} = P_{bb}\frac{m_a^2}{2}$ (4.26)

Thus, modulating power provides an average power $m_a^2/2$ times the average power from dc plate power source.For 100 percent modulation with sinusoidal modulating voltage, the modulation source is required to provide an average power of one-half of that delivered by the d.c plate power source. The modulating amplifier must accordingly be designed to deliver this large amount of power. The total average input power is then,

$P_i = P_{bb} + P_m = P_{bb}\left(1 + \frac{m_a^2}{2}\right)$ (4.27)

The a.c out power across the tank circuit is given by,

$P_o = \frac{1}{T_m}\int_o^{T_m} v_1.i_{p1}dt = \frac{1}{T_m}\int_o^{T_m} R_m.\ i_{P1}^2 dt$ (4.28)

Or $P_o = \frac{1}{T_m}\int_o^{T_m} R_O 2\left(\frac{k_iV_{bb}}{Q}\right)^2(1 + m_a \cos\ \omega_m t\ dt)^2 cos^2\ \omega_c t\ dt$

$= \left(\frac{k_iV_{bb}}{Q}\right)^2\frac{R_o}{T_m}\int_o^{T_m}(1 + m_a^2\cos^2\ \omega_m t + 2m_a\, cos\ \omega_c t)dt$ (4.29)

In Figure 4.29averaging of power over the audio frequency periodic time T_m has been done. Since,$\omega_o \gg \omega_m$, the average value of cos $2\omega_{c^t}$ is zero. Similarly average value of $2m_a$ cos $\omega_m t$ is zero. Hence,

$$P_o = \left(\frac{k_i V_{bb}}{Q}\right)^2 \frac{R_o}{T_m} \int_o^{T_m} (1 + m_a^2 cos\ \omega_m t) dt$$

By integrating, we get

$$P_o = R_o \left(\frac{k_i V_{bb}}{Q}\right)^2 \left(1 + \frac{m_a^2}{2}\right) \quad (4.30)$$

$$= \frac{R_o k_t^2}{Q^2 K_b} . P_{bb} \left(1 + \frac{m_a^2}{2}\right) \quad (4.31)$$

$$= P_c \left(1 + \frac{m_a^2}{2}\right) \quad (4.31a)$$

Where $P_c = R_o \frac{k_t^2}{Q^2 k_b} P_{bb}$ (4.31b)

$P_c$represents the unmodulated carrier power sideband power is $P_c m_a^2 / 2$.But Equ. (4.26) shows that the average power supplied by the modulating voltage source is $m_a^2 / 2$ times the P_{bb}, the average power from the plate supply source. We therefore, we can conclude that the d.c plate source supplies the carrier $P_c = R_o \frac{k_t^2}{Q^2 k_b} P_{bb}$

Whereas the modulating amplifier suppliers the average power $\left(P_m = \frac{m_a^2}{2} P_{bb}\right)$toproduce the sideband power$\frac{m^2}{2} P_c$.

The plate circuit efficiency of this linear series plate modulated amplifier is given by,

$$P_o = \frac{P_o}{P_i} = \frac{P_o}{P_{bb} + P_m} = \frac{(R_o\, k_t^2 / Q^2 k_b).\left(1 + \frac{m_a^2}{2}\right)}{k_{bb}\left(1 + \frac{m_a^2}{2}\right)} = \frac{R_o\, k_t^2}{Q^2 k_b} \quad (4.32)$$

Equ. (4.32) clearly shows that in a linear series plate modulation amplifier, plate circuit efficiency remains constant at $R_o\, k_t^2 / Q^2 k_b$ irrespective of the degree of modulation. For a given linear series plate modulated class Camplifier, we may simple calculate the plate circuit efficiency of this class C amplifier under unmodulated condition. The same value holds good when this class C amplifier is plate modulated. Plate dissipation, plate dissipation is given by,

$$P_p = P_i - P_o = P_i(1 - \eta_p) = (P_{pp} + P_m)(1 - \eta_p)$$

$$= P_{bb}\left(1 + \frac{m_a^2}{2}\right)(1 - \eta_p) \quad (4.33)$$

When modulation index is zero, $P_{po} = P_{bb}(1 - \eta_p)$ (4.34)

Equ. (4.33) may then be put as, $P_p = P_{po}\left(1 + \frac{m_a^2}{2}\right)$ (4.35)

Equ. (4.35) shows that plate dissipation with the power of modulation index m_a. Care must therefore be taken in choosing plate supply voltage and modulation power P_m so as not to exceed the maximum allowed plate dissipation P_p.

The following two systems of linear series plate modulation with class *C* modulation amplifier exists viz: Linear series plate modulation with class *A* modulating amplifier and Linear series plate modulation with class C modulating amplifier.

3.2.4. Linear Series Plate Modulation with Class *A* Modulating Amplifier

This uses small values of carrier power and a single operation for both the modulated amplifier and the modulating amplifier. Its overall efficiency of the system is then low but it is not a serious drawback at low carrier power level. Figure 4.6 shows the schematic circuit.

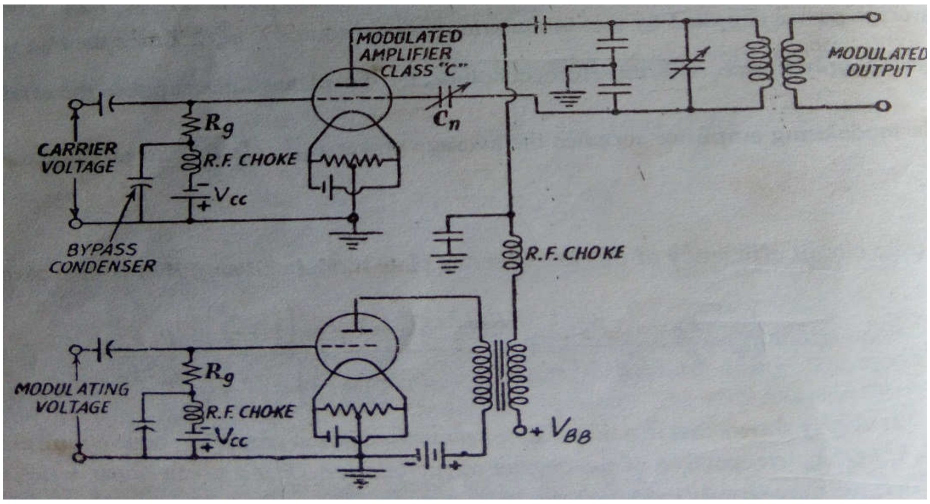


Figure. 4.6. Linear series plate modulation with class *A* modulating amplifier

Both modulating and modulated amplifiers use directly heating tungsten cathode in order to have large current ratings. To avoid uneven heating of filament, a high resistance is connected across the filament and circuit connection is made at the center point. An RF filter consisting of RF choke and a bypass capacitor is provided in the grid circuit to prevent the RF current from entering the cathode circuit. Suitable combination of fixed bias and grid leak bias is used to have desired linear modulation. Shunt feeding of the anode is preferred.

It is observed that average overall efficiency of this modulation system is very low and the variation in efficiency is high. In order to increase the overall efficiency and reduce the variation in overall efficiency, class B modulation amplifier is normally used.

3.2.5. Linear series plate modulation with Class *B* modulating amplifier

For high carrier powers, high overall efficiency is absolutely necessary. Hence, class B modulating amplifier is used. Further for large power handling capacity and for reduction of distortion push-pull operation is used in both the modulated amplifier and modulating amplifier. Figure 4.7 gives the basic circuit.

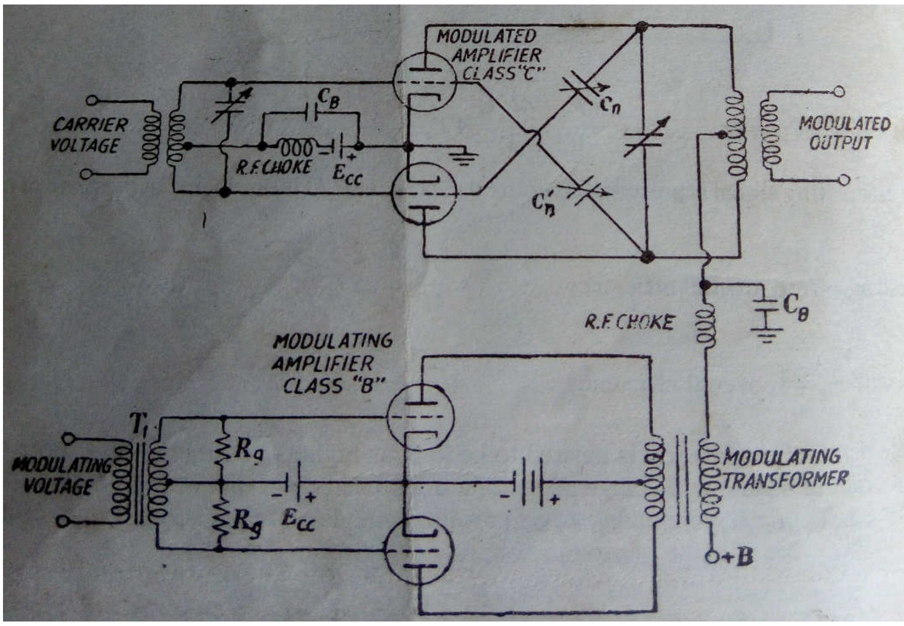


Figure. 4.7. Basic circuit of linear series plate modulation with class *B* Push-pull modulating amplifier

The modulation amplifier is required to feed large amount of the power to the modulated amplifier. To provide such a large amount of power, the grids in modulating push pull amplifier are required to be to be driven positive for a part of the cycle and the impendence looking back into the previous stage will be different stage for the case when the grid current flow and the grid current to not flow.

Nonlinear distortion will be caused then. Such a distortion is reduced by using for driving transformingT_1, a step-down transformer. As an additional remedy, resistors, R_2 (400 or 500 ohms) are connected across the secondary of driving transformer T_1.

Cross neutralization is used in the R.F modulation amplifier to neutralize the feedback effect of grid to-plate capacitance of amplifier tubes.Overall efficiency of this system including both the modulated and modulating amplifiers may be calculated for different values of modulation index. With this system of plate modulation, the overall efficiency is high and further the efficiency remains appreciably constant for various values of modulation index.

3.2.5.1.Advantages of Linear series plate modulation

Linear series plate modulation is widely used in high power radio transmitters. The merits of this system of modulation are:

1. Distortion is small

2. Plate circuit efficiency is high and
3. Adjustment is easy.

3.2.5.2. Disadvantages of Linear series plate modulation

The main drawback of this method of modulation is that a large amount of modulation power is required. This necessitates a heavy, bulky and costly modulation equipment.

3.2.6. Grid Bias Modulation

Grid bias modulation consists in feeding the audio modulating voltage in series with the fixed grid bias in the grid circuit of a class C amplifier. Figure 4.8 shows the basic circuit of grid bias modulated amplifier. In this case, we thus have three voltages in series in the grid circuit namely:

i. Fixed grid bias V_{cc},
ii. Modulating voltage v_m and
iii. Carrier voltage v_{gk}.

It assumed that modulation frequency ω_m is much smaller than the carrier frequency ω_c, then from the point of view of the carrier voltage, the modulation voltage v_m together than constitute the variation grid bias $v_m$together than constitute the variable grid bias v_{cc}.

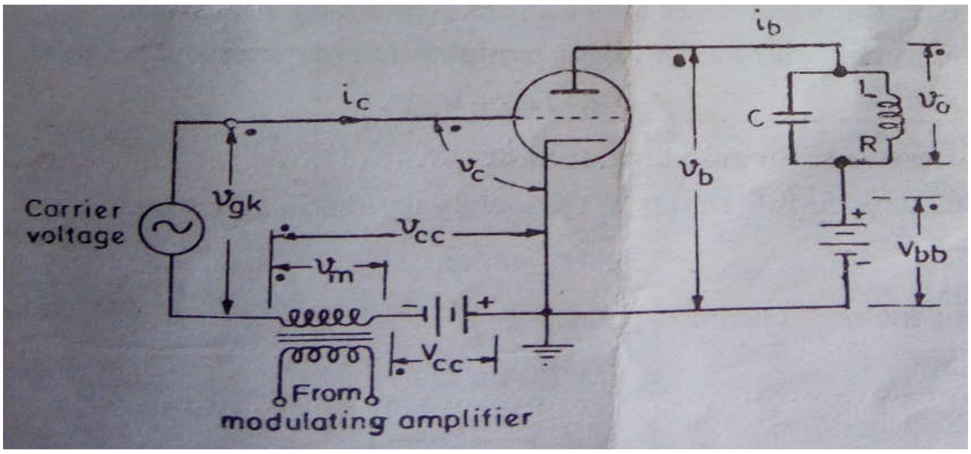


Figure4.8. Basic circuit of grid bias modulation.

3.2.6.1.Analysis of Grid Bias Modulation System

We must understand that the analysis of grid bias modulation amplified may be done in a way similar to that of linear series plate modulation.

Now, let the carrier voltage be given by,$v_{gk} = V_{gm} \sin\ \omega_c t$ (4.36)

The modulation voltage is,$v_m = V_{mm} cos\ \omega_c t$ (4.37)

Total varying grid v_{cc} is given by,$v_{cc} = V_{mm} cos\ \omega_c t$ (4.38)

Total grid-to-cathode voltage v_c is given by

$$v_c = v_{gk} + v_{cc}\ sin\ \omega_c t + V_{mm} cos\ \omega_c t + V_{cc} \tag{4.39}$$

Let the *r.m.s* value of tank current for linear modulation characteristic be given by,

$$I_t = k_i\ (v_{cc} - V_{co}) \tag{4.40}$$

Where $V_{co}$is grid-to-cathode voltage for zero tank current.

But the instantaneous tank current varies at frequency ω_c and is given by,

$$i_t = I_{tm} \sin\ \omega_c t \tag{4.41}$$

Where I_{tm} is the amplitude of the tank current.

Hence $i_t = \sqrt{2} I_t \sin\ \omega_c t$ (4.42)

Or$i_t = \sqrt{2} k_t . (v_{cc} - V_{co}) \sin\ \omega_c t$ (4.43)

$$= \sqrt{2} k_t . [V_{mm} cos\ \omega_{mt} + V_{cc} - V_{co}] \sin\ \omega_c t \tag{4.44}$$

When the modulation voltage$v_m$is zero, tank current is given by,

$$i_{to} = \sqrt{2} k_t [V_{cc} - V_{co}] \sin\ \omega_c t \tag{4.45}$$

$$i_{to} = I_{tmo} \sin\ \omega_c t \tag{4.46}$$

Where $I_{tmo}$is the amplitude of the tank current with zero modulation voltage and is given by,

$$I_{tmo} = \sqrt{2} k_t [V_{cc} - V_{co}] \tag{4.47}$$

Equ. (4.44) may be written as

$$i_t = \sqrt{2} k_t (v_{cc} - V_{co}) \left[1 + \frac{V_{mm} \cos\ \omega_c t}{V_{cc} - V_{co}}\right] \sin\ \omega_c t \tag{4.48}$$

$$i_t = I_{tmo} \left[1 + \frac{V_{mm}}{V_{cc} - V_{co}} cos\ \omega_c t\right] \sin\ \omega_c t \tag{4.49}$$

$$i_t = I_{tmo} [1 + m_a cos\ \omega_c t] \sin\ \omega_c t \tag{4.50}$$

Where $m_a$is the modulation index and is given by

$$m_a = \frac{V_{mm}}{V_{cc}-V_{co}} \tag{4.51}$$

Now since the grid bias is no longer constant but consists of a steady component V_{cc} and variable component v_m the d.c plate current I_b is no longer constant but varies in accordance with the variation of variable grid bias v_{cc} as given b the relation.

$$I_b = k_b(V_{cc} - V_{co}) \tag{4.52}$$

The steady value C.D plate current is given by,

$$I_{bb} = k_b(V_{cc} - V_{co}) \tag{4.53}$$

Substituting the value of v_{cc} as given by Equ. (4.38) intoEqu. (4.52), we get

$$I_b = k_b(V_{cc} - V_{mm} cos\ \omega_c t - V_{cc})$$

$$= k_b(V_{cc} - V_{mm})\left[1 + \frac{V_{mm} \cos\ \omega_c t}{V_{cc}-V_{co}}\right] \tag{4.54}$$

$$= k_b(V_{cc} - V_{mm})(1 + m_a cos\ \omega_c t) \tag{4.55}$$

$$= I_{bb}(+m_a cos\ \omega_c t) \tag{4.56}$$

Tank circuit voltage is the desired A.C output voltage and given by,

$$V_o = j\, XI_t \tag{4.57}$$

Where X is the reactance of either the inductor or capacitors at the resonant frequency ω_c Equ. (4.57)may be put as,

$$V_o = j\, X\, k_t(V_{cc} - V_{co})(1 + m_a cos\ \omega_c t) \tag{4.58}$$

The instantaneous output voltage is given by,

$$v_o = \sqrt{2}\text{X}\ k_t(Vv_{cc} - V_{co})(1 + m_a cos\ \omega_c t) \sin\ \omega_c t\ 90^o)$$

$$= \sqrt{2}k_t(V_{cc} - V_{co})(1 + m_a cos\ \omega_c t) \cos\ \omega_c t \tag{4.59}$$

D.C input power. The average input power from the D.C plate source is given by,

$$P_{bb} = \frac{1}{T_m}\int_o^{T_m} V_{bb} I_b\ dt \tag{4.60}$$

Where T_m is the periodic time of the modulating voltage v_m.

Putting the value of I_b in Fig (4.60), we get

$$P_{bb} = \frac{1}{T_m}\int_o^{T_m} V_{bb} I_{bb}(1 + m_a cos\ \omega_c t) dt$$

On integration, we get

$$P_{bb} = V_{bb}.I_{bb} \tag{4.61}$$

Thus the D.C input power from the plate supply source remains constant for all values of modulation indexm_a. The A.C output power is given by

$$P_o = \frac{1}{T_m}\int_o^{T_m} v_o i_p dt = \frac{1}{T_m}\int_o^{T_m} R_o.i_p^2 dt \tag{4.62}$$

Where i_p the a.c component of is plate current at fundamental frequency and R_o is the impedance of the tank at resonance of the frequency.

3.2.6.2.Characteristics of Grid Bias Modulation

As you might want to know, we have features associated with grid bias modulation. Now, the grid bias modulation has the following salient features:
i. The amount of power required from modulating amplifier is small as compared with plate modulation.
ii. For 100% modulation, the peak amplitude of modulated voltage is twice the amplitude of the unmodulated carrier and hence assuming the same permissible peak power for a tube, the maximum carrier amplitude that may be used in grid bias modulated amplifier is half of that permitted in the class *C* amplifier using the same tube. Thus, the carrier power that is obtainable from the modulated amplifier is roughly one-quarter of that obtainable for the same tube when used as class *C* modulated amplifier.

iii. Plate circuit efficiency at zero modulation of class *C* amplifier offers an increase of modulation index. At 100% modulation, the plate circuit efficiency is about 51%.

3.2.6.3.Performance comparison between Linear Series Plate and Grid Bias Modulations

We now want to make some basic comparisons between linear series plate and grid bias modulations. The following are the identified comparison regarding plate and grid bias modulation.

i. Plate Circuit Efficiency: Grid modulation amplifier has a low plate circuit efficiency of 34 to 51 per cent whereas the series plate modulated amplifier has large plate circuit efficiency of 75%.

ii. Power output in problem: This is small in grid bias modulation but large in plate modulation.

iii. Amount of modulation power: Grid bias modulation requires small modulating power whereas plate modulation requires large amount of modulating power.

Point i and ii above supports the plate modulation whereas iii supports grid bias modulation. As a result of modulation, the overall efficiency considering both the amplifier modulated, is more or less the same in the method of modulation. Further, both the methods are capable of giving almost 100% modulations with good linearity of modulation

Thus, there is no clear preference of one method over the other. However the circuit adjustments in grid bias modulation are more sensitive to changes in carrier voltage, plate supply voltage and load impedance and hence it is usually very difficult to achieve and maintain perfect linearity of modulation index.

From the consideration of ease of adjustment, high efficiency plate modulation is generally employed in amplitude modulation radio transmitters operating on medium and sort waves. However in spite of difficulty of adjustment, grid bias modulation of last RF power amplifier is used in high power television because of large width (about 7MHz) of the frequency band involved.

3.2.7. Cathode Modulation

In cathode modulation, the modulation voltage is introduced in the cathode circuit of the modulation amplifier as shown in Figure 4.9. The cathode circuit is common to both the grid circuit and plate circuit, hence the modulating voltage appears in both the grid and the plate circuit. Both the grid bias modulation and the plate modulation of the carrier, therefore, take place.

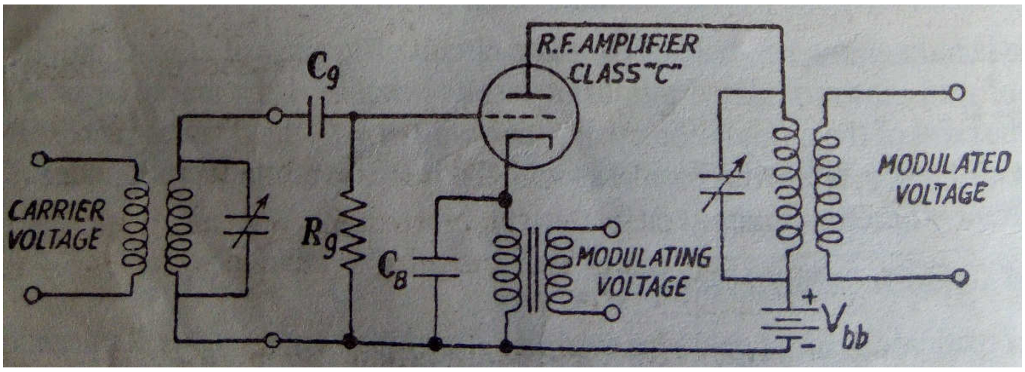


Figure 4.9. Basic circuit of cathode modulation.

Cathode modulation therefore has characteristics intermediate between those of plate modulation and grid bias modulation. Thus, the plate circuit efficiency, power output and modulation power requirement are less than those of plate modulation but more than those for grid bias modulation. The proportion of grid bias modulation and plate modulation depends upon the circuit adjustments.

Thus the proportion of plate modulation may be increased by (i) increasing the grid bias and (ii) by increasing the modulating voltage amplitude. With grid leak biasing arrangement, grid bias may be increased by increasing the value of grid leak resistance R_g. Thus, depending upon the relative proportion of gird bias modulation, and plate modulation, the cathode modulation may have (i) plate circuit efficiency ranging from about 40% corresponding to grid bias modulation at about 70% corresponding to plate modulation and (ii) modulation power requirement ranging from about 5% of the output carrier power corresponding to plate modulation. The exact values of plate circuit efficiency and modulation power requirement depend upon plate supply voltage, grid bias tank circuit impedance, modulation index, slope of constituent grid bias modulation and plate modulation curves and upon the degree of linearity of modulation.For optimum operation, grid bias is kept much greater than that for normal class C operation and the input carrier amplitude is nearly half. The cathode modulation amplifier is thus driven lightly as compared with conventional class C amplifier.

3.3.8. Suppressor Grid Modulation

This is a method of modulation that uses a pentode in class C modulation while applying the modulating voltage to the suppressor grid. The suppressor grid is biased sufficiently negatively and the modulation voltage and this bias are so adjusted that the modulation characteristics lies almost entirely in the region of negative suppressor grid voltage. The suppressor grid current is zero and hence modulation power required is zero. Figure 4.10 shows the basic circuit arrangement.

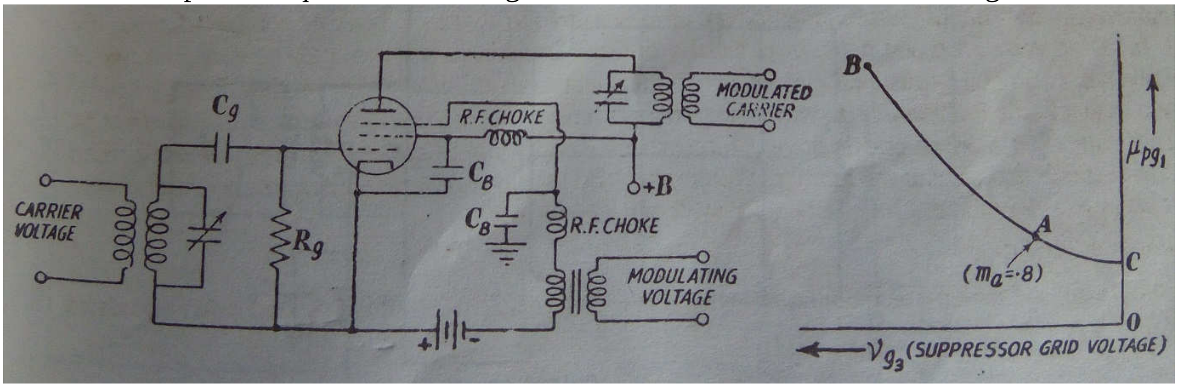


Figure 4.10.Basic circuit of suppressor grid Modulated amplifier with the modulation characteristic

This method of modulation in Figure 4.10 is based on the property that the amplification factor between plate and control grid remains a function of the suppressor grid voltage.

3.2.8.1. Advantages of Suppressor Grid Modulation

From the foregoing, the suppressor grid modulation has the following merits:

i. Low modulation power required by the suppressor is never positive.
ii. High linearity of modulation over the range of percentage modulation for which the modulations effective used. i.e upto maximum modulation.
iii. The carrier voltage is applied to the control grid whereas the modulation voltage is applied to the suppressor grid. Hence no limitation of the RF range occurs when selective circuits are used for separating the modulation frequency and carrier frequency current. This is an important advantage on long wave circuits.
iv. Neutralization is not required.
v. Adjustment of modulated amplifier is simple.

3.2.8.2.**Limitations of Suppressor Grid Modulation**

The limitations of this method of modulation are:

i. This may be used for low carrier power only, i.e.,up to about 500 watts
ii. The circuit is very sensitive to change in bias and the tube parameter. Distortion is caused thereby.

3.2.9. Screen Grid Modulation

This method of modulation is similar to grid bias modulation and suppressor grid modulation. However as compared with the other two methods, this method has the disadvantage that it requires somewhat greater modulating power but has the disadvantage that the screen gird input varies little during the modulation cycle. Thus the modulated amplifier presents to the modulating amplifier almost constant load impedance thereby keeping the distortion low.

3.2.10. Collector Modulation

Modern low AM transmitters use transistor for producing carrier power upto a few hundred watts. In this regard, the two types of AM modulation are i) the collector modulation and (ii) base modulation. These methods of modulation correspond respectively to the anode modulation and grid-bias modulation. Evidently collector modulation is preferred and more properly used. Figure 4.11 gives the basic circuit of collector modulated class C amplifier in which modulation is performed at the RF power amplifier stage. This modulated amplifier is always a push-pull amplifier. In this case, the leak type bias is used.

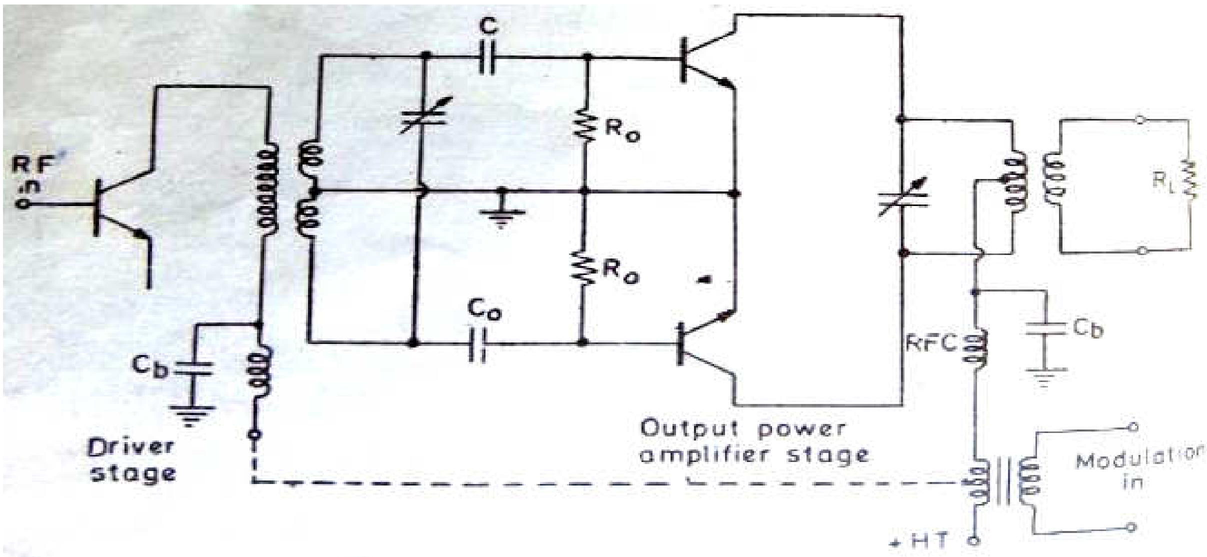


Figure 4.11. Collector modulation Schematics

Collector modulation is superior to base modulation for the following reason: (i) better linearity of modulation (ii) higher collector circuit efficiency and (iii) higher power output per transistor. However, it has the drawback that it needs higher modulation power. Further 100 percent modulation cannot be achieved in collector modulation system due to collector saturation. In several cases, therefore, we use a compound form of modulation.

Figure 4.11 shows one such arrangement. Connections made as shown by the dotted lines. This produces collector modulation of the output R.F power amplifier as well as the driver. Another alternative is to use simultaneously collector modulation and base modulation of the same stage. Here again leak-type bias is used but such bias may become excessive and the power output may drop. Accordingly simultaneous base and collector modulation on the same pattern as in Figure 4.11 is preferred. Simultaneously drain and gate modulation of FET amplifier is also sometimes used.

3.3. Classification of Square Law Modulation Methods

So far, we have introduced various types of amplitude modulation generation schemes. Again, in AM, we have the square law and switching modulation schemes. In Square law modulator, the circuit makes use of nonlinear current-voltage characteristics of diodes triodes and are, in general, suited for use of low voltage. Also, in this type of modulator, when the output device is not directly proportional to the input throughout the operation, the device is said to be non-linear. Important square law modulation methods include: Square law diode modulation, Van der Bijimodulation, balance modulation and switching modulation. These are discussed below.

3.3.1. Square low Diode Modulation

It utilizes the non-linear region of current voltage dynamic characteristic of a diode. This dynamic characteristic is highly nonlinear in the low voltage region as is obvious from Figure 4.11a.The A.C current may then be expressed as function of the A.C voltage as given by the following Taylor series.

$$i_a = a_1.v_s + a_2 v_s^2 \qquad (4.72)$$

Where a_1 and a_2 are Taylor series coefficient. i_a is the A.C anode current; v_s is the A.C anode voltage.Figure 4.11a shows the basic circuit arrangement of square law diode modulation with the Current voltage dynamic characteristic of a diode.Figure 4.11b shows the simplified diagram.

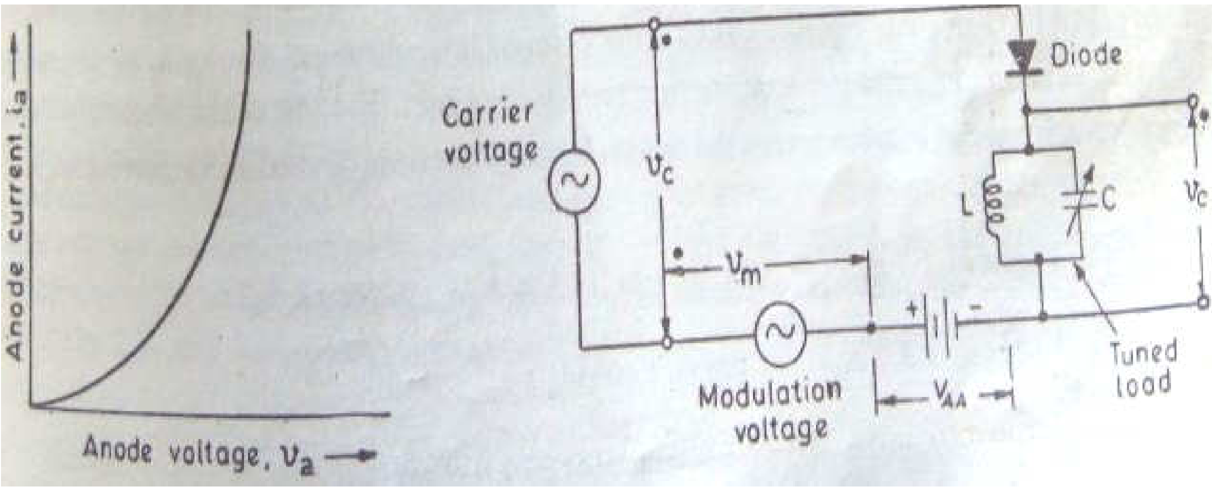


Figure 4.11a. Basic circuit arrangement of diode modulation

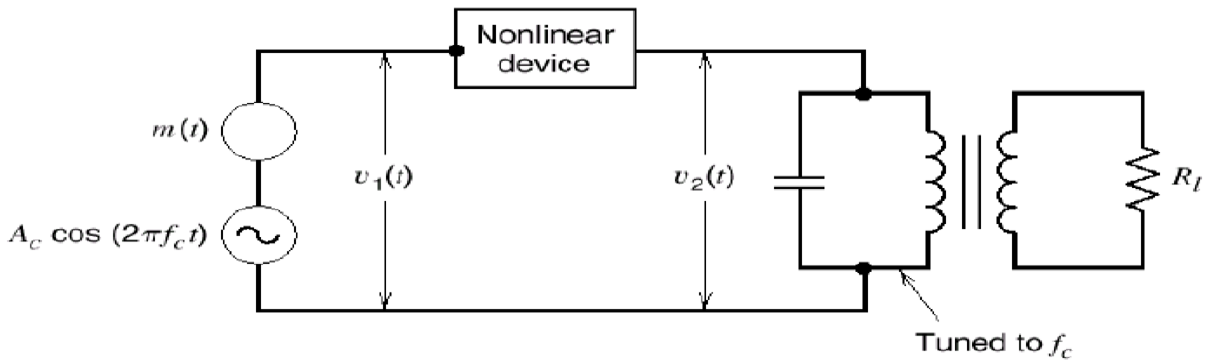


Figure 4.11b. Square Law modulator.

From Figure 4.11b, the input-output relation of a non-linear device can be express as

$V_0 = a_0 + a_1 V_{in} + a_2 V_{in}^2 + a_3 V_{in}^3 + a_4 V_{in}^4 + \cdots ...$

When the input is very small, the higher power terms can be neglected. Hence,, the output is Approximately given by $V_0 = a_0 + a_1 V_{in} + a_2 V_{in}^2$.

When the output is considered up to square of the input, the device is called a square law device and the square law modulator is shown in Figure 4.11b. Now from Figure 4.11a, Let the carrier voltage be given by,

$v_c = V_c \cos \omega_c t$ (4.73)

Let the modulation voltage be given by,

$v_m = V_m \cos \omega_c t$ (4.74)

Then the entire A.C plate voltage v_s is given by

$v_e = v_m + v_m = [V_c \cos \omega_c t + V_m \cos \omega_c t]$ (4.75)

Hence,

$i_a = a_1 . v_s + v_{s^2} = a_i[V_c \cos \omega_c t + V_m \cos \omega_c t] + a_2[V_m \cos \omega_c t + V_m \cos \omega_c t]^2$

$= a_1[V_c \cos \omega_c t + V_m \cos \omega_c t]$

$+ a_2[V_c^2 cos^2 \omega_c t + V_m^2 cos^2 \omega_m + 2V_c V_m \cos \omega_c t \cos \omega_c t]$ (4.76)

The various frequency components may be identified in Figure 4.12 as below

1. $i_a V_m \cos \omega_c t$ –Carrier frequency component of frequency ω_c

2. $a_2 V_m \cos \omega_c t$ –Modulation frequency component of frequency ω_m

3. $a_2 V_c^2 \cos^2 \omega_c t = a_2 V_c^2 \left[\frac{1 + \cos 2 \omega_c t}{2}\right]$.

This consistent of d.c component $a_2 V_c^2 / 2$and a component of frequency $2\omega_c$.

4.$a_2 V_c^2 \cos^2 \omega_c t = \frac{a_2 V_m^2}{2}[1 + \cos 2 \omega_c t]$.

This consistent of d.c component $a_2 V_c^2 / 2$and a component of frequency $2\omega_m$

5.$2a_2 V_c V_m \cos 2\omega_c t \cos 2 \omega_c t = a_2 V_c V_m [\cos(\omega_c \omega_m) + cos(\omega_c - \omega_m)t]$.

Here, $a_2 V_c V_m \cos(\omega_c - \omega_m))t$ is the upper sideband term of frequency$(\omega_c - \omega_m)$

Whereas$a_2 V_c V_m (\omega_c - \omega_m)t$ is the lower sideband term of frequency).

As such,we have in all 6 terms of different frequencies in addition to the d.c component. The load impedance is a tuned circuit which is tuned to carrier frequency ω_c.It responds to a narrow band of frequencies centered about the carrier frequency. Hence, the components which are developed in the output are terms of frequency $\omega_c(\omega_c - \omega_m)$ assuming that $\omega_m \ll \omega_c$. The rest of the frequency terms will not produced appreciable output voltage across the tuned circuit. Hence the desired anode current is given by,

$i_o = a_1 V_c \cos \omega_c t + a_2 V_c V_m cos(\omega_c + \omega_m)t + a_o V_c V_m cos(\omega_c - \omega_m)t$ (4.77)

$= a_1 V_c \cos \omega_c t + 2a_2 V_c V_m \cos \omega_c t \cos \omega_m t$

Or $$i_o = a_i V_c \left[1 + \frac{2a_2 \, V_m}{a_1}\right] \cos \, \omega_c t \quad (4.77a)$$

Or $$i_o = a_i V_c [1 + m_a \cos \, \omega_c t] \cos \, \omega_c t \quad (4.78)$$

Where modulation index,m_a=$\frac{2a_2 \, V_m}{a_1}$ (4.79)

The modulation output voltage is then given by,

$$v_o = i_o R \, t \quad (4.80)$$

Where Rt is the impedance of the tuned circuit at resonance.

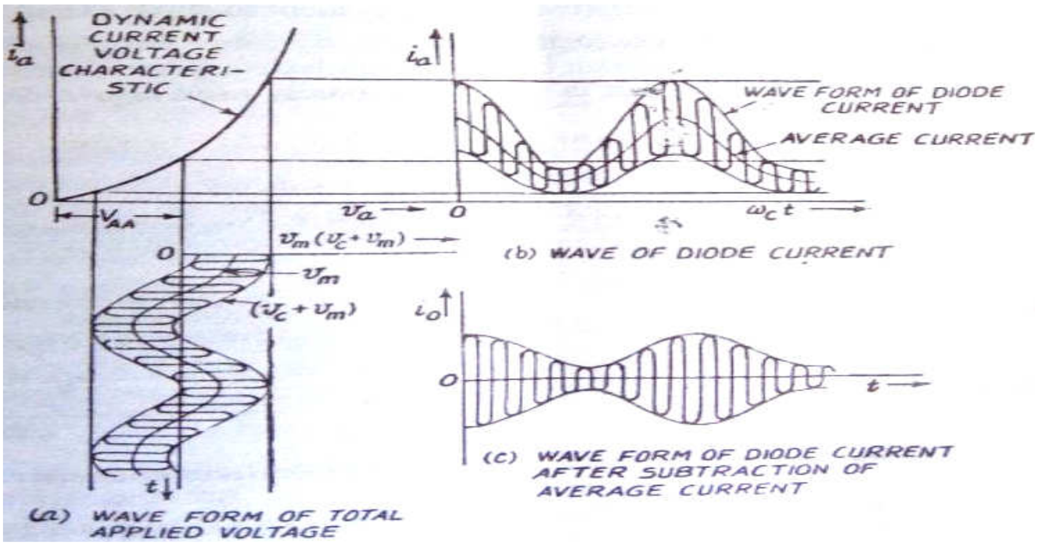


Figure. 4.12. Waveshapes of applied voltage and resulting currents in square law diode modulation

The operation of the diode square law modulation is illustrated diagrammatically in Figure 4.12. Because of the nonlinearity of the dynamic current-voltage characteristics of diode, the amplitude of the carrier component in the current waveform during the positive half cycle of modulating voltage to greater than the corresponding amplitude during the negative half cycle of the modulation voltage as shown in Figure.4.12. This waveform shown in Figure.4.12 corresponds to Equ. (4.76) for diode current. This current has an average component valuated over carrier frequency cycles. This average component consists of a steady component I_{dc} and component varying at the modulation frequency. Similar treatment applies in the case of a high vacuum diode of a semiconductor diode. However vacuum diode square law modulations are no longer used.

3.3.2. Van der Biji Modulation

From Section 3.9.1, we have established that the Square law modulation could be achieved by making use of some nonlinear-voltage characteristic of a vacuum diode. The carrier voltage of frequency ω_c and modulation voltage of frequency ω_m may each be applied either in the plate circuit or in the grid circuit. This presents the following four possibilities of square law modulation using triode.

i. Both carrier and modulating voltage applies in the grid circuit
ii. Both carrier and modulating voltage applied in the plate circuit.

iii. Modulating voltage in the grid circuit and the carrier voltage in the plate circuit.
iv. Modulating voltage in the plate circuit and the carrier voltage in the grid circuit.

Out of these methods, method (i) is commonly used and is called Van der Bijimodulation. Essentially, Van der Biji modulation is present, through in a very limited quantity, in every class *A* amplifier since nonlinearity of the dynamic transfer characteristic is always present to some extent.

The Van der Biji modulation uses the dynamic transfer characteristic of a class *A* amplifier whereas the grid bias modulation was once widely used though scarcely used these days. Figure 4.13 shows the basic circuit arrangement of Van der Biji Modulation with its dynamic transfer characteristic

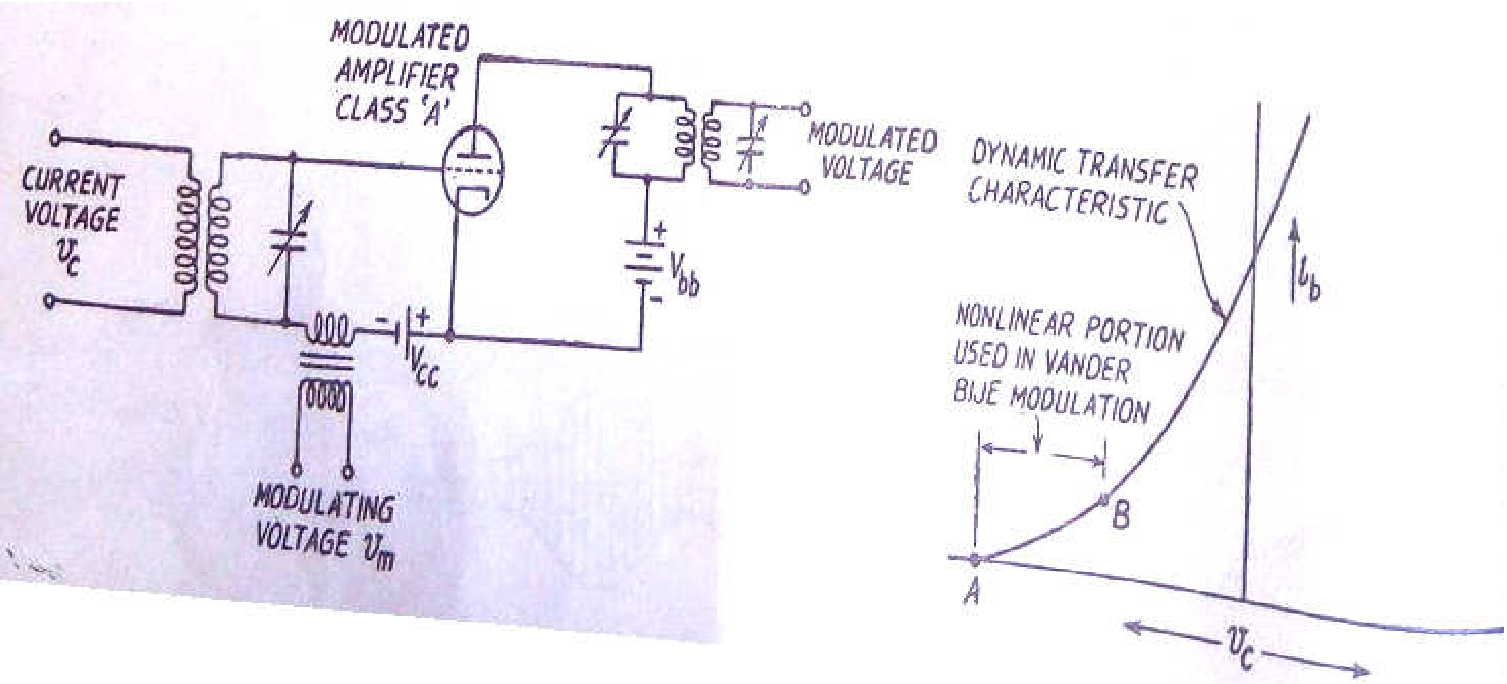


Figure 4.13. Basic circuit of Van der Biji Modulation with dynamic transfer characteristic.

Evidently both the carrier and the modulation voltage are applied in series in the grid circuit and the modulated voltage is developed across the tuned circuit in the plate circuit. On modulation, both the sidebands appear. Since in common radio broadcast practice, maximum modulation frequency of 5 kHz is used, the total frequency spectrum occupied by the modulation carrier voltage is 10 kHz and the tank circuit should have such selectivity as to include both sidebands up to 5kHz on either side. On the right hand side of the plot, the figure shows the dynamic transfer characteristic of the triode used in the circuit, the nonlinear portion AB of which is used in the modulation method.

Because of the nonlinear portion of the dynamic transfer characteristic used here, the AC plate current i_p may be expressed in terms of the A.C grid voltage $v_p$by the following Taylor series expressed:

$$i_p = a_1\, v_s + a_2\, v_s^2 \tag{4.81}$$

Where a_1 *and* $a_2$are Taylor series coefficients.

The total A.C grid voltage v_s is given by,
$v_s = v_c + v_m = V_c \cos\omega_c t + V_m \cos\omega_m t$
Where $v_c\ and\ v_m$are respectively the carrier and the modulation voltages.
Then,

$$i_p = a_1[V_c \cos\omega_c t + V_m \cos\ \omega_m t] + [V_c \cos\ \omega_c t + V_m \omega_m t]^2 \tag{4.82}$$

$$i_p = a_1 V_c \cos\omega_m t +\ V_m \cos\omega_m t + \frac{a_2 V_c^2}{2}(1 + \cos 2\,\omega_c t$$

$$+ \frac{a_2 V_m^2}{2}(1 + \cos 2\,\omega_c t) + a_2 V_m V_c[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t] \tag{4.83}$$

In general, the carrier frequency$\omega_c \gg \omega_m$so that the tuned circuit to carrier frequency $\omega_c$responds to the terms of frequency $\omega_c(\omega_c + \omega_m)$.The rest of the terms in Equ. (4.83) do not produce appreciation output across the tuned circuit. Hence the desired plate current component are given by,
$i_p = a_1 V_c \cos\omega_c t + a_2 V_m V_c[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$.

Or $$i_p = a_1\left[\frac{V_c[1+2a_2V_m]}{a_1}\right]\ \cos\omega_c t \tag{4.84}$$

or $$i_p = a_1 V_c[1 + m_a \cos\omega_m t]\ \cos\omega_c t \tag{4.85}$$

It is important to note that in Van der Biji modulation, grid current is very small so that little power is required from the source of carrier voltage and modulation voltage. This is an important advantage over the square law diode modulation where appreciable power from the carrier and modulation voltage source is required. The plate circuit efficiency of Van der Biji modulation is, however, small because of relatively large D.C plate current. Hence Van der Biji modulation is used for only low power applications such as in carrier current telephony or for measurement and control applications. For such low power applications low plate circuit efficiency is no serious drawback.Van der Biji modulation should be used only where percentage modulation is required at small power output.This is because, the ratio of the amplitudes of the undesired side frequencies to the desired side frequencies increase with the increase of modulation index.

3.4. Suppressed Carrier Balance Modulation

Recall from Equ. (3.3), we already established that the carrier $V_m \cos\ \omega_m t$ contain no information (or message). The information is contained in each of the sidebands. Accordingly, the carrier may be dropped or eliminated without losing any information. Furthermore, the carrier takes up a large proportion of the total modulated carrier power. In ordinary AM radio broadcast carrier is allowed along with the two sideband and the system is referred to as the double sideband (DSB) system. Carrier is necessary for reproduction of modulation signal in the detector stage of radio receiver. Now, in an ordinary AM broadcast, we allow the carrier to propagate and use a simple diode detector in the radio receiver. If we eliminate or suppress the carrier, the system becomes suppressed carrier double sideband system (DSB-SC). In this case, we need reinsert the carrier in the radio receiver and the circuit of the receive becomes complicated and costly. Hence, suppressed carrier DSB-SC may be used in point-to-point communication system.This suppressed carrier balanced modulators may be of the following types viz Suppressed carrier balanced modulator Transistors, Suppressed carrier balanced modulator Transistors, Suppressed Carrier Balanced Modulation using Field effect transistors (FETS) and Suppressed carrier Balance Modulation using two diodes. These are outside the scope of this work. Interested readers can refer to the established references. However, we still must note that the generation of suppressed carrier amplitude modulation voltage is done in balanced modulator which is briefly discussed below.

3.5. Balanced Modulator

In simple terms, a balanced modulator basically mixes the audio signal and the radio frequency carrier, but suppresses the carrier, leaving only the sidebands. The output from the balanced modulator is a double sideband suppressed carrier signal and it contains all the information that the AM signal has, but without the carrier. In order words, it is possible to generate an AM signal by taking the output from the balanced modulator and reinserting the carrier.To understand how a balanced modulator works it is necessary to revisit the amplitude modulationoperation in the previous units.

Figure 4.14a shows the basic circuit of a balanced modulator. It serves as a relatively simple method for generating a DSB-SC AM signal by using two conventional-AM modulators arranged in the configuration illustrated in Figure 4.14a using the square law AM modulator.

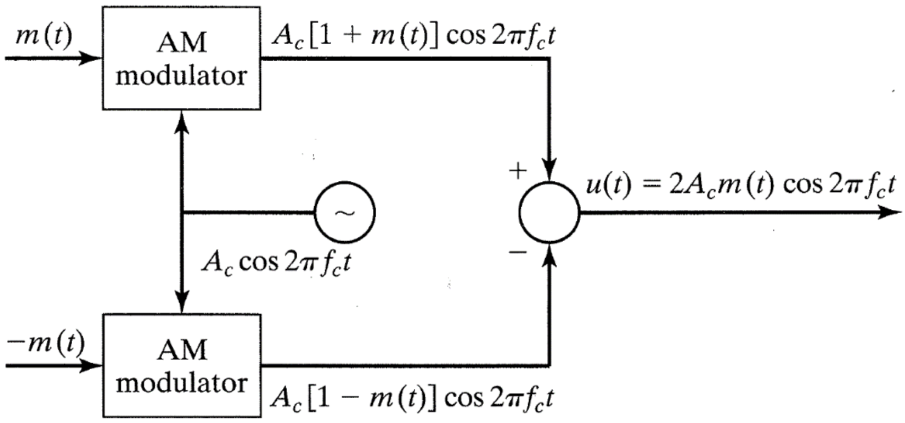


Figure 4.14a. Block diagram of a balanced modulator.

Now, the AM is a method of changing the strength of the carrier (amplitude) in sync with the modulating audio. While the power output does changes with modulation, because any AM modulator generates two sidebands, one above and one below the carrier. As power goes into these sidebands, the power output increases. The amplitude modulated signal, then, consists of a constant strength carrier and two sidebands. The sidebands carry the information and the carrier just goes along for the ride. The carrier can be removed at the transmitter and reinserted at the receiver to allow the transmitter to put all the power in the sidebands.Figure 4.14b shows the basic circuit of suppressed carrier balanced modulator using transistors.

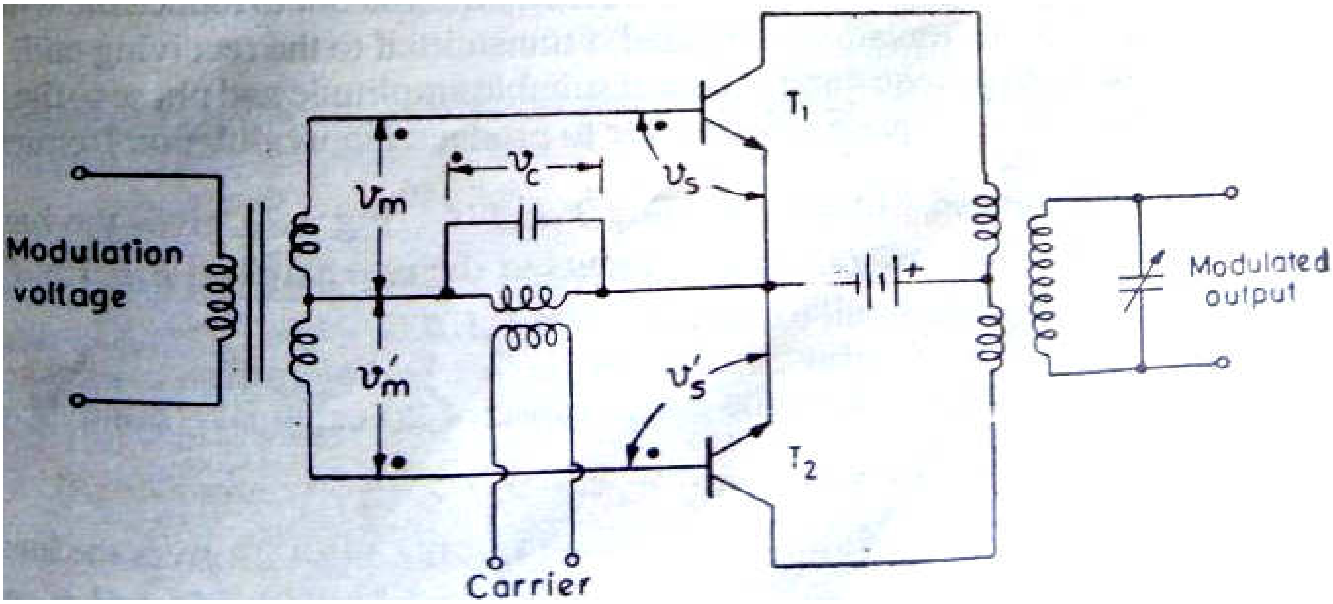


Figure. 4.14b. Basic circuit of suppressed carrier balanced modulator using transistors.

For a Suppressed carrier balanced modulator Transistors, Figure 4.14bgives the basic circuit. The carrier voltage is applied in parallel to the input of the two matched transistor T_1 and T_2 whereas the modulating voltage is applied in pushpull to the same two transistors. Suppression of carrier result in economy of power. Suppressed carrier balance modulator is commonly used in carrier current telephony in which one sideband is usually filtered out to reduce the width of the channel required for the transmission and the remaining sideband is transmitted to the receiving end. At the receiving end, a local oscillation add the carrier frequency voltage of suitable amplitude and phase to the sideband voltage and then the two together are fed to a square law detector to produce the modulation frequency voltage.For a Suppressed Carrier Balanced Modulation using FET, Figure 4.15 gives the basic circuit. Here again the carrier voltage v_c is applied in parallel to input of the matched FETs $T_1$and T_2 whereas the modulation voltage is applied in pushpull to the same two FETs.

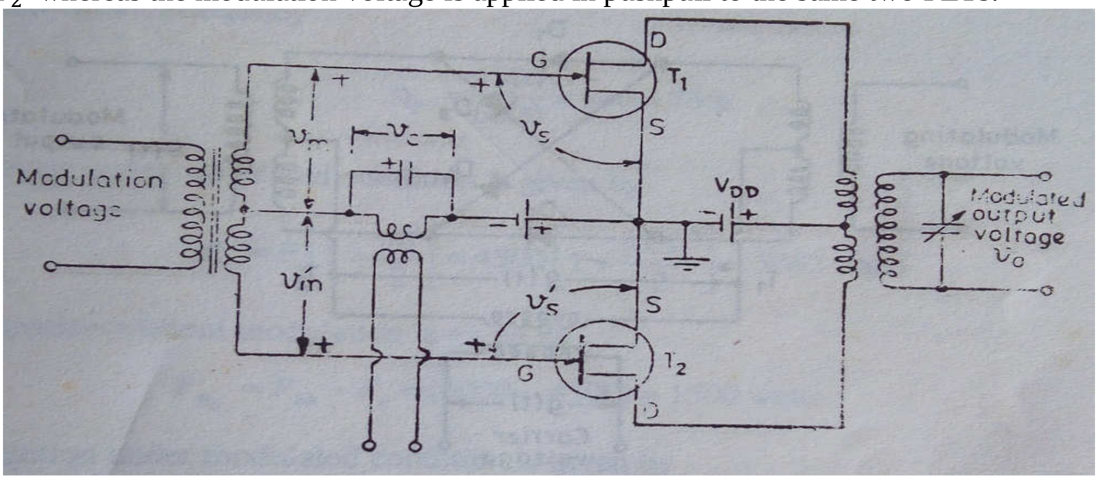


Figure. 4.15. Basic circuit of suppressed carrier balanced modulator using FET's.

For a Suppressed carrier Balance Modulation using Two Diodes, Figure 4.16 gives the basic circuit. In this case the carrier voltage $v_c$is applied in parallel to the two matched diodes $D_1$and D_2 while modulating voltage is applied in pull-push.

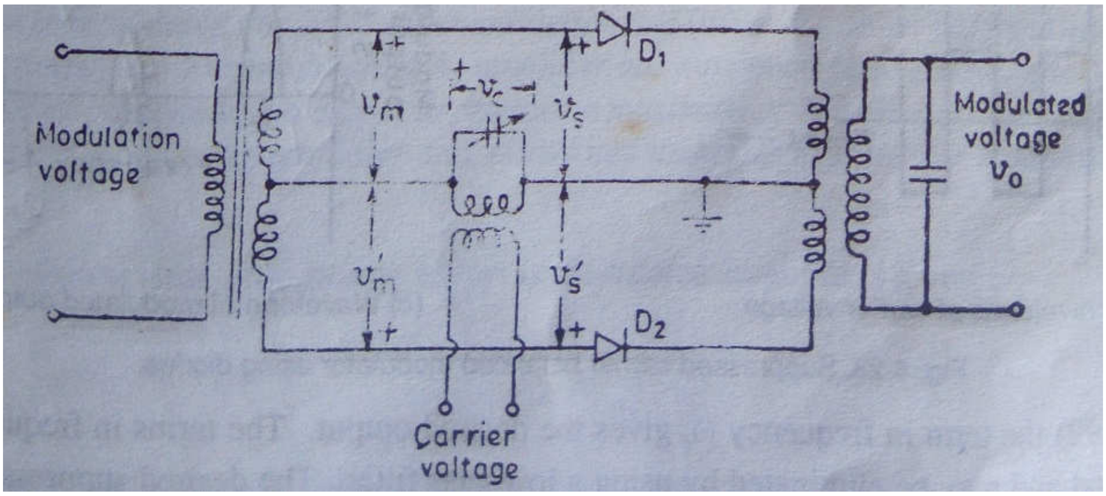


Figure. 4.16. Basic circuit of suppressed carrier balanced modulator using two diodes

The analysis is exactly similar to that for the circuit of Figure 4.15. In this circuit using two diodes we need a parallel tuned circuit in the output to remove the modulating frequency term from the output. However this removal of modulation frequency term my alternatively be achieved by using two more diodes to the diode balanced modulator of Figure 4.16 resulting in the circuit of Figure 4.17.

In the case of Suppressed Carrier Balanced Modulation using four diodes,Figure 4.17 gives the basic circuit. It uses four diode in a lattice configuration.The carrier *g(t)* has high amplitude. Then the carrier g(t) appearing across terminals $T_1$and T_2 is square wave signal having the same phase as g(t). During the positive half of *g(t),* diode $D_1$and D_2 conduct while diodes $D_3$and D_4 are baised to cut-off. Hence during this half cycle, signal *p(t)* reaches the output. During the other half cycle, Diodes $D_3$and D_4 conduct while diodes $D_1$and D_2 are baised off. Hence during this half cycle, singnal *p(t)* reaches the output with opposite sign. The plot of Figure 4.17 shows the waveforms of carrier voltage (left hand side) and the waveforms of modulated output voltage (right hand side).

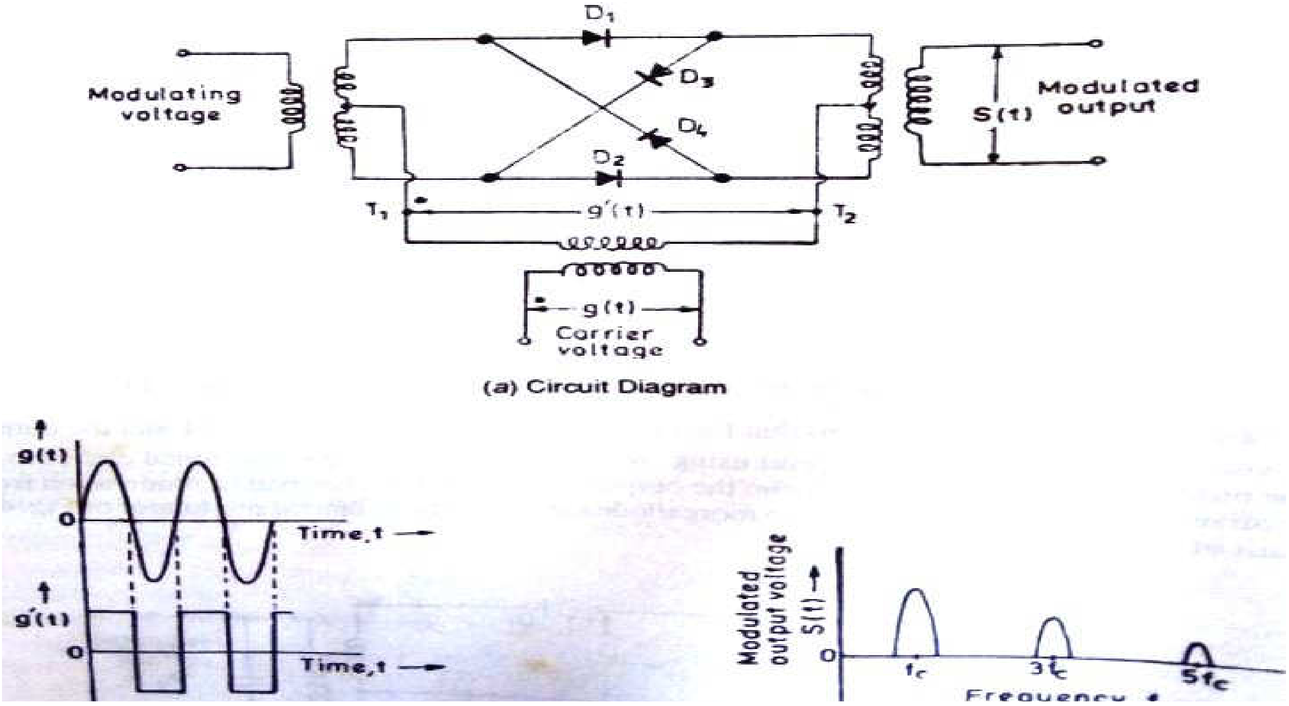


Figure. 4.17. Suppressed carrier balanced modulator using diodes

3.6. Ring Modulation

This is another form of balanced modulator. In this case, the carrier is applied to the mid-points of the input and output transformer. Ring modulation could be primarily seen as a signal-processing function in AM or an implementation of AM or frequency mixing, performed by multiplying two signals, where one is typically a sine wave or another simple waveform.Hence, a ring modulator is an electronic device for ring modulation, used for amplitude modulation or frequency mixing. It is referred to as "ring" modulation because the analog circuit of diodes originally used to implement this technique took the shape of a ring. This circuit is similar to a bridge rectifier, except that instead of the diodes facing "left" or "right", they go "clockwise" or "counterclockwise".

The carrier, which is AC, at a given time, makes one pair of diodes conduct, and reverse-biases the other pair. The conducting pair carries the signal from the left transformer secondary to the primary of the transformer at the right. If the left carrier terminal is positive, the top and bottom diodes conduct. If that terminal is negative, then the "side" diodes conduct, but create a polarity inversion between the transformers. This action is much like that of a double pole, double throw (DPDT) switch wired for reversing connections.

As shown in Figure 4.18a and b, the ring modulator includes an input stage, a ring of four diodes excited by a carrier signal, and an output stage. The input and output stages typically include transformers with center-taps towards the diode ring. It is important to note that while the diode ring has some similarities to a bridge rectifier the diodes in a ring modulator all point in the same clockwise or counter-clockwise direction. Other types of ring regulators are the Chopper and transistor based ring regulators

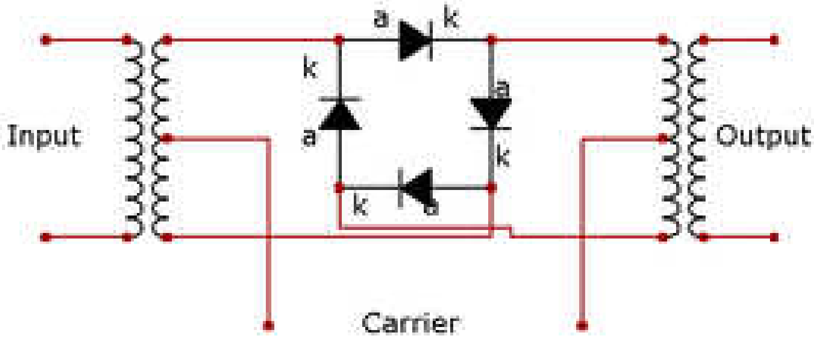


Figure 4.18a: Diode ring modulator.

Both Figure 4.18a and Figure 4.18billustrates another type of modulator (ring modulator) for generating a DSB-SC AM signal.A peculiar feature of the ring modulator is that it is bidirectional, i.e., the signal flow can be reversed allowing the same circuit with the same carrier to be used either as a modulator or demodulator, for example in low-cost radio transceivers. Ring modulators can frequency mix or heterodyne two waveforms, and output the sum and difference of the frequencies present in each waveform. This process of ring modulation produces a signal rich in partials. As well, neither the carrier nor the incoming signal are prominent in the outputs, and ideally, not at all.

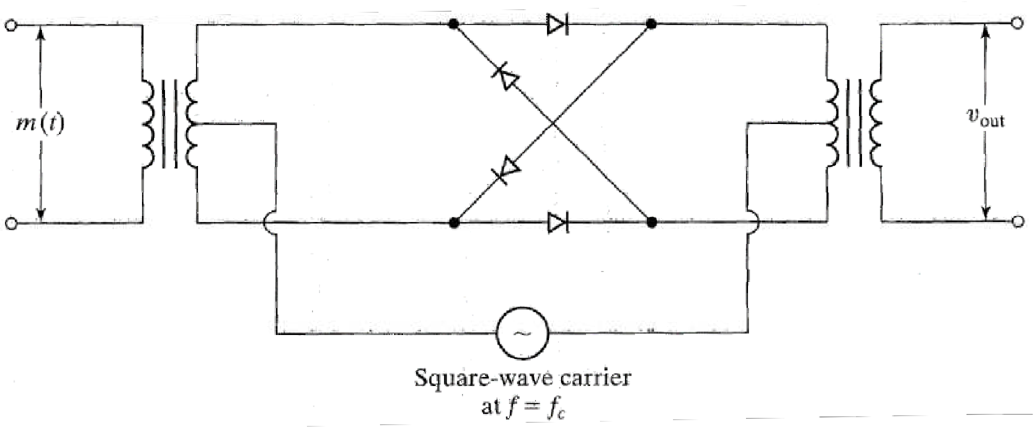


Figure 4.18b: Ring modulator for generating a DSB-SC AM signal

In this case, the switching of the diodes is controlled by a square wave of frequency *fc*, denoted as*c*(*t*), which is applied to the center taps of the two transformers

When *c*(*t*)> 0, the top and bottom diodes conduct, while the two diodes in the cross-arms are off. In this case, the message signal *m*(*t*)is multiplied by *+1*

When $c(t) < 0$, the diodes in the cross-arms of the ring conduct, while the other two diodes are switched off. In this case, the message signal $m(t)$is multiplied by *-1*.

Consequently, the operation of the ring modulator may be described mathematically as a multiplier of $m(t)$by the square-wave carrier $c(t)$, i.e.,$V_0(t) = M(t)C(t)$.

3.7. Switching Modulator

Another method for generating an AM-modulated signal is by means of a switching modulator. The generation of AM waves using the switching modulator could be explained by observing the switching modular diagram in Figure 4.19. The switching modulator using diode is depicted. The semiconductor diode is used as an ideal switch to which the signal $c(t) = A_c Cos(2\pi f_c t)$ and the information signal *m(t)* are simultaneously applied in Figure 4.19a and 4.19b.The total input for the diode at any instance is given by

$$c(t) = A_c Cos(2\pi f_c t) + M(t). \quad (4.86)$$

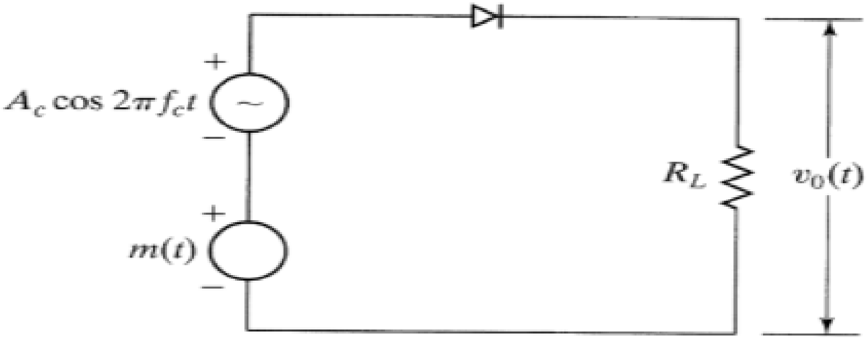


Figure 4.19a. Switching modulator and periodic switching signal.

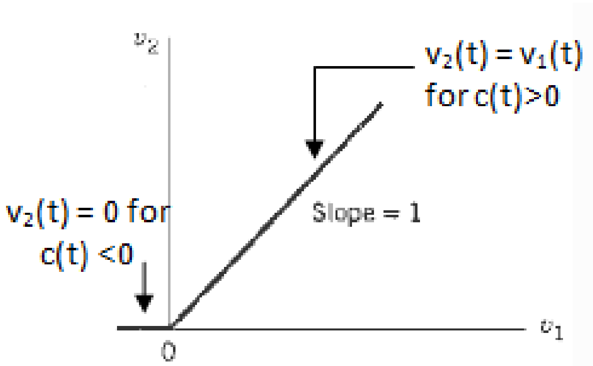


Figure 4.19b. Switching modulation characteristics.

The periodic function $s(t)$ can be represented in the Fourier series, while the desired AM-modulated signal is obtained by passing $V_0(t)$through a bandpass filter with the center frequency $f = f_c$and the bandwidth *2W.* At its output, we can then have the desired conventional AM signal.

4.0. Conclusion

We have seen that amplitude modulation (AM) classifications as used in electronic communication could be useful for transmitting information via a radio carrier wave. We have addressed all the stated objectives in this unit. We must note that without the established methods and techniques for amplitude modulation signal generation, we cannot appreciate sounds to be reproduced by a loudspeaker, or the light intensity of television pixels. A good understanding of the techniques of amplitude modulation methods has been laid so far.

5.0 Summary

So far, we have discussed the various techniques of amplitude modulation methods with their unique characteristics, merits and demerits. Some the AM methods previously explained includeLinear Modulation, Square Law Modulation, Linear Series Modulation, Cathode Modulation, Suppressed Modulation, Grid Bias Modulation, Screen Grid Modulation, Collector Modulation Van der Biji modulation, Suppressed carrier Balance Modulator, Ring and switching modulations. The Merits and demerits of linear series plate modulation as well as other types were discussed.

6.0 Tutor-Marked Assignment (TMA)

1. Explain the principal of grid bias modulation. Draw the basic circuit grid bias modulated amplifier and explain its working. Obtain expressions for D.C input power, A.C out power, plate efficiency and plate dissipation.

2. Discuss the relative performance of linear series plate modulation and grid bias modulation. Give the basic circuit cathode modulation and explain its working. Comment on the performance of the circuit relives to plat modulation and grid bias modulation.

3. Give the principle of square modulator. Draw the basic circuit of square law diode modulator. Describe the working of this circuit

4. Give the principle of Van der Biji Modulator. Give the basic circuit of this Modulator. What are the principle applications of the Modulator?

5. Give the principle of suppressed carrier balanced modulator. Draw the basic circuit of suppressed carrier balance modulator using transistor. Obtain expression for the output A.C voltage.What are its merits?

7.0 References/Further Readings

1. Seeley, S.: Electron Tube Circuits' McGraw Hill Book Company, New-York (1958).
2. Terman, F.E Electronic and Radio Engineering, McGraw Hill Book Company, New-York (1955).
3. Kennedy, G.: Electronic Communication System, Second edition, McGraw Hill Book Company, Kogakusha Ltd. Tokyo, 1977.

4. Ryder, J.D Electronic Fundamentals and Application, 5th edition prentice Hall of india private Ltd. New York
5. The ARRL Handbook for Radio Communication, American Radio Relay League, 2008, p. 9.30
6. Hayward, Campbell, and Larkin. Experimental Methods in RF Design, American Radio Relay League, 2003, p. 9.49
7.http://www.navodayaengg.in/wp-content/uploads/2015/09/U2-L3-Generation-of-AM-wave-square-law-modulator-switching-modulator1.pdf.

MODULE 3 FREQUENCY MODULATION/ULTRA HIGH IMPEDANCE SYSTEMS

Unit 1. Frequency Modulation Concepts
Unit 2 Frequency Modulation Transmitters
Unit 3Ultra High Frequency Tubes and Oscillators
Unit 4 High Impedance Circuits

UNIT 1 FREQUENCY MODULATION CONCEPTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
3.1 What is Frequency Modulation?
3.2 FM Wave Concepts
3.3 Waveform of Frequency Modulated Voltage
3.4 Mathematical Expression for Frequency Modulated Voltage
3.5 Frequency Spectrum of the FM wave
3.6 FM Bessel Functions Analysis
3.7 FM Bandwidth and Spectrum requirement
3.8 Phase Modulation
3. 8.1. Expression for Phase Modulated Voltage
3. 8.2. Comparison of Expressions for Phase Modulated Vs FM Voltages
3. 8.3. Comparison of Frequency Modulation and Amplitude Modulation
3.9 Noise and Frequency Modulation
3.9.1. Effect of Noise on carrier: Noise Triangle
3.9.2. Noise Triangle for $m_f > 1$
3.9.3. Pre-emphasis and De-emphasis
3.10. Alternative Sources of Interference
3.10.1. Adjacent Channel Interference
3.10.2. Co- channel Interference
3.11. Wideband FM and Narrowband FM
3.12. Stereophonic FM Multiplex Systems
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0 References/Further Readings.

1.0 INTRODUCTION

In the earlier modules, you have studied the various modes of AM which have some distinct characteristics. In these modes, usually each spectral component of the baseband signal occasions a rise to one or two spectral components in the modulated signal. These components are separated from the carrier by a frequency difference that is equal to frequency of the baseband component. Again, their modulators are such that the spectral components generated depend mainly on the baseband and carrier frequencies. In these types, the amplitude of the spectral components of the modulator output could depend on the amplitude of the input signal. Their system operations could be either additive, subtractive or even multiplicative, as such exhibits linear operations. This is why we call such type of modulation linear or amplitude modulations.In this unit, we shall look at another type of modulation devoid of the above described characteristics. This type is called angle modulation because in such modulating systems, in consonance with a carrier of constant amplitude, an angle which is a function of the baseband signal is made to respond to a baseband signal. Hence, angle modulation then involves staying at a constant carrier-amplitude while allowing the angle to fluctuate in proportion to the modulating signal. We shall look at frequency modulation as an angle modulation. Comparisons will be made with reference to linear modulation

2.0 Objectives

After going through this unit, you should be able to

- ✓ explain the concept ofFrequency modulation and compare with Angle modulation.
- ✓ discuss the various concepts FM signal wave.
- ✓ explain the merits and demerits of FM
- ✓ relate Stereophonic FM Multiplex Systems with Wideband FM and Narrowband FM
- ✓ discuss Pre-emphasis and De-emphasis in electronic communication systems
- ✓ relate noise triangle in relation to bandwidth optimization
- ✓ frequency spectrum computation using FM Bessel Functions Analysis
- ✓ solve numerical problems based on coherence.

5.0 Main Content

3.1 What is Frequency Modulation?

One might ask, what is? We define frequency modulation as a system whereby the amplitude of the carrier is made to be constant or fixed while its frequency is varied about its unmodulated frequency in a manner determined by the amplitude of the modulating signal. Thus the amplitude of the carrier does not change due to frequency modulation. This is an advantage since any incidental disturbance such as atmospheric disturbance or man-made static primarily appears in the form of variations of amplitude of the carrier voltage and may be eliminated in a frequency modulation receiver which is made insensitive to amplitude variation. Now, one might be wondering the implication this. Angle modulation has better detection performance than linear modulation though at the expense of greater channel bandwidth. It is also immune to channel non-linearity. This is because the information content is not conveyed in an envelope format which makes for a constant information flow devoid of distortion.

We can also say that Frequency modulation (FM) involves the encoding of information in a carrier wave by varying the instantaneous frequency of the wave. This contrasts with amplitude modulation, in which the amplitude of the carrier wave varies, while the frequency remains constant. In analog frequency modulation, such as FM radio broadcasting of an audio signal representing voice or music, the instantaneous frequency deviation, the difference between the frequency of the carrier and its center frequency, is proportional to the modulating signal.

Frequency modulation can be classified as narrowband if the change in the carrier frequency is about the same as the signal frequency, or as wideband if the change in the carrier frequency is much higher (modulation index >1) than the signal frequency. For example, narrowband FM is used for two way radio systems such as Family Radio Service, in which the carrier is allowed to deviate only 2.5 kHz above and below the center frequency with speech signals of no more than 3.5 kHz bandwidth. Wideband FM is used for FM broadcasting, in which music and speech are transmitted with up to 75 kHz deviation from the center frequency and carry audio with up to a 20-kHz bandwidth and subcarriers up to 92 kHz.

Figure1.1a shown the carrier voltage $v_{c.}$ for simplicity voltage is also assumed to be sinusoidal and let it be given by Equ 3.3 as previously established in this course work (Module 2, Unit 1).

i.e., $v_m = V_{m\cos\omega_m t}$

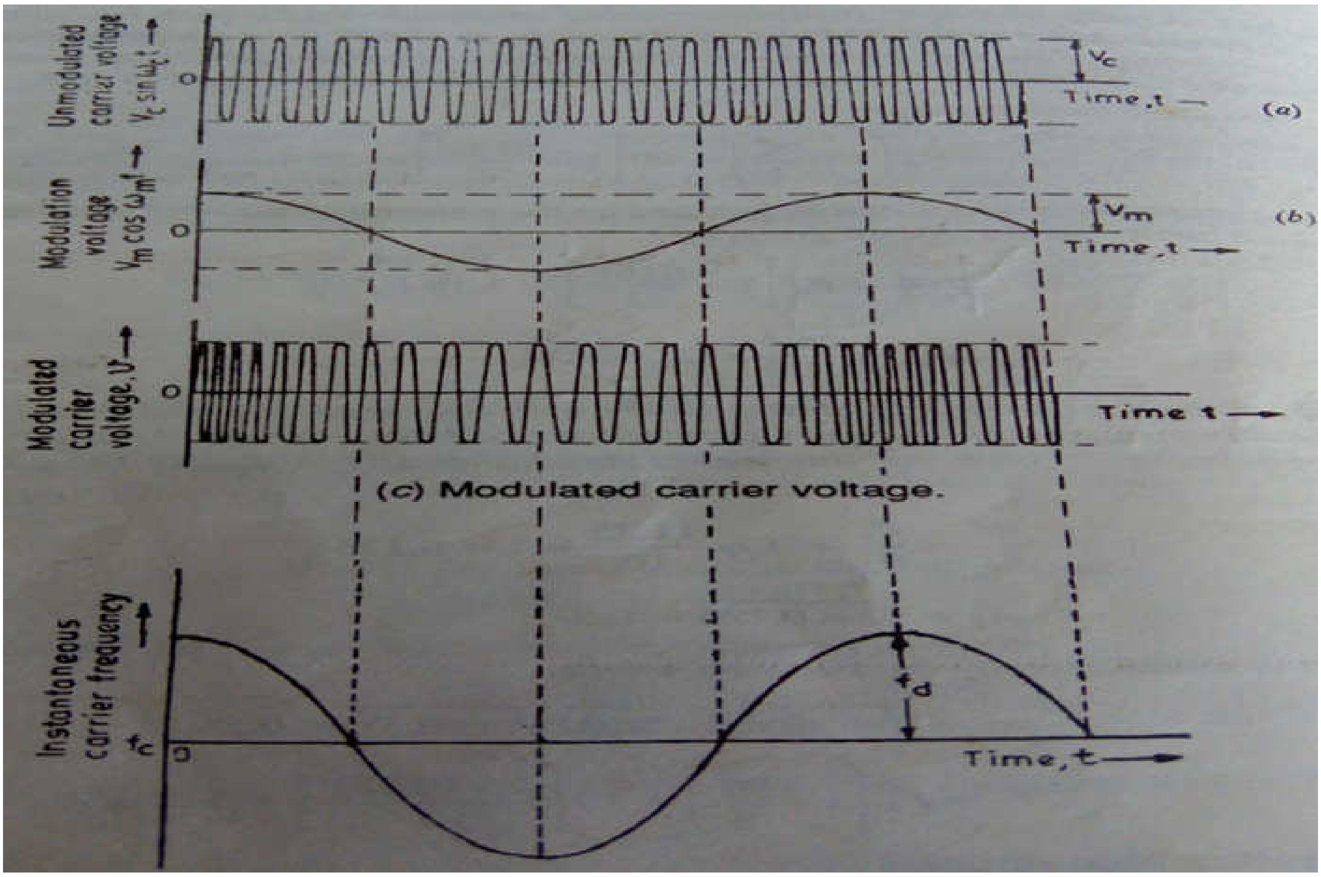


Figure1.1. Waveform of frequency modulated carrier voltage.

Figure1.1b shows the modulating voltage $v_{m.}$Figure 1.1c shows the resulting frequency modulated carrier voltage. Figure1.1d shows the variation of instantaneous carrier frequency with time. It may be seen that this frequency variation is identical in form with the variation with time of the modulating voltage. Evidently, the frequency variation called the frequency deviation is proportional to the instantaneous value of the modulating voltage.

The rate at which this frequency variation takes place is obviously equal to the modulating frequency. The amplitude of the frequency modulated carrier remains consists at all time during frequency modulation. Before proceeding further, we shall first describe critical terminologies that go with FM considering Figure 1.1.

3.2. FM Wave Signal Concepts

Considering Figure 1.1, it is very key to describe the convection terms associated with FM for ease of understanding and application.FM signals can be generated using either direct or indirect frequency modulation:

- Direct FM modulation can be achieved by directly feeding the message into the input of a voltage-controlled oscillator.
- For indirect FM modulation, the message signal is integrated to generate a phase-modulated signal. This is used to modulate a crystal-controlled oscillator, and the result is passed through a frequency multiplier to produce an FM signal. In this modulation, narrow band FM is generated leading to wide band FM later and hence the modulation is known as indirect FM modulation.

i. Frequency Deviation – This is referred to as the peak difference between the instantaneous frequency of a FM wave and the carrier frequency in a cycle of modulation. It is proportional to the amplitude of the modulating signal and may be increased by increasing the amplitude of the modulating signal.FM applications use peak deviations of 75 kHz (200 kHz spacing), 5 kHz (25 kHz spacing), 2.5 kHz (12.5 kHz spacing), and 2 kHz (8.33 kHz spacing).

ii. System Deviation- It is obvious that in any kind of frequency modulated communication system, there is a benched marked maximum frequency deviation referred to as rated system deviation. This captures the maximum possible amplitude of modulating function which can be accepted by the system. An over-modulation occurs when there is a large frequency variation. If the system deviation is referred to as f_d , the frequency deviation for other than maximum modulation signal is denoted by $K\, f_d$ where K is the ratio of the amplitude of the information or modulating signal and the maximum rated amplitude for the system.

iii. Modulation index- As in other modulation systems, the modulation index indicates by how much the modulated variable varies around its unmodulated level. It relates to variations in the carrier frequency. It is defined as the ration of the frequency deviation to the modulating frequency given by $\beta = \frac{K\, fd}{fm}$.

3.3. Waveform of Frequency Modulated Voltage

Let the carrier voltage v_c have amplitude v_c and frequency ω_c radians. This has already been given by Equ. 3.3 previously in Module 2, Unit 1 as $v_c = V_c \sin \omega_{c^t}$.
The above equation is the case without a phase angleϕ

3.4. Mathematical Expression for Frequency Modulated Voltage

Let the sinusoidal modulating voltage be given by the expression given in Equ 3.4in Module 2,

$$v_m = V_m \cos \omega_{m^t}$$

Where ω_m is the angular frequency of the modulating voltage in rad/sec and V_m is its amplitude in volts.
Let the carrier voltage be given by,

$$v_c = V_c \sin\ (\omega_{c^t} + \theta) \qquad (1.1)$$

Where ω_c is the angular frequency of the carrier in rad/sec?
$V_c$is the amplitude of the carrier in volts and θ_c is the phase angle in radians.
Let

$$\phi = \omega_{c^t} + \theta \qquad (1.2)$$

In Equ. (3.31),ϕ is the total instantaneous phase angle of the carrier voltage so that Equ. 1.1 may be put as,

$$v_c = V_c \sin \phi \qquad (1.3)$$

Obviously the angular frequency ω_c is related to the phase angle ϕ by the relation,

$$\omega_c = \frac{d\phi}{dt} \qquad (1.4)$$

Regarding FM, the frequency of the carrier no longer remains constant but varies with time in accordance with the instantaneous value of the modulating voltage. Thus, the frequency of the carrier voltage after frequency modulation is given by,

$$\omega = \omega_c + \kappa_f . v_m \qquad (1.5)$$

$$= \omega_c + \kappa_{f.} V_m \cos \omega_m t \qquad (1.6)$$

Where k_f is the constant of proportionality.

Integration of Equ. 1.6yields the phase angle of the modulated carrier voltage. Thus we get

$$\phi = \int \omega dt = \int [\omega_c + \kappa_{f.} V_m \cos \omega_m t]\, dt \qquad (1.7)$$

Or $\phi = \omega_c t + k_f V_m \frac{1}{\omega_m} \sin \omega_m t + \phi_1$ (1.8)

Where ϕ_1 is the constant of integration and represent a constant phase angle?

Angle ϕ_1 may be neglected in the following analysis since it is insignificant in the modulation process.

Hence the frequency modulated carrier voltage is given by,

$$v = V_c\, sin\, [\omega_c t + k_f \frac{V_m}{\omega_m} \sin \omega_m t] \qquad (1.9)$$

From Equ. 1.6 (3.35) instantaneous frequency modulated carrier voltage in Hz is given by,

$$f = \frac{\omega}{2\pi} = f_c + k_f \frac{V_m}{2\pi} \cos \omega_m t \qquad (1.10)$$

The maximum value of frequency is given by,

$$f_{max} = f_c + k_{f.} \frac{V_M}{2\pi} \qquad (1.11)$$

The minimum value of frequency is given by,

$$f_{min} = f_c - k_{f.} \frac{V_m}{2\pi} \qquad (1.12)$$

Thus the frequency deviation, i.e. the maximum variation in frequency from the mean value is given by,

$$f_d = f_{max} - f_c = f_c - f_{min} = k_f \frac{V_m}{2\pi} \qquad (1.13)$$

Modulation index m_f is the ratio of frequency deviation to modulation frequency and is also indicated by σ

Thus,

$$\delta = m_f = \frac{f_d}{f_m} = \frac{\omega_d}{\omega_m} = \frac{k_{f.} V_m}{\omega_m} \qquad (1.14)$$

Thus the expression for the frequency modulated voltage is given by,

$$v = V_c \sin\, (\omega_c t + m_f \sin \omega_m t) \qquad (1.15)$$

It may be noted that as the modulating frequency ω_m decreases and the modulating voltage amplitude remains constant, i.e. ω_d remains constant, the modulation index increases. This forms the basis for distinguishing frequency modulation from phase modulation.

Example 1.
In an FM system, the frequency deviation is 5 kHz when the modulating frequency is 400 Hz and the audio modulating voltage is 2V. Compute the modulation index. Also compute the frequency deviation and the modulation index if (i) AF voltage is increases to 6 V keeping modulation frequency unaltered (ii) AF voltage is increased to 8 V while the modulation frequency is reduced to 200Hz.

Solution.
With AF frequency 400 Hz and AF voltage 2 V, $f_d = 5kHz$
Hence modulation index

$$m_f = \frac{f_d}{f_m} = \frac{5kHz}{400Hz} = 12.5$$

Recall that $f_d \propto V_m.$
Hence ,

$$\frac{f_d}{V_m} = \frac{5kHz}{2\,volts} = 2.5 \text{ kHz/volt (of modulating voltage)}$$

Case 1 $V_m = 6$ V
Hence $f_d = 2.5 \times 6 = 15kHz$
Hence $m_f = \frac{f_d}{f_m} = \frac{15kHz}{400Hz} = 37.5$
Case11. $V_m = 8\,V; f_m = 200Hz$
Hence $f_d = 2.5 \times 8 = 20kHz$
Hence $m_f = \frac{f_d}{f_m} = \frac{20\,kHz}{200\,Hz} = 100.$

Example 2
An FM wave is represented by the voltage equation, $v = 20\sin(5 \times 10^8 t + 4\sin\ 1500\,t)$. Find the carrier and modulating frequencies, the modulation index and the maximum deviation of the FM.What power will this FM voltage dissipate in a 20 ohm resistor?

Solution.
FM voltage is given by,$V = V_c\, sin\,[\omega_c t + \frac{\omega_d}{\omega_m}\sin\omega_m t]$
Obviously then $f_c = \frac{5\times10^8}{2\pi}\ Hz = 7.96 \times 10^7\ Hz, f_m = \frac{1500}{2\pi}$Hz =238.7 Hz
Modulation index $m_f = 4.$
Frequency deviation $f_d = m_f.f_m = 4 \times 238.7\ Hz$

$$P = \frac{V_{rms}^2}{R} = \frac{(20/\sqrt{2})^2}{20} = \frac{400}{2\times20} = 10 \text{ Walts.}$$

3.5. Frequency Spectrum of the FM wave
In amplitude modulation, only two sideband terms are produced namely the upper sideband term and the lower sideband term. Situation is more complex in F.M. when a carrier voltage v_c is

frequency modulated by a single sinusoidal voltage $v_{m,}$ the modulated carrier voltage is given by Equ. 1.15 reproduced here:

$v = v_c \sin(\omega_{c^t} + \delta \sin \omega_{c^t})$ (1.16)

This equation involves sine of a sine. Hence the solution involves the use of Bessel functions. A simplified explanation of Bessel functions is given in Section 3.6 below.

3.6. FM Bessel Functions Analysis

For the case of a carrier modulated by a single sine wave, the resulting frequency spectrum can be calculated using Bessel functions of the first kind as a function of the sideband number and the modulation index. The carrier and sideband amplitudes are illustrated for different modulation indices of FM signals. For particular values of the modulation index, the carrier amplitude becomes zero and all the signal power is in the sidebands.Since the sidebands are on both sides of the carrier, their count is doubled, and then multiplied by the modulating frequency to find the bandwidth.

Knowing the modulation index, you can compute the number and amplitudes of the significant side-bands. This is done through a complex mathematical process known as the Bessel functions.

The typical Bessel function graph is shown in Table 1.1a. The left column gives the modulation index. The remaining column indicates the relative amplitudes of the carrier and the various pairs of side-bands. Any side-bands with relative carrier amplitude of less than 1% (0.01) have been eliminated. The total bandwidth of an FM signal can be determined by knowing the modulation index and using the Table.

For example, assuming the modulation index is 2. Referring to Table 1.1b (3.1b), you can see that this produces significant pairs of side-bands. The Bandwith can then be determined with the simple formula

BW = 2Fm x Number of significant side-bands.

Using the example above and assuming a highest modulating frequency of 2.5KHz, the bandwidth of the FM signal is given as

BW = 2(2.5)(4) = 20KHz.

An FM signal with a modulation index of 2 and a highest modulating frequency of 2.5KHz will then occupy a bandwidth of 20KHz.

Table 1.1a. Bessel function of the first kind

Modulation Index	Sidebands (Pairs)																
	Carrier	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th
0.00	1.00	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
0.25	0.98	0.12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
0.5	0.94	0.24	0.03	--	--	--	--	--	--	--	--	--	--	--	--	--	--
1.0	0.77	0.44	0.11	0.02	--	--	--	--	--	--	--	--	--	--	--	--	--
1.5	0.51	0.56	0.23	0.06	0.01	--	--	--	--	--	--	--	--	--	--	--	--
2.0	0.22	0.58	0.35	0.13	0.03	--	--	--	--	--	--	--	--	--	--	--	--
2.5	-0.06	0.50	0.45	0.22	0.07	0.02	--	--	--	--	--	--	--	--	--	--	--
3.0	-0.20	0.34	0.49	0.31	0.13	0.04	0.01	--	--	--	--	--	--	--	--	--	--
4.0	-0.40	-0.07	0.36	0.43	0.28	0.13	0.05	0.02	--	--	--	--	--	--	--	--	--
5.0	-0.18	-0.33	0.06	0.36	0.39	0.26	0.13	0.05	0.02		--	--	--	--	--	--	--
6.0	0.15	-0.28	-0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02	--	--	--	--	--	--	--
7.0	0.30	0.00	-0.30	-0.17	0.16	0.35	0.34	0.23	0.13	0.08	0.02	--	--	--	--	--	--
8.0	0.17	0.23	-0.11	-0.29	-0.10	0.19	0.34	0.32	0.22	0.13	0.06	0	--	--	--	--	--
9.0	-0.09	0.24	0.14	-0.18	-0.27	-0.06	0.20	0.33	0.30	0.21	0.12	0.1	0.03	0.01	--	--	--
10.0	-0.25	0.04	0.25	0.06	-0.22	-0.23	-0.01	0.22	0.31	0.29	0.20	0.1	0.06	0.03	0.01	--	--
12.0	-0.05	-0.22	-0.08	0.20	0.18	-0.07	-0.24	-0.17	0.05	0.23	0.30	0.3	0.20	0.12	0.07	0.03	0.01
15.0	-0.01	0.21	0.04	0.19	-0.12	0.13	0.21	0.03	-0.17	-0.22	-0.09	0.10	0.24	0.20	0.25	0.18	0.12

Table 1.1b.Bessel Function with Sideband amplitude

Modulation index	Sideband amplitude																
	Carrier	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0.00	1.00																
0.25	0.98	0.12															
0.5	0.94	0.24	0.03														
1.0	0.77	0.44	0.11	0.02													
1.5	0.51	0.56	0.23	0.06	0.01												
2.0	0.22	0.58	0.35	0.13	0.03												
2.41	0	0.52	0.43	0.20	0.06	0.02											
2.5	−0.05	0.50	0.45	0.22	0.07	0.02	0.01										
3.0	−0.26	0.34	0.49	0.31	0.13	0.04	0.01										
4.0	−0.40	−0.07	0.36	0.43	0.28	0.13	0.05	0.02									
5.0	−0.18	−0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02								
5.53	0	−0.34	−0.13	0.25	0.40	0.32	0.19	0.09	0.03	0.01							
6.0	0.15	−0.28	−0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02							
7.0	0.30	0.00	−0.30	−0.17	0.16	0.35	0.34	0.23	0.13	0.06	0.02						
8.0	0.17	0.23	−0.11	−0.29	−0.10	0.19	0.34	0.32	0.22	0.13	0.06	0.03					
8.65	0	0.27	0.06	−0.24	−0.23	0.03	0.26	0.34	0.28	0.18	0.10	0.05	0.02				
9.0	−0.09	0.25	0.14	−0.18	−0.27	−0.06	0.20	0.33	0.31	0.21	0.12	0.06	0.03	0.01			
10.0	−0.25	0.04	0.25	0.06	−0.22	−0.23	−0.01	0.22	0.32	0.29	0.21	0.12	0.06	0.03	0.01		
12.0	0.05	−0.22	−0.08	0.20	0.18	−0.07	−0.24	−0.17	0.05	0.23	0.30	0.27	0.20	0.12	0.07	0.03	0.01

On using Bessel functions, it may be shown that Equ.1.16 may be expanded to yield,

$v = V_c[\ J_0(\delta)\sin\omega_{ct} + J_1(\delta)\{\sin(\omega_c + \omega_m)^t - \sin(\omega_c - \omega_m)t\}$
$+ J_{2\,(\delta)}\{\sin(\omega_c + 2\omega_m)t + \sin(\omega_c - 2\omega_m)\,t\}$
$+ J_3(\delta)\{\sin(\omega_c + 3\omega_m)\,t - \sin(\omega_m - 3\omega_m)\,t\}$
$+J_4(\delta)\{\sin(\omega_c + 4\omega_m)t + \sin(\omega_c - 4\omega_m)\,t\}$
+………etc. (1.17)

Equ.1.17reveals that the sinusoidal carrier voltage after frequency modulation by another sinusoidal voltage consists of the following frequency terms,

(i) Carrier voltage reduced in magnitude by the factor $J_0(\delta)$.
(ii) Infinite number of sideband terms on both lower and upper frequency sides of the carrier frequency at intervals equal to the modulation frequency. The amplitude of these sideband terms are V_c multiplied by various Bessel function of the first and different orders denoted by the subscripts.

Bessel function $J_n\,(\delta)$ is given by,

$$J_n(\delta) = (\frac{\delta}{2})^n\left[\frac{1}{n!} - \frac{(\delta/2)^2}{1!(n+1)!} + \frac{(\delta/2)^4}{2!(n+2)!} + \frac{(\delta/2)^6}{3!(n+3)!} + \cdots\right] \quad (1.18)$$

In order to find the amplitude of a given pair of sideband terms and the magnitude of the carrier, it is necessary to known the value of the corresponding Bessel function. It is, however, not necessary to evaluate the Bessel function using Equ.1.18 since the magnitude of Bessel function of this type are readily available (in Table form) as in Table 1.1 or in graphical form as in Figure1.2.

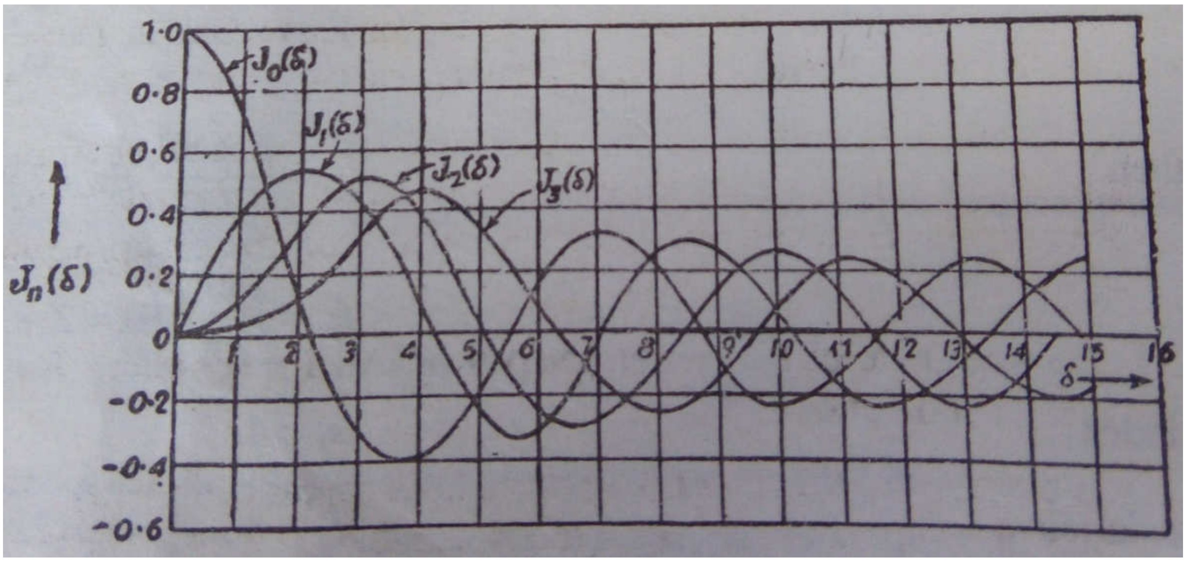


Figure 1.2. Bessel functions of the first kind and different order.

Recall that the frequency modulation can be classified as narrowband if the change in the carrier frequency is about the same as the signal frequency, or as wideband if the change in the carrier frequency is much higher (modulation index >1) than the signal frequency.

For example, narrowband FM is used for two way radio systems such as Family Radio Service, in which the carrier is allowed to deviate only 2.5 kHz above and below the center frequency with speech signals of no more than 3.5 kHz bandwidth. Wideband FM is used for FM broadcasting, in which music and speech are transmitted with up to 75 kHz deviation from the center frequency and carry audio with up to a 20-kHz bandwidth and subcarriers up to 92 kHz.The following conclusions are drawn from the foregoing study and from Table 1.1b and Figure1.2.

(i) In AM only three frequencies namely the carrier and the two sidebands are involved. Fm, on the other hand, has carrier and an infinite number of sideband terms recurring at frequency interval of $f_{m.}$

(ii) The J coefficients, in general, decrease with increase of order n but not in a simple way. Figure 1.2. Shows that the values of J fluctuate on both sides of zero and diminish gradually. Each J coefficient represents the amplitude of the corresponding pair of sideband terms. Hence the amplitudes of the sideband terms also eventually decrease but not past a certain value of $n.$ The modulation index δ (or m_f) thus determines the number of the significant sidebands, i.e. sidebands having amplitude at least 1% of the unmodulated carrier amplitude $V_{C.}$

(iii)The sidebands at equal frequency intervals from f_c have equal amplitudes. Thus the sideband distribution is symmetrical about the carrier frequency.

Table 1.1c. Bessel Function of n order

x	n or Order														
(m_f)	J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}	J_{11}	J_{12}	J_{13}	J_{14}
0.00	1.00	—	—	—	—	—	—	—	—	—	—	—	—	—	—
0.25	0.98	0.12	—	—	—	—	—	—	—	—	—	—	—	—	—
0.5	0.94	0.24	0.03	—	—	—	—	—	—	—	—	—	—	—	—
1.0	0.77	0.44	0.11	0.02	—	—	—	—	—	—	—	—	—	—	—
1.5	0.51	0.56	0.23	0.06	0.01	—	—	—	—	—	—	—	—	—	—
2.0	0.22	0.58	0.35	0.13	0.03	—	—	—	—	—	—	—	—	—	—
2.5	−0.05	0.50	0.45	0.22	0.07	0.02	—	—	—	—	—	—	—	—	—
3.0	−0.26	0.34	0.49	0.31	0.13	0.04	0.01	—	—	—	—	—	—	—	—
4.0	−0.40	−0.07	0.36	0.43	0.28	0.13	0.05	0.02	—	—	—	—	—	—	—
5.0	−0.18	−0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02	—	—	—	—	—	—
6.0	0.15	−0.28	−0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02	—	—	—	—	—
7.0	0.30	0.00	−0.30	−0.17	0.16	0.35	0.34	0.23	0.13	0.06	0.02	—	—	—	—
8.0	0.17	0.23	−0.11	−0.29	−0.10	0.19	0.34	0.32	0.22	0.13	0.06	0.03	—	—	—
9.0	−0.09	0.24	0.14	−0.18	−0.27	0.06	0.20	0.33	0.30	0.21	0.12	0.06	0.03	0.01	—
10.0	−0.25	0.04	0.25	0.06	−0.22	−0.23	−0.01	0.22	0.31	0.29	0.20	0.12	0.06	0.03	0.01
12.0	0.05	0.22	0.08	0.20	0.18	−0.07	−0.42	−0.17	−0.05	0.23	0.30	0.27	0.20	0.12	0.17
15.0	−0.01	0.21	0.04	−0.19	−0.12	0.13	0.21	0.03	−0.17	−0.22	−0.09	0.10	0.24	0.28	0.25

(iv)From Table 1.1c, we find that as δ increases, value of any particular J coefficient say J_{12} also increases. But δ is inversely proportional to the modulating frequency. Hence the relative amplitude of the distant sideband increases when the modulation frequency is reduced.

(v) In AM, with the increases of depth of modulation, the sideband power and hence the total power increase. In FM, on the other hand, the total transmitted power remains constant. However, with increased value of modulation index δ (orm_f), the required bandwidth for relatively undistorted signal gets increased.
(vi) From Equ. 1.17, it is evident that the theoretical bandwidth required in FM is infinite. In practice. However, bandwidth used is one including all the significant sidebands under most exacting condition. This implies that using maximum deviation by the highest modulation frequency, no significant sidebands are excluded.
(vii) In AM, the amplitude of the carrier component remains constant. But in FM, the carrier component is J_0 which a function is ofδ.
(viii) In FM, it is possible for the carrier component to disappear completely. From Figure. 1.2, we find that this happens for values of δ approximately equal to 24, 5.5, 8.6, 11.8 etc. these values of δ are called Eigen values.

3.7. FM Bandwidth and Spectrum requirement

The students must understand that we have regulatory bodies for FM signal propagation and any violation attracts serious sanctions. For example, the international regulations of FM broadcast have prescribed the following values for FM signals:

i. Maximum frequency deviations $f_d = \pm 75$ kHz;
ii. Allowable bandwidth per channel =200 kHz; (iii) frequency stability of carrier =$\pm$ 2 kHz.

From Table 3.1c, it is possible to find the magnitude of the carrier and each sideband term for any specific value of modulation indexδ. Thus considering the modulating frequency of 15 kHz, if the frequency deviation is $\pm$ 75 kHz, the deviation ratio δ =75/15=5. The Bessel functions that are then involved in representing the carrier and the sideband terms are J_n (5) where n varies from zero to infinity. Value of J_n (5) as a function of n may be obtained from graph such as in Figure 1.2J_n (5) as a function of n is plotted in Figure. 1.3.

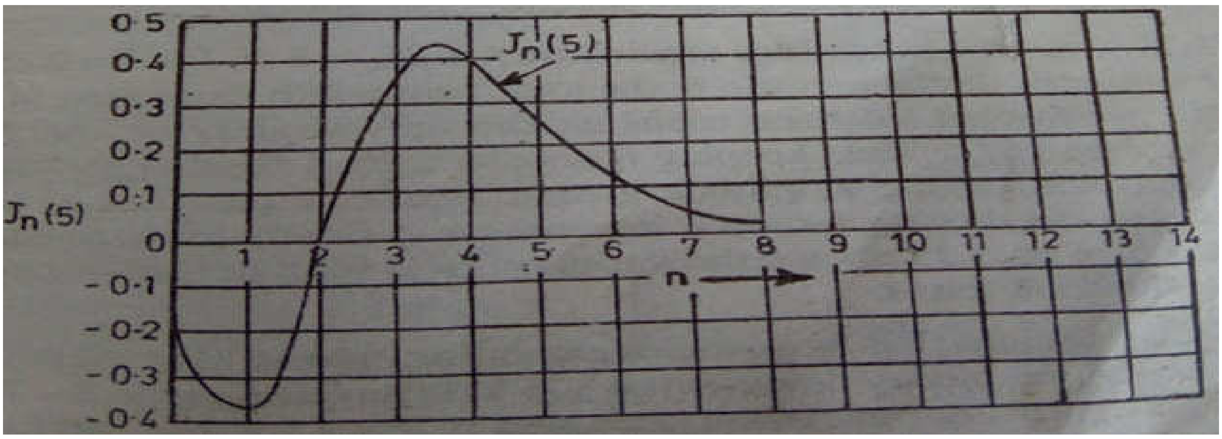


Figure 1.3. J_n (5) as function of n.

It is found that from Figure 1.3, J_0 (5) is about -0.18 amplifying that the carrier is about 0.18 $V_{c.}$ similar J_1 (5) is about -0.33 amplifying that the first order sideband terms have amplitudes of 0.33 V_c each function J_2 (5) is very small, being about +0.05. Further J_3 (5) is about 0.36 and J_4 (5) is about 0.39. Beyond the found order sideband terms, the amplitudes of the sideband terms fall off rapidly and all sideband terms beyond the eighth have amplitudes less than 1% of the unmodulated carrier amplitude $V_{c,}$ i.e. they are not significant at may be neglected. For deviation ratio of 5, the significant sideband terms extend upto the eighth, i.e. up to 8x15 = 120 kHz on either side of the carrier for modulating frequency of 15 kHz.

3.8. Phase Modulation

So far, we have looked at amplitude and frequency modulation. We shall now look at phase modulation. Phase modulation consists in varying the phase angle of a carrier voltage in accordance with the instantaneous value of the modulating voltage. In this regard, the amplitude and frequency of the carrier voltage remains unaltered after phase modulation. Like we have done in the previous discussions, let us now establish the expressions for Phase modulation voltage.

3.8.1. Expression for Phase Modulated Voltage

From Equ 1.1, let the carrier voltage be given as

$v_c = v_c \sin\ (\omega_{c^t} + \theta_o)$

And from Module 2, unit 1, let the modulating voltage be,

$v_m = V_m \sin \omega_{m^t}$

If the instantaneous phase of the carrier before modulation is given by,

$$\phi_c = \omega_{c^t} + \theta_0 \qquad (1.19)$$

After phase modulation, the instantaneous phase of the carrier is given by,

$$\theta\ (t) = \omega_{c^t} + \theta_0 + k_p v_m \qquad (1.20)$$

$$= \ \omega_{c^t} + \theta_0 + k_p V_m \sin \omega_{c^t} \qquad (1.21)$$

The phase modulated carrier voltage is then given by,

$$v = V_c \sin\ [\omega_{c^t} + \theta_0 + k_p V_m \sin \omega_{m^t}] \qquad (1.22)$$

In phase modulation process, the constant phase angle θ_0 plays no part and hence for the sake simplification θ_0 may be omitted. Then the modulated carrier voltage is given by,

$$v = V_c \sin\ [\omega_{c^t} + k_p V_m\ \sin \omega_{m^t}] \qquad (1.23)$$

Clearly, the maximum phase deviation is$k_{p.} V_m$. This may be indicated by $\emptyset_{m.}$ then the modulated voltage may be put as,

$$v = V_c \sin[\omega_{c^t} + \phi_m \sin \omega_{m^t}] \qquad (1.24)$$

or

$$v = V_c \sin\ [\omega_{c^t} + m_p \sin \omega_{m^t}] \qquad (1.25)$$

Where $m_p = \phi_m$ is the modulation index for phase modulation.

3.8.2. Comparison of Expressions for phase Modulated and Frequency Modulated Voltages

There are technical points that must be understood with regards to phase modulated and frequency modulated voltages. Now, considering Equ. 1.24 for phase modulated voltage, this is identical with Equ.1.15 for frequency modulated voltage except that instead of deviation ratio δ

(or m_f) in frequency modulation, we have maximum phase deviation $\phi_{m.}$ accordingly the frequency spectrum of the phase modulated voltage is similar to that of frequency modulated voltage. But the different is that:
In phase modulation,

$$\phi_m = k_p V_m \tag{1.26}$$

In frequency modulation,

$$\delta = \frac{k_f.V_m}{\omega_m} \tag{1.27}$$

From Equs. 1.26 and 1.27, we can see that in phase modulation, phase deviation ϕ_m is independent of modulating frequency $f_{m,}$ whereas in frequency modulation, deviation ratio δ is inversely proportional to modulating frequency $f_{m.}$hence, in phase modulation, for all values of modulating frequency, the phase deviation $\phi_m$remains constant.

However, for any single modulating frequency $f_{m,}$ the spectral distribution is similar to that in frequency modulation, i.e. the sideband terms appear at interval of $f_{m,}$ and have similar relative amplitude.
In order to compute the effect of variation of modulating frequency in the cases of frequency and phase modulation, we take the following illustration: let ϕ_m in the case of phase modulation be adjusted to be 750 radians. Evidently this situation for phase modulation corresponds to the condition of deviation ratio δ =750 in frequency modulation.

Let this deviation ratio δ=750 be obtained at a modulating frequency of 100 Hz. The significant sideband in frequency modulation will the extend upto 750th order. In the phase modulation as well as frequency modulation, the significant bandwidth at modulating frequency of 100Hz. What is occupied by the channel is 2×750× 0.1=150 kHz.

If now the modulating frequency is increased to say the maximum modulating frequency of 15 kHz, keeping V_m constant then in phase modulating $\emptyset_m$ still remains unaltered at the value 750 radians and significant bandwidth occupied by the channel is =$2 \times 750 \times 15 = 22{,}500\ kHz.$ On the other hand, in frequency modulation, for modulating frequency of 15 kHz keeping V_m unaltered, the deviation ratio δ reduces to (750 × 0.1 /15) =5.

For this value of deviation ratio δ =5, the significant sideband in frequency modulation extend upto the eighth order.
Hence the significant bandwidth occupied in frequency modulation with modulating frequency of 15 kHz is = 2 × 8 × 15 = 240 kHz. This bandwidth is very small as compared with the bandwidth of 22,500 kHz required in phase modulation. Thus, it is concluded that keeping the amplitude V_m of modulating voltage constant, as the modulating frequency f_m is increased, the significant bandwidth in phase modulation increases proportional to the modulating frequency whereas in frequency modulation the bandwidth requirement increases only slightly.

In another explanation, in phase modulation the significant bandwidth do not converge as the order of modulation frequency is increased whereas in frequency modulation, the significant, sidebands converge rapidly in spite of increases of modulating frequency. This constitutes one significant advantage of frequency modulation over phase modulation. It may be noted, however,

that if V_m is varied to keep δ and $\emptyset_m$ identical for all modulating frequencies, then both frequency modulation and phase modulation produce identical sidebands.
From Equ. 1.26 and 1.27, it is seen that expressions for δ and $\phi_m$for FM (frequency modulation) and PM (phase modulation) respectively are exactly similar in form except that he term ω_m appears in the denominator of expression for δ. This provides a simple means of covering PM to FM. All that is required to be done is to convert the modulating voltage to a form where ω_m appears in the denominator and then to use this modified modulating voltage to phase modulate the carrier. Term ω_m may be made to appear in denominator of the expression for modulation voltage by simple integration. This principle has been utilized in the Armstrong method of frequency modulation.

The above consideration lead to the following practical effect: if an FM signal is receiver on a PM receiver, the bass frequencies have considerably more deviation (of phase) than a PM transmitter would have given them. The output of a PM receiver is proportional to phase deviation. Hence the signal would appear bass booted. Alternatively PM signal received by an FM receiver would appear to be lacking in bass. This deficiency can, of course, be corrected by bass-boosting of the modulating signal prior to phase modulating. This forms the practical different between phase modulation and frequency modulation. But it is quite evident that one type of signal can be obtained from the other very simply.

Example 1
A20 MHz 5V carrier is modulated by a 500 Hz sine wave. The maximum frequency deviation is 15 kHz and the same modulation index is obtained for both FM and PM. Write expression for this modulated wave for (a) FM and (b) PM. Next if the modulating frequency is increased to 3 Hz, other things remaining the same, write new expressions for (c) FM and (d) PM.
Solution
The carrier frequency ω_c and the modulation frequency ω_m in radians/sec are
$\omega_c = 2\,\pi\ \times 20\ \times 10^6 = 1.25\ \times 10^8$ Radians/sec
$\omega_m = 2\pi\ \times 500 = 3141$ Radians/sec.
The modulation index is,$m = m_f = m_p = \frac{15000}{500} = 30$
Hence the expression for FM wave is,
$v = 5\,[1.25\ \times 10^8 t + 30 \sin 3141\ t]$
The expression for PM wave is,
$v = 5\,[1.25\ \times 10^8 t + 30 \sin 3141\ t]$.

The two expressions are identical since $m_f = m_{p.}$ when the modulation frequency is increased from 500 Hz to 3 kHz, i.e. made 6 times, the modulation index $m_{p.}$ for PM remains unaltered while the modulation indexm_f for FM reduces 6 fold (from 30 to 5). Hence the revised expressions for modulated carrier voltage are:

$\mathrm{FM{:}\, v = 5\,[1.25\ \times 10^8 t + 5 \sin\ 3141\ t]}$

$\mathrm{PM{:}\, v = 5\,[1.25\ \times 10^8\ t + 30 \sin\ 3141\ t]}$.

Demodulation in FM is very unique. Many FM detector circuits exist. A common method for recovering the information signal is through a Foster-Seeley discriminator. A phase-locked loop can be used as an FM demodulator. Slope detection demodulates an FM signal by using a tuned circuit which has its resonant frequency slightly offset from the carrier.

As the frequency rises and falls the tuned circuit provides changing amplitude of response, converting FM to AM. AM receivers may detect some FM transmissions by this means, although it does not provide an efficient means of detection for FM broadcasts. We shall now make comparison between FM and AM as regards communication systems below.

3.8.3 Comparison of Frequency Modulation and Amplitude Modulation

Let us now establish the merits, demerits and comparisons of both FM and AM with regard to electronic communication systems. We shall start with frequency modulation below.

The frequency modulation has the following advantages:

i. The amplitude of the frequency modulated wave is independent of the depth of modulation whereas in amplitude modulation, it is dependent on the modulation index. This permits the use of low-level modulation in FM transmitter and use of efficient class C amplifiers in all stages following the modulator. Further since all amplifiers handle constant power; the average power handle equals the peak power. In transmitter, the maximum power is four times the average power. Finally in FM, all the transmitted power useful whereas in AM, most of the power is carrier which does not contain any information.
ii. In FM there is a large decrease in noise and hence increase in signal-to-noise ratio. This result from the following two reasons: (a) there is less noise at carrier frequency at which FM is used (typically VHF and UHF) and (b) FM receivers can use amplitude limiters to remove all amplitude variations caused by noise.
iii. In FM noise may be further reduced by increase deviation. AM does not possess this feature.
iv. International Ratio consultative committee (CCIR) of the I.T.U. allows for a guard band between commercial FM stations. Thus there is less adjacent channel interference than in AM.
v. FM broadcast transmitters operate in the upper VHF range and in the UHF range. At these high frequencies, there is less noise than in the MF and HF ranges used for AM broadcast.
vi. Since FM broadcast takes place in the VHF and UHF ranges, the propagation used is space wave propagation. The radius of operation is limited to slightly more than the line of sight. This permits use` of several independent FM transmitters on the same frequency with negligible interference. This is not possible in AM.

The following are the disadvantages of FM

i. A much wider channel typically 200 kHz is needed in FM as against only 10 kHz in AM broadcast. This forms serious limitation of FM.
ii. FM transmitting and receiving equipments particularly for modulation and demodulation tend to be more complex and hence costly.
iii. In FM, the reception using conventional method is limited to line of sight. Thus the area of reception of FM is much smaller than for FM. This restriction is not due to the intrinsic properties of FM but due to the carrier frequencies (in VHF and UHF range) employed for its transmission. This is, no doubt, a disadvantage for FM mobile communications over a wide area but forms an advantage for co-channel allocations.

3.9. Noise and Frequency Modulation

It is very pertinent to note that the FM is much more immune to noise than AM and is significantly more immune than phase modulation. We hereunder examine the effect of noise on a carrier to establish the above facts and to determine the extent of noise improvement.

3.9.1. Effect of Noise on carrier (Noise Triangle)

When you consider a single noise frequency. This could normally affect the output of a receiver only if it falls within its passband. In that case, the carrier and noise voltages will mix and the difference frequency, if audible, will interfere with the reception of the wanted signal.Considering the single noise voltage vectorially, the noise vector gets superimposed on the carrier, rotating about it with relative angular velocity $(\omega_n - \omega_c)$ as shown in Figure1.4. The maximum deviation in the resultant amplitude from the average value is $V_n$whereas the maximum phase deviation is

$\phi = sin^{-1}(V_n/V_c)$. (1.58)

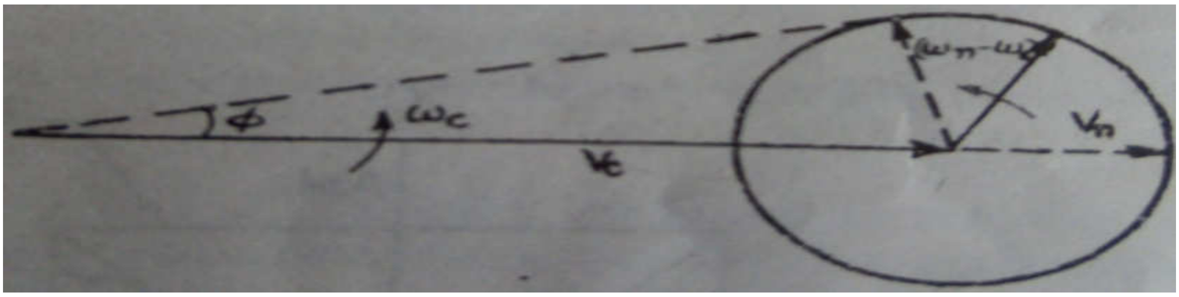


Figure1.4. Vector effect of noise on carrier.

Let the noise voltage amplitude be one-fourth of the carrier voltage amplitude. Then, for AM, the modulation $m_a = V_n/V_c = \frac{1}{4} = 0.25$ while the maximum phase deviation is ϕ =sin 0.25 /1 =14.5°. Let us also assume here that AM receiver responds only to amplitude channel and does not respond to phase changes.

We further assume that the FM receiver responds only to frequency changes and does not respond to amplitude changes since the amplitude limiter in FM receiver removes all the amplitude variations. We now proceed to assess the influence of phase changes on FM receiver and that of a amplitude changes on AM receiver.

We make this comparison under the most severe condition for FM. Let the modulating frequency be 15 kHz and let us assume for the sake of simplicity that the modulation index for both AM and FM be unity. Then in AM receiver noise-to-signal voltage ration will be 0.25/1 =0.25. Concerning FM, we convert the modulation index from unity devices to radians. Thus the ration is$14.5^\circ/57.3^\circ$ =0.253. Thus the noise-to-signal ration in FM is just slightly worse than in the case of AM.

We next study the performance when the modulation frequency has been altered from 15 kHz to the lowest value say 30 Hz. In AM, as the noise different frequency $(\omega_n - \omega_c)$and the modulating frequency are reduced from 15 kHz to 30Hz, there appears no difference in the relative noise, carrier and the modulating voltage amplitudes. In order words, in AM variation in the noise and modulating frequencies do not vary the noise-to signal ratio. In FM, on the other hand, since the ratio of noise to carrier voltage remains constant, the value of modulation index, i.e. maximum phase deviation due to noise also remains constant.

Thus, while the modulation index due to noise remains constant (as the noise sideband frequency is reduced), the modulation index caused by the signal goes on increasing in proportion to the reduction in modulation frequency. Hence in FM, the noise-to-signal ratio goes on reducing with modulation frequency. At the lowest modulation frequency of 30 Hz, the noise-to-signal ratio in FM is (0.253 x 30 /15000) =0.000505. Thus the noise-to-signal ratio reduces from 25.3 percent at 15 kHz to 0.05 per cent at 30 Hz.

We assume the noise frequency components to be evenly spread across the pass band of the receiver. Hence it is evident that the noise output from the receiver decreases uniformly with noise sideband frequency for FM. On the other hand, in AM it remains constant. Figure 3.13. (a) illustrates these noise sideband distributions for AM and FM. The triangular noise distribution for FM is referred as to the noise triangle. The noise sideband distribution for AM is a rectangle as shown in Figure 1.5a. From Figure 1.5a, we may conclude that the average voltage improvement for FM under these conditions is 3:1. Such a conclusion is valid for average audio frequency at which FM noise voltage appears to be half the AM noise voltage. In actual practice, however, the situation is more complex and the improvement obtainable in FM over AM is only a voltage ratio of $\sqrt{3}$: 1, i.e. power ratio of 3:1 or about 4.75 dB.

We have assumed in the beginning that the noise voltage is lower than the signal voltage. When two signals are simultaneously received, the amplitude limiter gets actuated by the stronger signal and it tends to reject the weaker signal. Accordingly if peak noise voltage exceeds the signal voltages, the signal will get excluded by the limiter. With very low signal-to-noise ratio, therefore, AM is superior to FM. The exact value of signal-to-noise voltage ratio at which FM becomes superior to AM depends on the value of FM modulation index. However, in general, FM becomes superior to AM when signal-to-noise voltage ratio becomes 4(12 dB) or more at the amplitude limiter level.

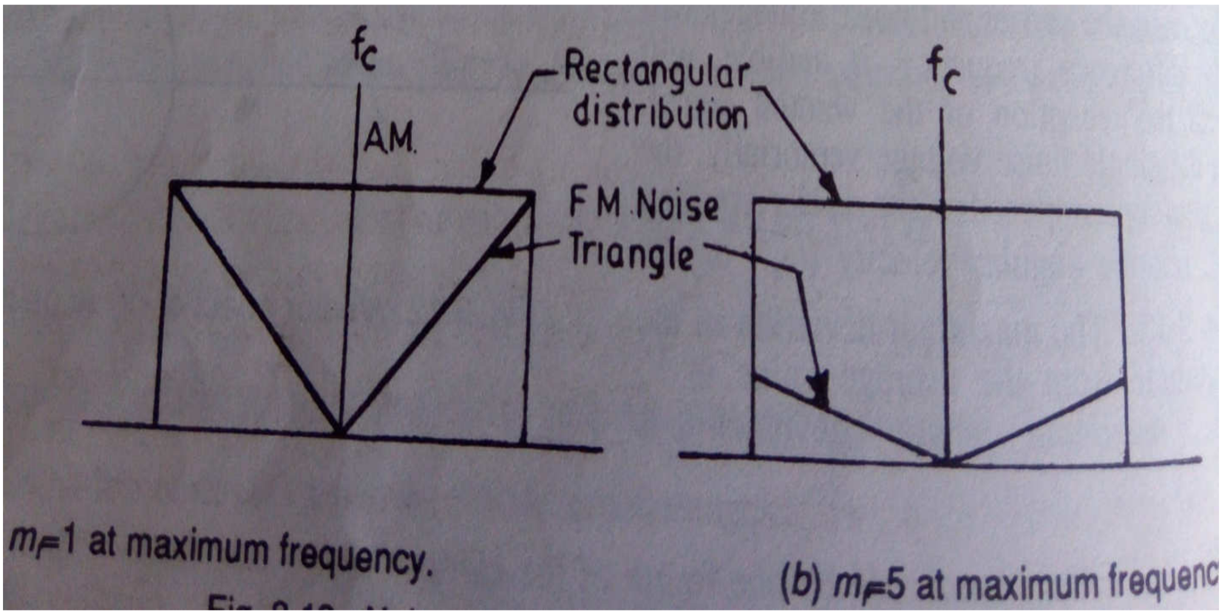


Figure 1.5. Noise sideband distribution in AM and FM.

3.9.2. Noise Triangle for $m_f > 1$

You must recall that in AM, the maximum permissible value of amplitude modulation is 1, i.e,m_a=1. In FM, there is no such limit. In FM, the limit is not the maximum frequency deviation. Thus for FM VHF broadcast, maximum frequency deviation is limited to 75 kHz. Hence using even the highest modulation frequency of 15 kHz, the modulation index in FM broadcast is as high as 5. At lower modulating frequencies, the modulation index is correspondingly higher. Thus, with modulation frequency of 1kHz, m_f is 75. The signal-to-noise voltage ratio in the output of the limiter in FM receiver will get increased in proportion to the modulation index. Thus, with m_f=5 (the highest permitted $m_f for f_m$=15 kHz), the signal-to-noise improvement is 5:1 in voltages and 25:1 in power (14 dB). No such improvement is possible for AM. With sufficient signal-to-noise ratio at the receiver input as assumed earlier, overall improvement secured in FM over AM is (4.75+14) = 18.75 DB. Figure 1.5b shows the noise triangle form_f=5.

From the above considerations, it becomes evident that in FM, we may use reduced bandwidth and thereby achieve higher signal-to-noise ratio. Such a trading of bandwidth is not possible in AM. It may also be noticed that just the increase of deviation (and hence the system bandwidth) in FM, does not necessarily mean that more random noise will be admitted. In fact this extra random noise produces no effect if the noise sideband frequency lie output the pass band of the receiver. Hence from this consideration, the maximum deviation and hence bandwidth, may be increased without fear. Phase modulation also has all the properties of FM except the noise triangle. Noise now phase modulates the carrier but there is no improvement as modulating and noise sideband frequencies are lowered. Thus under identical conditions, FM will be 4.7 dB better than PM regarding noise. It is for this reason that frequency modulation is preferred to phase modulation in practical transmitters.

In practice, however, in FM, bandwidth and maximum deviation cannot be increased indefinitely. Thus when a pulse is applied to a turned circuit, its peak amplitude is proportional to the square root of the bandwidth of the circuit. Similarly when a noise impulse is applied to the tuned circuit in the IF amplifier of an FM receiver, a large noise pulse results because of the unduly large bandwidth needed to accommodate the high deviation. When the magnitudes of noise pulses exceed about one-half of the carrier amplitude at the amplitude limiter, then the limiter function fails. When the noise pulse magnitude exceeds the carrier amplitude, the noise so to say captures the signal. The maximum deviation of 75 kHz is a compromise between the two extreme conditions described above.
It may be proved that when impulse noise amplitude$V_n < 0.5\ V_{c,}$ this impulse noise gets reduced in FM to the same extent as random noise. AM communication receivers use amplitude limiters. Such a limiter does not limit random noise at all and limits impulse by about 10 dB. Thus the FM system is better than the AM system in this regard as well.

3.9.3. Pre-emphasis and De-emphasis

We shall explain the meaning of these concepts in this Section and even in subsequent sections. Now, the noise triangle of Figure1.1 shows that the noise produces greater effect on the higher modulating frequencies than on the lower ones. It is, therefore, considered desirable that the higher modulating frequencies be artificially boosted up at the transmitter before modulation and correspondingly cut at the receiver after demodulation.This greatly improves the noise immunity at these higher modulation frequencies. This boosting of the higher modulation frequencies at the transmitter in any desired manner is called pre-emphasis while the relative attenuation of these higher modulation frequencies at the demodulator output in the receiver is called de-emphasis.

Figure 1.6ashows a typical pre-emphasis circuit while Figure 1.6b shows the corresponding de-emphasis circuit. The per-emphasis in USA FM broadcasting and in the sound transmission accompanying television has been standardized at 75μs whereas several other services, such as European and Australian broadcasting and TV sound transmission, use pre-emphasis of 50 μs. When using 75 μs pre-emphasis at transmitter, corresponding 75 μs de-emphasis must be used at the receiver. This is necessary in order that the relative amplitudes of the modulation frequency terms unaltered. Figure1.6ashows a C-R circuit of time constant RC =50 μs used for de-emphasis. These values of L,R and C may be altered to obtain pre-emphasis and de-emphasis of 75 μs. A 50 μs pre-emphasis corresponds to a frequency responds which is 3 dB up at the frequency whose time constant RC is 50 μs. This frequency is given by $f = R/2\pi L$ and is, therefore 3180Hz.

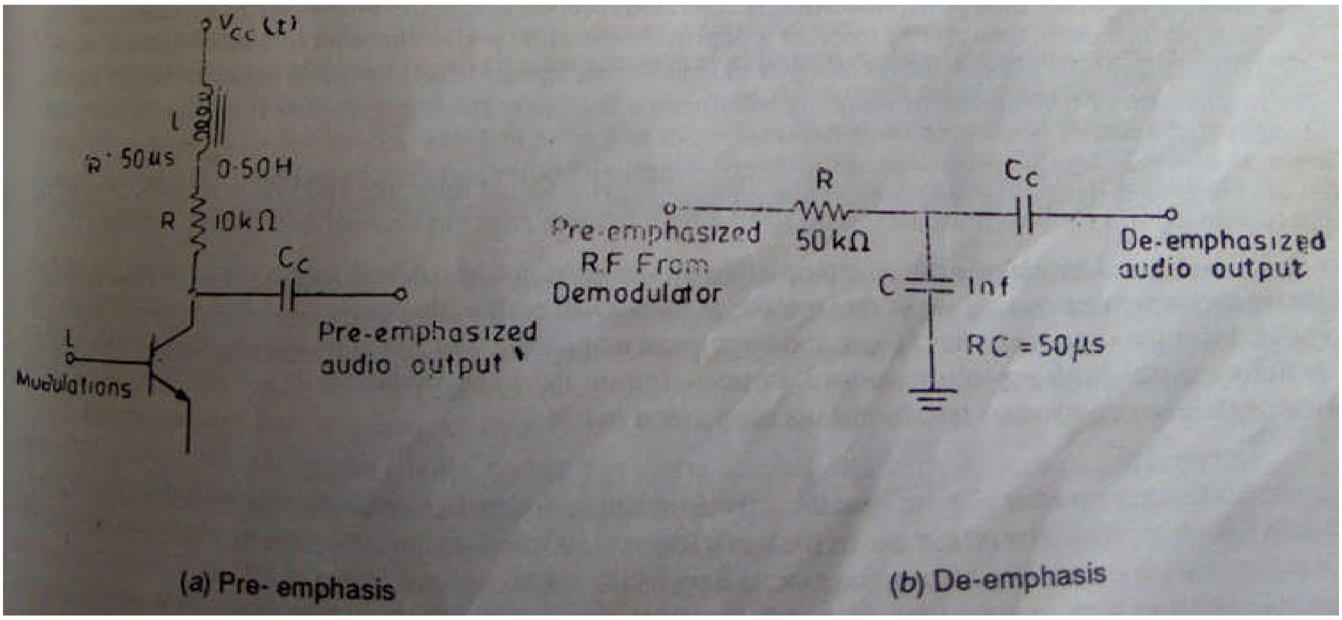


Figure1.6. 50µs pre-emphasis and de-emphasis circuits

Figure 1.7 shows this pre-emphasis curve for 50 μs A 50 μs de-emphasis corresponds to frequency response which is 3 dB down at the same frequency 3180 Hz ($3180\ Hz = \frac{1}{2\pi RC}$). Figure 1.7shows this de-emphasis curve also for 50 μs is 2120 Hz. We shall discuss the role of pre-emphasis below.

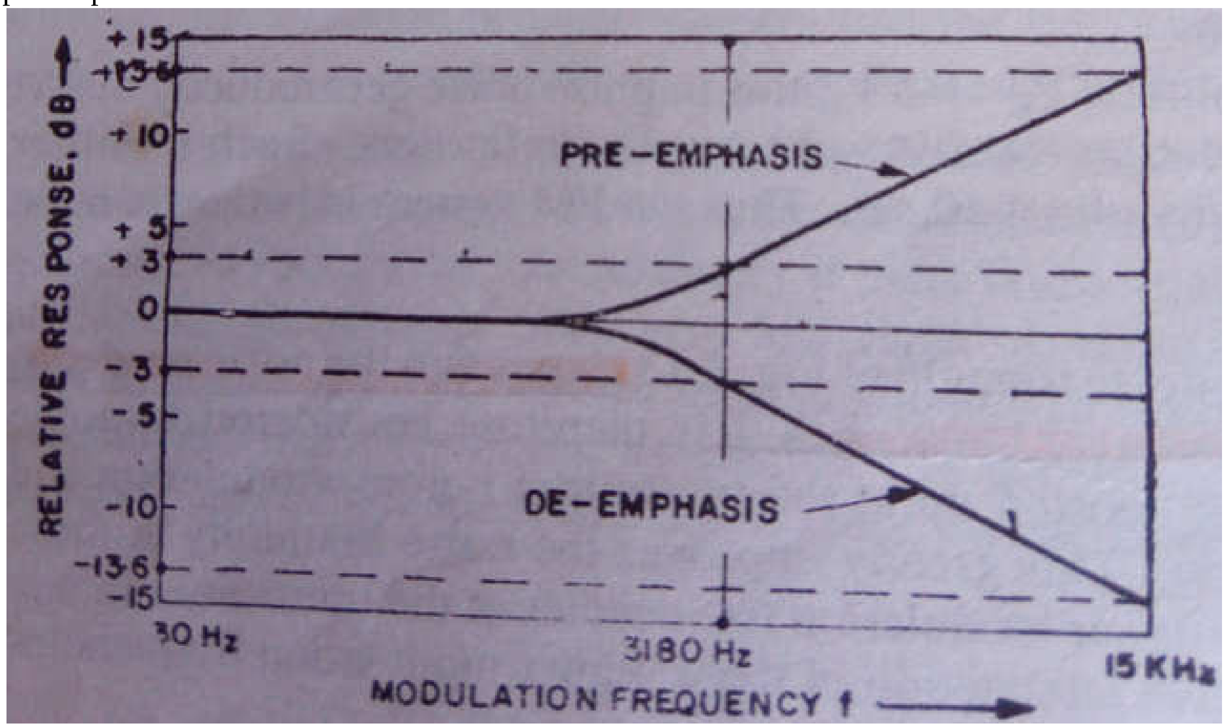


Figure 1.7. 50 μs pre-emphasis and de-emphasis curves.

Assuming there are two modulating signals having the same initial amplitude. Let one of these (at higher frequency) be pre-emphasized to say twice the amplitude whereas the order (at lower frequency) be left unaffected. Then the receiver will naturally de-emphasize the first signal by factor2 in order to ensure that both the signals have the same amplitude in the output of the receiver.
However, before demodulation, i.e. while susceptible to interference by noise the emphasized signals has twice the deviation it would have had without pre-emphasis and is, therefore, more immune to noise.Emphasis gives an improvement of about 4.5 dB in signal-to-noise ratio and relatively greater improvement with 75 μs emphasis.

It is important that care be taken such that the higher modulating frequencies are not over emphasized, or else over-modulation may take place with frequency deviation exceeding 75 kHz resulting in distortion. In practice, the order of pre-emphasis used is a compromise between protection for high modulating frequencies on one hand and the risk of over modulation on the great as in FM, since the modulating frequencies in AM are not unequally affected by noise.

3.10. Alternative Sources of Interference

Apart from noise, other forms interference are found in radio receivers such as (i) image frequency, (ii) transmitters operating on an adjacent channel and (iii) transmitters operating on the same channel. The first of the above mentioned interference is discussed later. The other two types are discussed here.

3.10.1. Adjacent Channel Interference

FM provides not only improvement in the S/N ratio but also greater discrimination against all other interfering signals, no matter what their source. We have already seen in the preceding section that FM with maximum deviation of 75 kHz and 50 μs pre-emphasis provides noise rejection of at least 23 dB better than AM. Accordingly if an AM receiver needs S/N ration of only 37 (=60 -23) dB. This result irrespective of whether the interfering signal is due to noise or adjacent channel signal. The mechanism of FM limiter of reducing interference is exactly the same whether it is noise or adjacent channel signal.
Further each FM broadcast channel occupies 200 kHz. Out of this, only 180 kHz is actually used while the remaining 20 kHz constitute the guard band which reduces adjacent channel interference further.

3.10.2. Co- channel Interference

The amplitude limiter used in FM receiver passes the stronger signal but eliminates the weaker one. It is for this reason, as mentioned earlier, that noise reduction is obtained in FM provided that the signal is at least twice the noise peak amplitude. For the same reason, a relatively weaker interfering signal from any other transmitter operating on the same frequency as the desired one, will be attenuated. Thus co-channel interference is suppressed in FM.

The possibility of co-channel interference arises in practice when a mobile receiver travels from one transmitter towards another operating on the same frequency. Interesting phenomenon of capture takes place. Thus in FM, as the mobile receiver moves from one transmitter to the second, the second transmitter is virtually inaudible causing practically no interference so long as the signal voltage from the second transmitter is less than about half of that from the first.

Beyond this point, the second transmitter becomes quite audible in the background and eventually predominates thereby excluding the first transmitter. Thus, the mobile receiver gets captured the second transmitter. When the receiver is in the transition region, i.e. roughly in the center zone and fading takes place, then signals from the two transmitters are alternately stronger. Thus the receiver is captured alternately by one transmitter and then the other. This switching from one transmitter to the other is most annoying and does not happen in AM system. In AM receiver, capture effect is not obtainable. In this case, as the mobile receiver travels from one transmitter to the second, the nearer transmitter would always predominate while the other one would be heard as quite significant interference although it may be very distant.

3.11.Wideband FM and Narrowband FM

Wideband FM is one in which the modulation index normally exceeds unity. This is the one which has so far been discussed. In wideband broadcast FM the modulating frequencies extend from 30 Hz to 15 kHz while the maximum permissible deviation is 75 kHz. Hence the maximum modulation index ranges from 5 to2500. In narrowband FM, the modulation index is usually about unity since the maximum modulating frequency is usually 3 kHz and the maximum deviation is usually 5 kHz.

The bandwidth used in any FM system depends on the application. Large frequency deviation and consequent large bandwidth has the advantage that noise is better suppressed. Care must; however, be taken to ensure that impulse noise peaks do not become excessive. Wideband FM system however, need large bandwidth, typically 15 times that of narrow bandwidth system. From these considerations, the wideband FM system are used in entertainment broadcasting whereas narrowband FM systems are used for communications.

Thus, narrowband FM is used by the mobile communications services such as police wireless, ambulances, taxicabs, short range VHF ship-to-shore sources and defenses. In all such cases, higher audio frequencies are attenuated as in the case of long distance telephone services but the resulting truncated speech is still perfectly clear and intelligible. The maximum deviation permitted in such services ranges from 5 to 10 kHz. Narrowband systems with even lower deviations are sometimes used. Pre-emphasis and de-emphasis are used in such narrow band FM system also, as in other FM systems.

3.12. Stereophonic FM Multiplex Systems

In stereo FM system, enough information is sent to the receiver to enable it to reproduce the original stereo material. Such a stereo FM system came into commercial usage in 1961, several years after commercial monaural FM system. This stereo FM system had to be made compatible with the existing monaural FM system. This resulted in an unduly complicated stereo FM system. Had we switched to FM stereo system straightway, the system would have been considerably simpler. The situation is similar to that existing for TV system in which colour TV came later than the monochrome TV. Thus, from the consideration of compatibility with the existing commercial monaural FM system, it is not possible in stereo FM to use a two channel system with a left channel and a right channel transmitted simultaneously and independently.

Figure 1.8 gives the block diagram of the stereo FM multiplex generator with optional subsidiary communication Authorization (SCA), and follows the standards laid down by FCC in 1961. Hence the two channel outputs L and R are fed to a matrix which produces sum (L+R) and the difference (L-R) signals. The sum signals (L+R) of frequency 50 Hz. To 15 kHz modulates the carrier in the same way as the signal in a monaural transmission. The sum signal is received by the monaural FM receiver tuned to the stereo transmission and reproduced at its output as if it were the complete signal sent.

The different signal (L-R) after demodulation in a stereo receiver and on being added to the sum signal (L + R) produces the left channel signal while the difference the sum (L +R) and the difference (L – R) signals produces the right channel. We here study as to how the difference signal is impressed on the carrier.

Both the sum (L + R) and the difference (L – R) signal occupy the same frequency range of 50 Hz to 15 kHz. If impressed together on modulator, these two signals, being in the same frequency range will get mixed up. To avoid this, the difference signal is shifted in frequency from 50 -15000 Hz to a higher value. Such a stacking of signals in different parts of the frequency spectrum is referred to as frequency multiplexing and hence the name of this system. The difference signal amplitude modulates a sub-carrier at 38 kHz in a suppressed carrier balanced modulator. At the output of the balanced modulator, the subcarrier gets suppressed while the two sidebands alone are obtained extend upto 15 kHz on either side of the sub-carrier of 38 kHz and thereby occupy frequency range extending from 23 kHz to 53 kHz. The sideband is added to the sum signal (L + R) occupying frequency range 50-15000 Hz and a 19 kHz sub-carrier signal. This combined signal then frequency modulates the carrier. No interference between the sum and the difference channel signal thus occurs since they are stacked at different frequency slots. This frequency modulated carrier is then transmitted and receiver in the corresponding stereo FM receiver. In the monaural receiver, the audio frequency band corresponding to this difference channel (23-53 kHz) is filtered out and discarded. In a stereo FM receiver, on the other hand, the wanted difference signal is extracted. To facilitate the tricky extraction of difference signal and the demodulation process, a sub-carrier of 15 kHz (half the suppressed sub-carrier frequency) is used for demodulation. The sum and difference signals are then added in one combining network and subtracted in another combining network to yield the left and the right channels. These two separate bands of signals are amplified in separate chains of audio amplifier and reproduced as the two channels of the system.

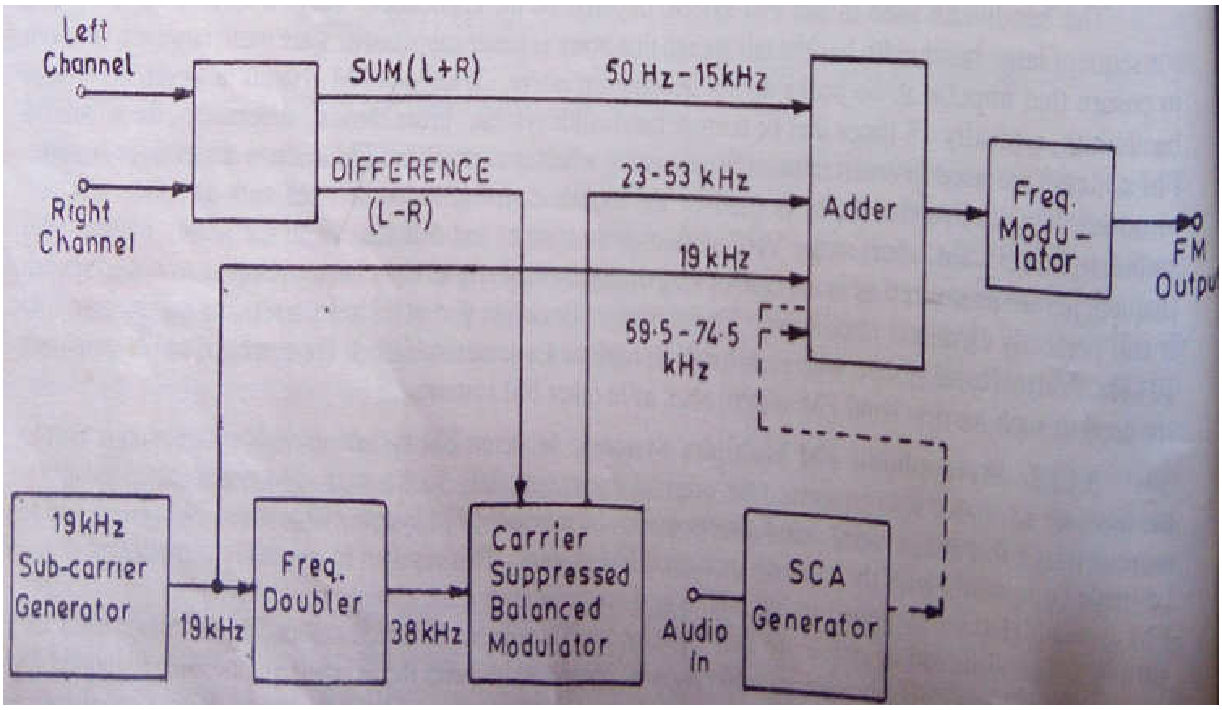


Figure 1.8. Stereo FM multiplex generator.

We can see that the Stereo FM multiplexing used here differs from conventional multiplexing in the fact that here the lowest audio frequency is 50 Hz compared with 300 Hz normally encountered in communication voice channel multiplexing. This low minimum frequency of 50 Hz makes it difficult to suppress the unwanted sideband without affecting the wanted one. The second problem which arises is that it is difficult to extract the pilot carrier in the receiver. However it is imperative to use some form of carrier to ensure that the receiver has a stable reference frequency for demodulation failing which distortion of the difference signal occurs.
The 38 kHz sub-carrier is generated from 19 kHz oscillator for a specific reason. It avoids the difficulty of having to extract the pilot carrier from among the close sideband frequencies in the receiver.
As shown in Figure 1.8, the output of 19 kHz sub-carrier generator is added to the sum and the difference signals in the output adder preceding the frequency modulator. This frequency of 19 kHz neatly fits into the space between the top of the sum signal and bottom of the difference signal and is at the same time far away from each of them. Hence, there is no difficulty in extracting it in receiver and also in suppressing the unwanted sideband without affecting the wanted one. Thus both the problems mentioned above are overcome through use of 19 kHz sub-carrier.

In the receiver, the 19 kHz signal is doubled in frequency and is then reinserted as the carrier for the difference signal. This 38 kHz sub-carrier is reinserted at 10% level which level is adequate but at the same time not so large as to draw undue power from the sum and difference signals or to cause over modulation.

In the system described above, a subsidiary communication Authorization (SCA) signal may also be transmitted shown by dashed line in Figure1.8. This SCA may then provide a second medium quality transmission used as background music in restaurants, stores etc. The SCA transmission uses a sub-carrier of 67 kHz frequency modulated to a depth of ± 7.5 kHz by the audio signal resulting in frequency band extending from 59.5 to 74.5 kHz which fits in the frequency spectrum sufficiently above the difference signal so as not to interfere with it. Figure 1.9 shows the overall frequency allocation within the modulating signal of an FM stereo multiple transmissions with SCA. However, care has been taken to keep the amplitude of the sum and the difference signals about 10% in the presence of SCA failing which over modulation of the main carrier may result. Figure 1.9 shows the spectrum of stereo FM multiplex modulating signal with optional SCA.

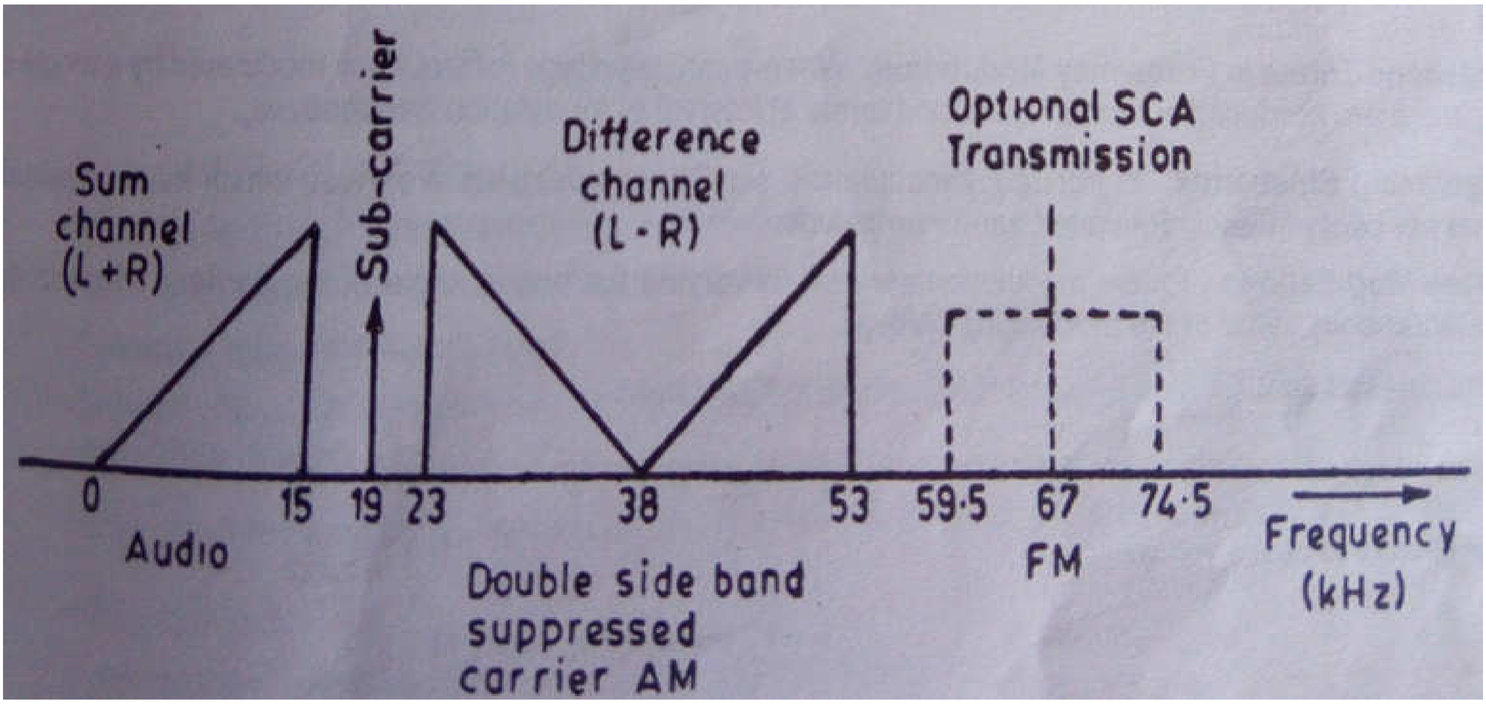


Figure 1.9. Spectrum diagram of stereo FM multiplex modulating signal.

4.0. Conclusion

In this unit, we have understood the meaning of Frequency Modulation, FM Wave Concepts, and Waveform of Frequency Modulated Voltage while showing the mathematical expression for Frequency Modulated Voltage. Issues such as frequency Spectrum of the FM wave, FM Bessel Functions Analysis as well as FM Bandwidth/Spectrum requirement were highlighted. Also, we looked at phase modulation and its Expressions for Phase Modulated Voltage. Comparison with respect to the expressions for phase and frequency modulated Voltages were discussed. We also, highlighted the comparison of Frequency and Amplitude Modulation. Other areas such as Noise and Frequency Modulation, the effect of noise on carrier, Pre-emphasis and De-emphasis circuits and interferences issues were analysis. We looked at Wideband and Narrowband FM and ended the unit with the Stereophonic FM Multiplex Systems

5.0 Summary

This Section has discussed Modulation as the process by which some characteristic, usually amplitude, frequency or phase angle of a voltage (called the carrier voltage) is varied in accordance with the instantaneous value of some other voltage, called modulating signal. We established that In amplitude modulation, the amplitude of the carrier varies in accordance with the instantaneous value of the modulating voltage. The following were established from the study,

- Modulating voltage $v_m = V_m \cos \omega_{m^t}$
- Carrier voltage $v_c = V_c \cos \omega_{c^t}$
- Modulated carrier voltage $v = V_c[1 + m_a \cos \omega_{m^t}] \cos \omega_{c^t}$
- Modulation index $m_a = \frac{K_a V_m}{V_c} = \frac{v_{cmax} - V_{cmin}}{v_{cmax} + V_{cmin}}$
- Sidebands produced in AM
  $v = V_c \cos \omega_{c^t} + \frac{m_a}{2} V_c \cos(\omega_c + \omega_m)\, t + \frac{m_a}{2} V_c \cos\, (\omega_c - \omega_{m^t})$
- Power Relations in AM: $P_t = P_{carrier} + P_{LSB} + P_{USB}$
  $P_t = P_c + \frac{m_a^2}{4} P_c + \frac{m_a^2}{4} P_c = P_c \left[1 + \frac{1}{2} m_a^2\right]$
- Current in Amplitude Modulation

$$I_t = I_c \sqrt{1 + \frac{m_a^2}{2}}$$

Other important points highlighted from the study include:

- Modulation Index- This is the fraction by which the amplitude of the carrier changes an amplitude modulation.
- Frequency Modulation- Frequency modulation involves varying the frequency of the carrier voltage in accordance with the instantaneous value of the modulating voltage.
- Modulating voltage: $v_m = V_m \cos \omega_{m^t}$
- Carrier voltage:$v_c = V_c \sin\, (\omega_{c^t} + \theta)$
- Modulated carrier voltage: $v = V_c \sin\, (\omega_{c^t} + m_f Sin\, \omega_{c^t})$
- Deviation Ratio δ. It is the ratio of the frequency deviation to modulation frequency and forms the modulation index m_f for frequency modulation.
  $\delta = m_f = \frac{K_f . V_m}{\omega_m}$
- Sideband Term in Frequency Modulation - When a carrier voltage is frequency modulated by a single sinusoidal voltage $v_{m,}$ there are produced numerous sideband terms at interval of modulation frequencyω_m.
- Significant Sidebands- In frequency modulation significant sideband are those which have amplitude at least equal to one per cent of the unmodulated carrier amplitude.
- Phase modulations - Phase modulation consists in varying the phase angle of the carrier voltage in accordance with the instantaneous value of the modulating voltage.
- Modulating voltage is given by the equation $v_m = V_m \sin \omega_{m^t}$
- Carrier voltage is given by the equation$v_\circ = V_c \sin\, (\omega_{c^t} + \theta)$
- Phase modulated carrier voltage is given by the equation $v = V_c \sin\, (\omega_{c^t} + \phi_m \sin \omega_{m^t})$
- Modulation index is given by $\phi_m = m_p = k_p V_m$

Also, we saw the merits and demerits of FM as summarized below.

- **Merits of FM.**

i. The amplitude of the frequency modulated wave remains unaffected.
ii. In FM there is a large decrease in noise and hence increase in S/N ratio.
iii. In FM, noise may be further reduced by increasing deviation.
iv. In FM, frequency allocation allows for a guard band. This reduces adjacent channel interference.
v. In the UHF bands where Fm operates, there is less noise than in the HF or MF bands.
vi. FM permits use of several independent transmitters on the same frequency with negligible interference.

- **Disadvantages of FM**

i. A much wider channel, typically 200 kHz, is needed.
ii. Transmitting and receiving equipments are complex and costly.
iii. Reception using conventional methods is limited to line of sight.

- **Pre-emphasis in FM.** In FM transmitter the higher modulation frequencies are boosted up before FM modulation using typically 50 μs L-R network.
- **De-emphasis**. When pre-emphasis is used in FM receiver at the output of detector, higher modulation frequencies are relatively attenuated to bring them back to their original relative values. Use of pre-emphasis and de-emphasis result in improved S/N ratio for higher modulation frequencies.
- **Adjacent Channel Interference in FM.** In FM system, use of limiter results in automatic reduction in adjacent channel interference. Adjacent channel interference in FM is also reduced by the guard band provided in FM broadcast channel allocation.
- **Co-channel interference in FM.** Use of amplitude limiter in FM results in interference reduction provided that desired signal channel is reasonably stronger than the undesired co-channel signal.
- **Wideband FM.** It is used for broadcast. Typically the modulating frequencies extent from 30 Hz 15 kHz. Modulation index exceeds unity. Maximum permissible deviation is =75 kHz.
- **Narrowband FM-** We established that in narrowband FM:

- Modulation index is usually about unity
- The maximum modulating frequency is usually 3 kHz and
- Maximum frequency deviation is usually = 5 kHz. It is used by mobile communication services.

- **Stereophonic FM Multiplex System-** This simply means that itdoes not use two separate channels, rather the sum of the two channels is sent as one signal and the difference as the other signal. The sum signal modulates the FM carrier and is received by monaural receiver and reproduced at output. The difference signal amplitude modulates a sub-carrier at 38 kHz. Which sub-carrier is then suppressed? The sidebands extending from 23 to 53 kHz then frequency modulates the carrier along with the sum signal.

1.0 Tutor-Marked Assignment (TMA)

1. Explain the concept of significant sidebands in frequency modulation.
2. Show the expression for the sinusoidal carrier voltage which has been phase modulated by another sinusoidal modulating voltage.
3. Prove that the significant sidebands converge rapidly despite the increase of modulation frequency in FM but not so in phase modulation.
4. Discuss the principle merits and limitations of FM?

5. Explain the effect of noise on carrier in FM system.
6. Why is it necessary to employ pre-emphasis and de-emphasis in FM system?
7. Draw typical pre-emphasis and de-emphasis circuits.
8. Explain how co-channel interference gets reduced in FM. Also, explain how co-channel interference gets reduced in FM system provided that the desired signal is stronger than the co-channel interfering signal.
9. Enumerate the salient features of wideband FM system.
10. Outline the typical applications of narrowband FM system.
11. With a well labeled block diagram, explain the principle of stereophonic FM multiplex system.

7.0 References/Further Readings

1. Sturley, k.R (1971). Frequency Modulated Ratio. George Newness Ltd., London, 1958.
2. Taub, H., and D.L. Schilling: Principle of communication Systems," McGraw Hill Book company, New York.
3. Mandel M(1973): Principles Electronic communications," Prentice Hall inc., Eaglewood cliffs, NJ.,.
4. Kennedy G (1977); "Electronic communication Systems," McGraw Hill Kogakusha Ltd., Tokyo.
5. Termans F.E.: "Electronics and Ratio Engineering," McGraw Hill Book Company, New York, 1955.
6. Ryder, J.D(1978): "Electronics Fundamentals and Applications," 5th edition, prentice Hall of india Private Ltd., New Deihi
7. Stremler, F.G.: Introduction to communication system (1977)," Addison Wesley Publishing Co., Reading, Mass,.
8. Holmes, Thom (2008). "Early Computer Music". Electronic and experimental music: technology, music, and culture (3rd ed.). Taylor & Francis. pp. 257–8. ISBN 0-415-95781-8. Retrieved 2011-06-04.
9. Stan Gibilisco (2002). Teach yourself electricity and electronics. McGraw-Hill Professional. p. 477. ISBN 978-0-07-137730-0.
10. David B. Rutledge (1999). The Electronics of Radio. Cambridge University Press. p. 310. ISBN 978-0-521-64645-1.
11. B. Boashash, editor, "Time-Frequency Signal Analysis and Processing – A Comprehensive Reference", Elsevier Science, Oxford, 2003; ISBN0-08-044335-4
12. T.G. Thomas, S. C. Sekhar Communication Theory, Tata-McGraw Hill 2005, ISBN0-07-059091-5 page 136.
13. Der, Lawrence, Ph.D., Frequency Modulation (FM) Tutorial, http://www.silabs.com/Marcom%20Documents/Resources/FMTutorial.pdf, Silicon Laboratories, Inc., accessed 2013 February 24, p. 5
14. B. P. Lathi, Communication Systems, John Wiley and Sons, (1968).ISBN0-471-51832-8, p, 214–217.
15. H. P. Westman, ed. (1970). Reference Data for Radio Engineers (Fifth ed.). Howard W. Sams & Co. p. 21-11.
16. Alan Bloom (2010). "Chapter 8. Modulation". In H. Ward Silver and Mark J. Wilson (Eds). The ARRL Handbook for Radio Communications. American Radio Relay League. p. 8.7. ISBN 978-0-87259-146-2.

17. Communication Systems (2001). 4th Ed, Simon Haykin.
18. FM Systems of Exceptional Bandwidth" Proc. IEEE vol 112, no. 9, p. 1664, September 1965.
19. A. Michael Noll (2001). Principles of modern communications technology. Artech House. p. 104. ISBN 978-1-58053-284-6.
20. Armstrong, E. H. (May 1936). A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation". Proceedings of the IRE. IRE. **24** (5): 689–740. doi:10.1109/JRPROC.1936.227383.

UNIT 2 FREQUENCY MODULATION TRANSMITTERS

1.0 Introduction
2.0 Objectives
3.0 Main Content
3.1. Methods of Frequency Modulation
3.1.1. FM Reactance Tube Modulation
3.1.2. Capacitive Reactance Tube.
3.1.3. Capacitance Reactance Tube Modulator
3.1.4. Expression for Modulation Index in Capacitive Reactance Tube Modulated Oscillator
3.1.5. Inductive Reactance Tube Modulator
3.1.6. Distortion in Reactance Tube Frequency Modulators
3.2. FM Reactance FET
3.3. Capacitive FET
3.4. Capacitive RC Reactance FET Modulator
3.5. Reactance Modulator Using BJT
3.6 Frequency Modulator using Varactor Diode
3.7. Simple FM Transmitter using Reactance Modulator
3.8. Frequency Drift in Reactance Modulator FM Transmitter
3.9. Stabilized Reactance Modulator FM Transmitter
3.10. Reactance Modulator FM Transmitter using AFC Frequency Stabilization.
3.11. Need for Frequency Mixing
3.12. Federal Telecommunication Laboratory Frequency Stabilization Method
3.13. Armstrong Method of Frequency Modulation
3.14. Pre-Emphasis and De-Emphasis
3.15. Armstrong F.M. Transmitter
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0 References/Further Readings

1.0 INTRODUCTION

We shall start this section by saying that in electronics telecommunication, a typical transmitter/ or radio transmitter is an electronic systemwhich produces radio wave propagation with an antenna. The transmitter itself generates a radio frequency alternating current, which is applied to the antenna. When excited by this alternating current the antenna radiates radio waves.Transmitters are necessary component parts of all electronic devices that communicate by radio, such as radio and television broadcasting stations, cell phones, walkie-talkies, wireless computer networks, Bluetooth enabled devices, garage door openers, two-way radios in aircraft, ships, spacecraft, radar sets and navigational beacons. It is very pertinent to note that term transmitter is limited to equipment that generates radio waves for communication purposes; or radiolocation, such as radar and navigational transmitters. Generators of radio waves for heating or industrial purposes, such as microwave ovens or diathermy equipment, are not usually called transmitters even though they often have similar circuits.The term is popularly used more specifically to refer to a broadcast transmitter, a transmitter used in broadcasting, as in FM radio transmitteror television transmitter. This usage typically includes the transmitter proper, the antenna, and often the building it is housed in.An unrelated use of the term is in industrial process control, where a "transmitter" is a telemetry device which converts measurements from a sensor into a signal, and sends it, usually via wires, to be received by some display or control device located a distance away.

In the previous unit, we established that the Frequency modulation (FM) has numerous advantages over AM such as better fidelity and noise immunity. However, it is much more complex to both modulate and demodulate a carrier wave with FM, and AM predates it by several decades. The primary function of a frequency modulation generator is to produce a carrier with variable output frequency, the frequency variation being proportional to the instantaneous value of the modulating voltage. The other requirements are: (i) the unmodulated carrier frequency should not get altered with the insertion modulation system and (ii) the frequency deviation should be independent of modulation frequency. In case the modulation system does not meet these requirements, correction should be introduced during the modulation forces. This unit shall look at the methods of frequency modulation for FM transmitters.

3.0 Objectives

After going through this unit, you should be able to

- ✓ understand the concept of frequency modulation
- ✓ distinguish direct from indirect modulation techniques
- ✓ relate modulation index with mathematical expressions.
- ✓ explain frequency stabilization in FM.
- ✓ explain frequency drift in reactance modulator FM Transmitter
- ✓ analyze the need for frequency mixing
- ✓ explain Armstrong method of frequency modulation
- ✓ explain the role of pre-emphasis and de-emphasis in FM broadcasting.

3.0 Main Content

3.1. Methods of Frequency Modulation

In general, there are two major methods of producing frequency modulation. These may be classified under two broad categories highlighted as directmethods and indirect methods
Now, you may ask Why the direct methods?
Well, In a direct FM system the instantaneous frequency is directly varied with the information signal. To vary the frequency of the carrier is to use an Oscillator whose resonant frequency is determined by components that can be varied. The oscillator frequency is thus changed by the modulating signal amplitude. Direct method of producing FM essentially utilizes an electron device such as an electron tube, Field Effect Transistor, Bipolar Junction Transistor or varactor diode which offers a voltage variable reactance, either capacitive or inductive. Such a device is placed directly in shunt with the L-C tuned circuit of an oscillator which synchronizes frequencies. The oscillator frequency then varies in accordance with the modulating voltage applied to the device resulting in frequency modulation.
By proper adjustment of circuit parameters and operating conditions, an almost linear relation may be obtained between the frequency deviation and the instantaneous modulating voltage.
There are five direct methods of producing frequency modulation. These include the following types:
i. Frequency modulation using reactance tube, FM-RT
ii . Frequency modulation using reactance Field Effect Transistor FM-RFET

iii. Frequency modulation using reactance Bipolar Junction Transistor, FM-RBJT
iv. Frequency modulation using varactor diode VD.FM-VD

In FM, the various types of indirect method consist (i). Pre-distorting the modulation voltage in an integrator to make its amplitude vary inversely with its frequency and (ii) making this modified modulating voltage to phase modulates the carrier. This method of modulation was suggested by Armstrong and hence is referred to as the Armstrong method of frequency modulation. We shall now discuss the various direct methods of producing frequency modulation below.

3.1.1. FM Reactance Tube Modulation

A reactance modulator changes the frequency of the tank circuit of the oscillator by changing its reactance. This is accomplished by a combination of a resistor, a condenser, and a vacuum tube (the modulator) connected across the tank circuit of oscillator and so adjusted as to act as a variable inductance or capacitance. This is a classical method of producing frequency modulation in electronic communication systems. Prior to the advent of semiconductor devices, this method was most popularly used. However, today it has been largely replaced by the varactor diode method. This method makes use of a reactance tube, i.e. an electron tube which offers between its plate and cathode terminals, a reactance varying in accordance with the modulating voltage applied between its control grid and cathode. This reactance tube is placed in shunt with the tuned circuit of an oscillator. This oscillator frequency then varies in accordance with the modulating voltage resulting in frequency modulation. By proper adjustment of circuit parameters and operating conditions, an almost linear relation may be obtained between the frequency deviation and the instantaneous modulating voltage.

The reactance tube may be either a capacitive reactance tube or an inductive reactance tube depending upon whether it offers a capacitive reactance or an inductive reactance across its plate-to-cathode circuit. Further, this reactive tube operation may be achieved by use of either an R-C network or an R-L networks. However R -C network is preferred. In context, we consider only the reactance tubes, using R-C network.

3.1.2. Capacitive Reactance Tube

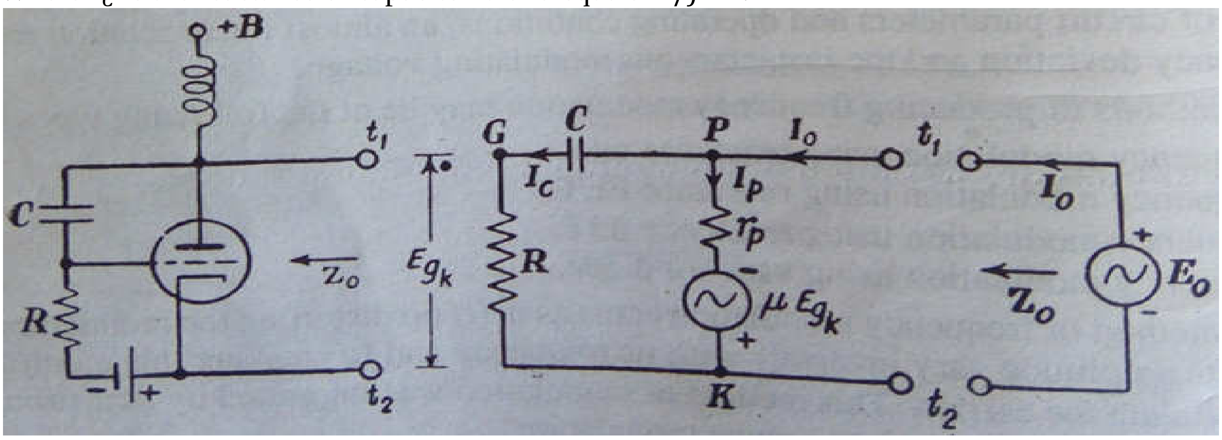
Now, the reactance tube is nothing but a pentode using an RC phase splitting network so arranged that when an A.C voltage is applied between plate and cathode, the resulting A.C. plate current is in phase quadrature with this applied A.C. voltage. Thus, the tube behaves as a reactance to the A.C. Voltage applied between its plate and cathode. If the plate current leads the applied A.C voltage applied by 90° then the reactance tube behaves as a capacitance. On the other hand, if the plate current lags behind the applied A.C voltage by 90°, then the reactance tube behaves as an inductor.

Figure 2.1a gives the basic circuit-of a capacitive reactance tube while Figure 2.1b gives it's A.C. equivalent circuit.In Figure 2.1a, if R $\ll \frac{1}{\omega C}$, then current I_c flowing through R and C (assuming that grid draws no current) leads the applied voltage by 90°. The grid-to-cathode voltage V_{gk}, being equal to I_c R, also leads the applied voltage by 90°. Plate current I_p being in phase with V_{gk} also leads the applied voltage by 90°: The reactance tube thus behaves as a capacitance.

The value of effective capacitance Co offered by the reactance tube may be calculated. Thus in the A.C. equivalent circuit of Figure 2.1b, let the tube offer an impedance Z_0 to a voltage V_0 applied between its plate the cathode, i.e., between terminals t_1 and t_2. Grid current is assumed to be zero.
Let $I_o$be the current drawn by the reactance tube circuit from the applied voltage V_o.
Let I_c be current that flows through the capacitor C. Then I_c is given by,

$$I_c = \frac{V_o}{R+\frac{1}{j\omega_C}} = \frac{V_o}{R-jX_C} \qquad (2.1)$$

WhereX_C is the reactance of capacitor *C* and equals $1/j\omega C$.



(a) Basic circuit arrangement. (b) A.C. equivalent circuit.
Figure2.1. Basic arrangement and A.C equivalent circuit of a capacitive reactance tube.

Grid-to-cathode voltage V_{gk} is given by.

$$V_{gk} = R.I_c = R\frac{V_o}{R-jX_c} \qquad (2.2)$$

The A.C current I_p is given by

$$I_p = \frac{V_o+\mu V_{gk}}{r_p} \quad \text{or} \qquad (2.3)$$

$$I_p = \frac{V_o}{r_p} + \frac{\mu}{r_p}.R\frac{V_o}{R-jX_c} \qquad (2.4)$$

Hence,$I_c = I_p + I_c = \frac{V_o}{r_p} + g_m R\frac{V_o}{R-jX_c} + \frac{V_o}{R-jX_c}$ (2.5)

Hence output terminal admittance is given by,

$$\frac{\boldsymbol{Y_o=\frac{1}{Z_o}=I_o}}{\boldsymbol{V\ sub\ o}} = \frac{1}{r_p} + \frac{g_m R}{R-jX_c} + \frac{1}{R-jX_c} \quad \text{or} \qquad (2.6)$$

$$Y_o = \frac{1}{\boldsymbol{r_p}} + \frac{1}{R-j\frac{1}{\omega}C} + \frac{1}{\left(\frac{1}{g_m}\right)-j\frac{1}{g_m R\omega C}} \qquad (2.7)$$

Equ. (2.7) suggests that so far as the output terminal impedance concerned the capacitive reactance tube of Figure2.7.
If, however, both $r_p$and $(R-jX_c)$ are large compared with the impedance $\left(\frac{\boldsymbol{1}}{g_m} - j\frac{1}{g_m R\omega C}\right)$, then these may be neglected.

Further, if $\frac{1}{\omega CR}$ » 1, i.e. $\frac{1}{g_m R\omega C}$ » $\frac{1}{g_m}$, then $1/g_m$ may be neglected as compared with reactance $1/g_m\omega CR$. Hence the output terminal impedance Z_o of this reactance tubes becomes simply $j/g_m\omega CR$. This is then purely capacitive and is caused by a capacitance $g_m CR$. Thus the reactance tube itself behaves as a capacitance of $g_m CR$. Figure 2.2 shows the A.C. equivalent circuit of capacitive reactance tube of Figure. 2.1.

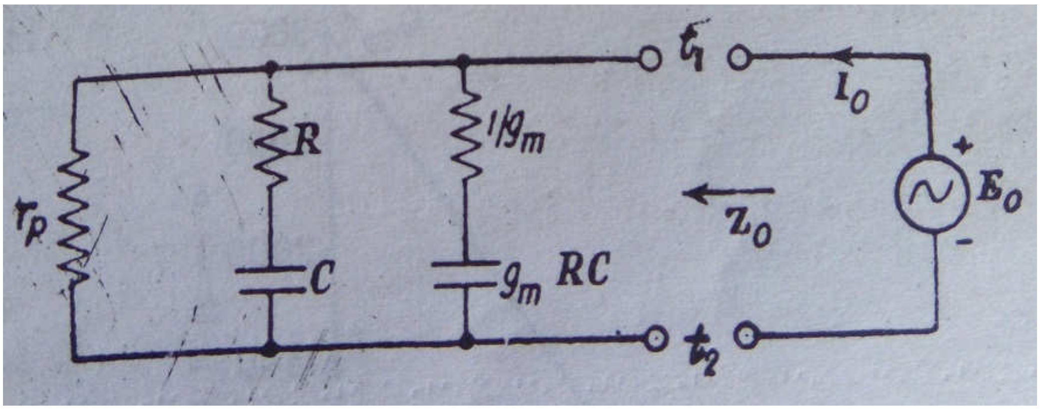


Figure2.2. A.C. equivalent circuit of capacitive reactance tube

3.1.3. Capacitance Reactance Tube Modulator

The schematic representation of the capacitive reactance tube is depicted in Figure2.1. This is obtained by keeping *C* and *R* constant and by varying the mutual conductance g_m in accordance with the modulating voltage. In this case, the capacitance offered by the reactance tube may be varied and hence frequency modulating may be produced.

Figure2.3 shows the basic circuit arrangement of the capacitance reactance tube modulated oscillator.The reactance tube is placed in shunt with the tuned circuit of the tuned plate oscillator. The modulating voltage V_m is applied at the control grid of reactance tube.

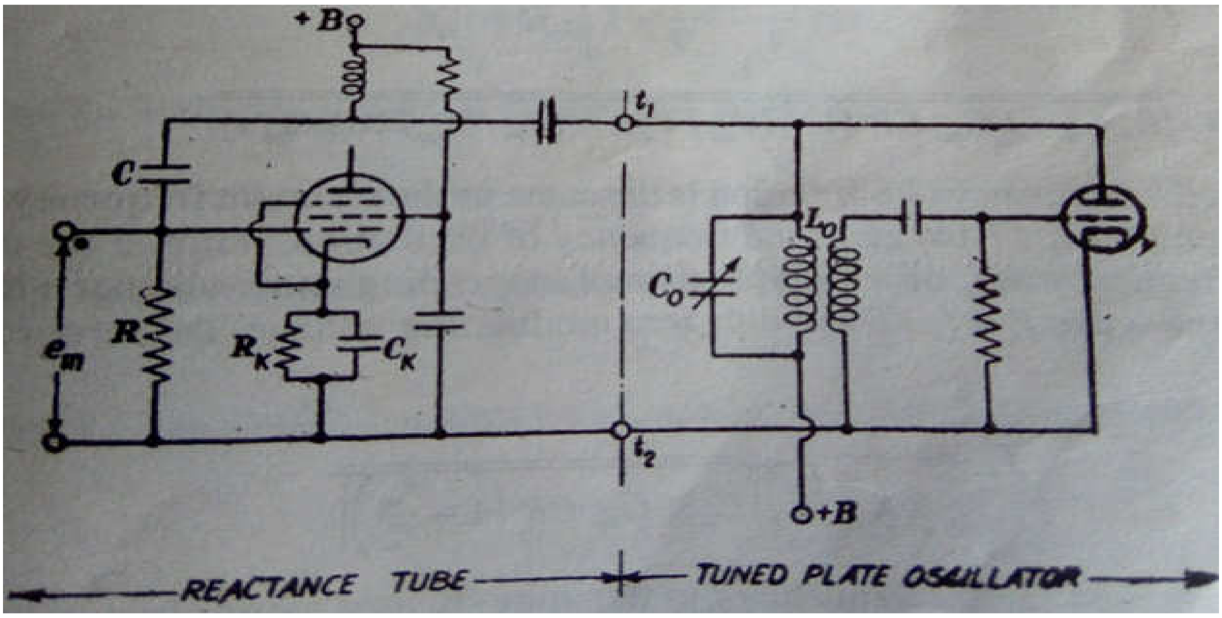


Figure2.3. Basic circuit of capacitive reactance tube modulated oscillator.

3.1.4. Expression for Modulation Index in Capacitive Reactance Tube Modulated Oscillator

We shall now establish an approximate expression for deviation ratio considering a capacitive reactance tube modulator oscillator for FM below. Now, Figure 2.4 gives the curve showing the variation of g_m of the reactance tube with the grid potential V_c.

The g_m versus grid voltage curve of Figure2.4 is linear over most of the negative grid voltage region. This linear portion may be extended or projected on both side to cut the g_m and V_c axes at points G_{co} and V_{co} respectively as shown.

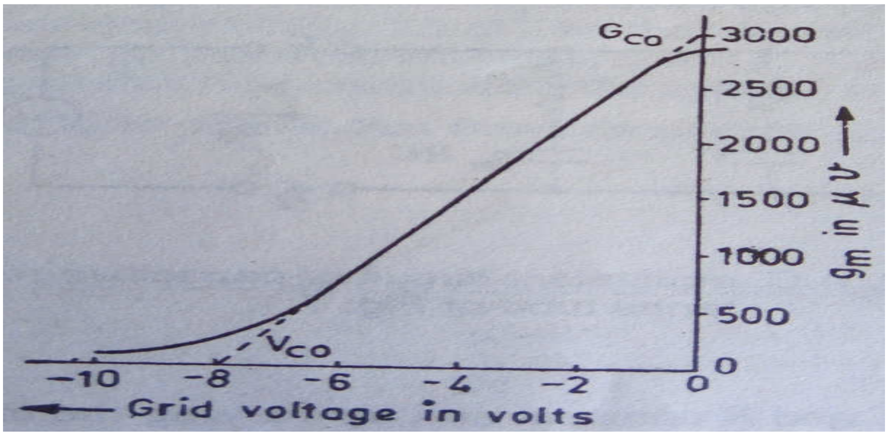


Figure2.4. Variation of g_m with grid voltage.

Then g_m for any grid voltage V_c is given by the relation,

$$g_m = \frac{G_{co}}{V_{co}} v_c + G_{co} \quad (2.8)$$

But

$$v_c = V_{cc} + V_m \quad (2.9)$$

$$= V_{cc} + V_m \cos \omega_m t \quad (2.10)$$

Hence,

$$g_m = -G_{co} \frac{V_{cc}}{V_{co}} + G_{co} - \frac{G_{co}}{V_{co}} . V_m \cos \omega_m t \quad (2.11)$$

Hence effective output capacitance is approximately given by,

$$C_e = g_m$$

$$CR = CRG_{co} \left[1 - \frac{V_{co}}{V_{co}} - \frac{V_m}{V_{co}} cos \omega_m t\right] \quad (2.12)$$

Hence, resonant frequency of the tank circuit given by,

$$f_o = \frac{1}{2\pi \sqrt{L_o(C_o + C_e)}} \quad (2.13)$$

$$\text{or} f_o = \frac{1}{2\pi \sqrt{L_o C_o + LG_{co}\ CR\ [1 - V_{co}/V_{co}) - (V_m/V_{co})\ cos\ \omega_m t]}} \quad (2.14)$$

Let us assume that the frequency of oscillation is the same as the resonant frequency of the tank circuit. Then frequency f_o as given by Equ. (2.14) gives the frequency of oscillation.

Further the carrier frequency f_c may be either the same frequency as f_o or a multiple thereof, depending upon whether a harmonic generator is used or not Let us' assume that $f_c = f_o$, then with zero modulating voltage, the carrier frequency is given by,

$$f_o = \frac{1}{2\pi \sqrt{L_o\left[C_o + G_o\ CR\left(1 - \frac{V_{co}}{V_{co}}\right)\right]}} \quad (2.15)$$

The ratio of the modulated carrier frequency f to the unmodulated carrier frequency f is then given

$$\frac{f_o}{f_c} = \frac{\sqrt{C_o + G_{co}\ CR\ [1 - (V_{cc}\ /\ V_{co})]}}{\sqrt{\sqrt{C_o + G_{co}\ CR\ [1 - (V_{cc}\ /\ V_{co}) - (V_m\ /\ V_{co})\ cos\ \omega_m t}}} \quad (2.16)$$

$$= \frac{1}{\sqrt{\frac{1 - G_o\ CR(V_m\ /\ V_{co})\ cos\ \omega_m t}{C_o + C_{co}\ CR\ [1 - (V_{co}\ /\ V_{co})]}}}$$

$$= \frac{1}{\sqrt{1 - \frac{v_m \cos \omega_m t}{(C_o V_{co} \ / \ G_{co} \ CR) + (V_{co} - V_{cc})}}} \qquad (2.17)$$

The expression on the right hand side of Equ. (2.17) may be expanded by the binomial theorem. Further assuming the frequency change to be small, only the first term in the expansion maybe considered. Accordingly, Equ. (2.17) may be put as,

$$\frac{f}{f_c} = 1 - \frac{1}{2}\left[\frac{V_m \cos \omega_m t}{(C_o V_{co} / C_c \ CR) + (V_{co} - V_{cc})}\right] \qquad (2.18)$$

$$\text{or} = f_c[1 + \frac{1}{2} \frac{V_m}{(C_o V_{co} \ / \ G_{co} \ CR) + (V_{co} - V_{cc})} \cos \omega_m t \qquad (2.19)$$

$$f = f_c \ [1 = m_f \cos \omega_m t] \qquad (2.20)$$

Where m_f is the modulation index and is given by,

$$m = \frac{1}{2} . \frac{V_m}{(C_o V_{co} \ / \ G_{co} . CR)} + (V_{co} - V_{cc}).$$

3.1.5. Inductive Reactance Tube Modulator

The electron tube of Figure 2.1a may be made to offer inductive reactance between its anode and cathode, instead of a capacitive reactance by exchanging the positions of *R*and *C* letting R » 1/00 C. It may be shown that the tube then behaves and inductor of value of CR/ g_m. This inductive reactance tube may then be connected across the tank circuit of an oscillator exactly in the manner shown in Figure2.3. Then keeping C and R constant and on varying the modulating voltage applied at the 'control grid of the reactance tube, g_m, of the tube varies, effective inductance CR/ $g_{m,}$ or the tube varies and hence the frequency of oscillation of the oscillator varies. ~ results in the desired frequency modulation. It may be proved that g_m of the inductive tube is given by,

$$g_m = G_{co}\left[1 - \frac{V_{cc}}{V_{co}} - \frac{V_m}{V_{co}} \cos \omega_m t\right] \qquad (2.22)$$

Hence, effective inductance offered by the tube is given by,

$$L_o = \frac{CR}{g_m} = \frac{CR}{G_{co}[1 - (V_{cc} / V_{co}) - (V_m \ / \ V_{co}) \ cos \ \omega_m t]} \qquad (2.23)$$

Hence, it may be proved that the frequency of oscillation of modulation/is given by,

$$f = f_o \ [1 + m_f \cos \omega_m t] \qquad (2.14)$$

Where f_c is the unmodulated carrier frequency and m_f is the modulation index and is given by,

$$m_1 = \frac{1}{2} . \frac{V_m}{(V_{cc} \ CR \ / \ G_{co} L_o) + + (V_{cc} - V_{co})} \qquad (2.25)$$

Where$L_o$is the inductance of the tank circuit coil with modulation.

3.1.6. Distortion in Reactance Tube Frequency Modulators

Recall that in capacitive reactance tube modulator of Figure2.3, the inductance $L_o$for the tuned circuit is kept constant while the total shunt capacitance varies as equivalent capacitance offered by the reactance tube varies. Now this equivalent capacitance is proportional tog_m, while the variations ing_m, are proportional to variations in the modulating voltage. These relations are only approximately true. Thus,$g_m - v_c$curve of Figure2.4departs from linearity. Also the relation $C_c = g_m CR$ is only approximately true. Hence, distortion is caused in the modulated voltage. However, these two causes do not introduce excessive distortion.

The more serious cause of distortion in reactance tube modulators the fact that the resonant frequency of the tuned circuit of modulated oscillator is given by the relations 1/2 $\pi\sqrt{LC}$so that even ifcapacitance change is proportional to modulating voltage, the resultant frequency change is not strictly proportional to the change in modulating voltage. In other words Equ. (2.18) is only approximately true. The higher frequency terms are usually quite significant. The percentage bandwidth is given by the expression below.

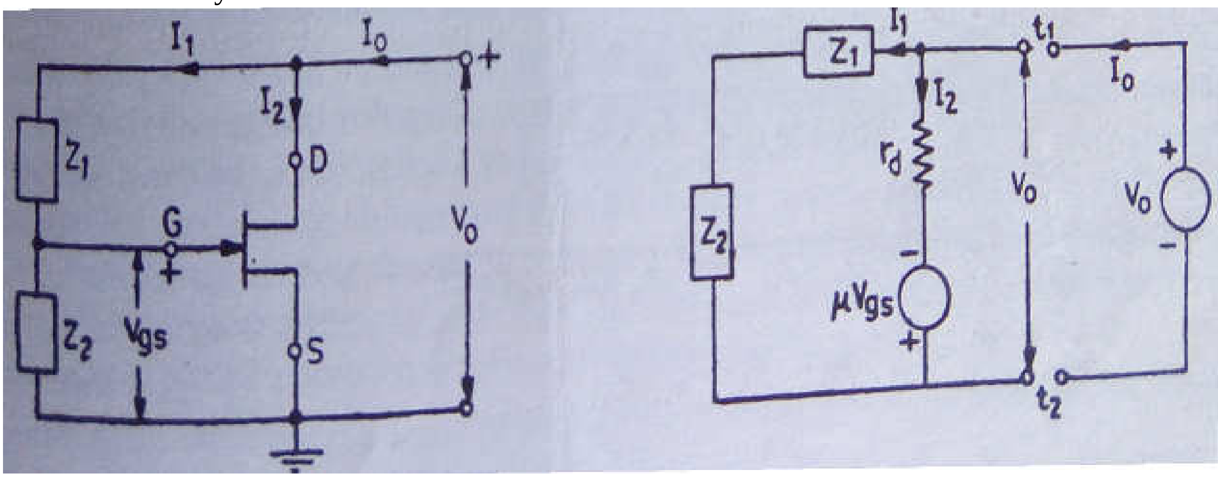
$$\text{Percentage bandwidth} = \frac{\text{Peak} - \text{to} - \text{peak frequency swing}}{\text{Centre frequency}}$$

Accordingly in applications like sub-carrier frequency distortion where large percentage bandwidths are involved, it is not possible to use reactance tube modulation to produce the desired large percentage bandwidth directly because then the distortion is very large. However, reactance tube modulation may still be used to produce small percentage bandwidth at first and the consequent small percentage distortion. Subsequently, frequency conversion may be used to increase the percentage bandwidth. This method does not increase the percentage distortion.

3.2.FM Reactance FET

Figure2.5agives the basic circuit of a reactance field effect Transistor (FET) while Figure 2.5b gives it's A.C equivalent circuit. The circuit shown in Figure. 2.5a is the basic circuit of a reactance FET which may be connected across the tank circuit of the oscillator to be frequency modulated. The FET may be made to behave as capacitance or as an inductance depending on the nature of circuit elements Z_1 and Z_2. The FET acts as a capacitance if Z_1 is a capacitor and $Z_2$is a resistor and R»X_c.Furthermore, instead ofR-C network, we may use R-L network. Then for obtaining capacitive reactance, Z_1 is a resistor and Z_2 is inductor and X_L » R while for obtaining inductive reactance Z_1 is a resistor and R »X_L. Any one of these four combinations may be used to obtain variable reactance.

Table 2.2shows the four different arrangements of reactance FET. It also gives (i) the pre-requisites for FET to behave as almost pure reactance and (ii) the formulae for output reactance. The basic prerequisite in each case is that the drain current must be much greater than the bias network current.Out of the four configurations listed in Table 2.2, the first one using RC network is most commonly used.



(a) Basic circuit. (b) A.C. equivalent circuit

Figure 2.5. Basic circuit of a reactance FET and it's A.C equivalent circuit.

Table 2.2. Four Different Arrangement of Reactance FET

Circuit Arrangement	Z_1	Z_2	Condition $Z_1 \gg Z_2$	Reactance Formula
R C Capacitive	C	R	$X_c \gg R$	$C_e = g_m RC$
R C Inductive	R	C	$R \gg X_c$	$L_e = \frac{RC}{g_m}$
R L Capacitive	R	L	$R \gg X_L$	$C_e = \frac{g_m L}{R}$
R L Inductive	L	R	$X_L \gg R$	$L_e = \frac{L}{g_m R}$

With reference to the equivalent circuit in Figure 2.5b, let the FET offer an impedance Z_o to voltage V_o applied between its drain and source i.e. between terminals t_1 and t_2. Gate current is assumed to be zero. Let I_o be the current drawn by the reactance FET from the applied source V_o.From Figure 2.5b for open circuit output,

$$I_1 = \frac{V_0}{Z_1+Z_2} \tag{2.27}$$

$$I_2 = \frac{V_0+\mu V_{gs}}{r_d} \tag{2.28}$$

$$V_{gs} = V_0 \frac{Z_2}{Z_1+Z_2} \tag{2.29}$$

Hence ,$I_0 = I_1 + I_2 = \frac{V_o}{Z_1+Z_2} + \frac{V_0+\mu V_{gs}}{r_d} = \frac{V_o}{Z_1+Z_2} + \frac{V_o}{r_d} + g_m V_0 \cdot \frac{Z_2}{Z_1+Z_2} = V_o = \left[\frac{1}{Z_1+Z_2} + \frac{1}{r_d} + \frac{g_m Z_2}{Z_1+Z_2}\right]$

Hence,

$$Y_0 = \frac{I_o}{V_o} = \frac{1}{Z_1+Z_2} + \frac{1}{R_d} + \frac{1}{g_m\,[(Z_1/Z_2)+1]} \tag{2.30}$$

Let, $Z_1 = Z_{gd} \quad and\ Z_2 = Z_{gs}$

Then, $Y_o = \frac{1}{Z_{gd}+Z_{gs}} + \frac{1}{r_d} + \frac{1}{(1/g_m)\,[Z_{gd}\,/Z_{gs})+1]}$ (2.31)

Equ. (2.31) forms the general equation valid for all the four configurations listed in Table 2.2.

3.3. Capacitive FET

In this case, it is assumed that the bias network current I_1 as depicted in Figure 2.6is negligible compared to the drain current I_2. Hence, the impedance of the bias network (R-C combination) must be large enough to be ignored. We further assume that dynamic drain resistance r_d is very large and hence may be neglected in the shunt are in the circuit of Figure. 2.6. It represents the circuit corresponding to Equ. (2.32).

$Z_{gd} = \frac{1}{j\omega C}$ and $Z_{gs} = R$

Hence, $Y_o = \frac{1}{R+(1/j\omega C)} + \frac{1}{r_d} + \frac{1}{(1/g_m)\,[1+(1/j\omega CR)]}$ (2.32)

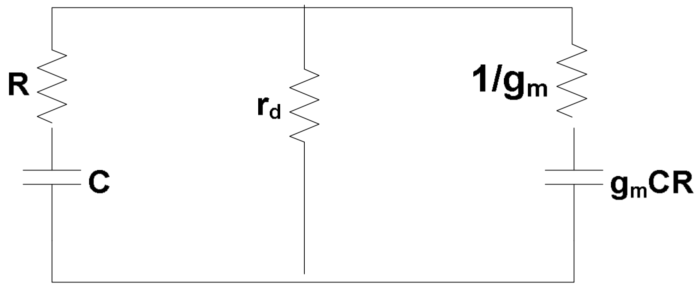


Figure 2.6. Circuit equivalent to capacitance FET.

The equivalent circuit for FET then reduces to $1/g_m$ in series with capacitance $g_m CR$.

Thus, $Z_o = \frac{1}{g_m} + \frac{1}{j\,\omega\,g_m\,CR} = \frac{1}{g_m}\left[1 - j\frac{X_c}{R}\right]$ (2.33)

If $X_c \gg R$, Equ. (2.33) reduces to the following form

$Z_o = -j\frac{X_c}{g_m R}$ (2.34)

Then the impedance offered by FET is the capacitive reactance of a capacitor C_c given by

$C_e = g_m CR$ (2.35)

From Equ. (2.35), some very useful inferences can be drawn:

i. The equivalent capacitance C_e depends on the device transconductance g_m and may be varied
ii. The C_e may originally be adjusted to any suitable desired value by varying the components R and C

iii. The term $g_m CR$ has the dimensions of capacitance.

We have assumed that $X_c >> R$. If, however, X_c is not much greater than R, the gate voltage is not longer exactly 90^0 out of phase with the applied voltage V_o, nor with the drain current I_d. Hence the input impedance is no longer purely reactive and has a resistance component $1/g_m$ as seen from Equ. (2.35). This component $1/g_m$ varies with the applied modulating voltage and appears directly across the tank circuit of the master varying its Q and hence varying the output voltage. This results in a small amount of undesired amplitude modulation. This is true for all types of reactance modulators. If amplitude modulation is undesirably large, an amplitude limiter is placed at the output of the modulated oscillator.

3.4. Capacitive RC Reactance FET Modulator

Another FM modulator is discussed here and it is known as the capacitive RC reactance FET. This type comes in shunt with the tuned circuit master oscillator, in this case a Clapp oscillator. Figure2.7 gives the basic circuit. The Clapp oscillator is most popularly used. However, we may use any other oscillator such as Hartley, Colpitt or tuned-drain oscillator. These oscillators require two- tuned circuits such as tuned gate-tuned drain oscillator.

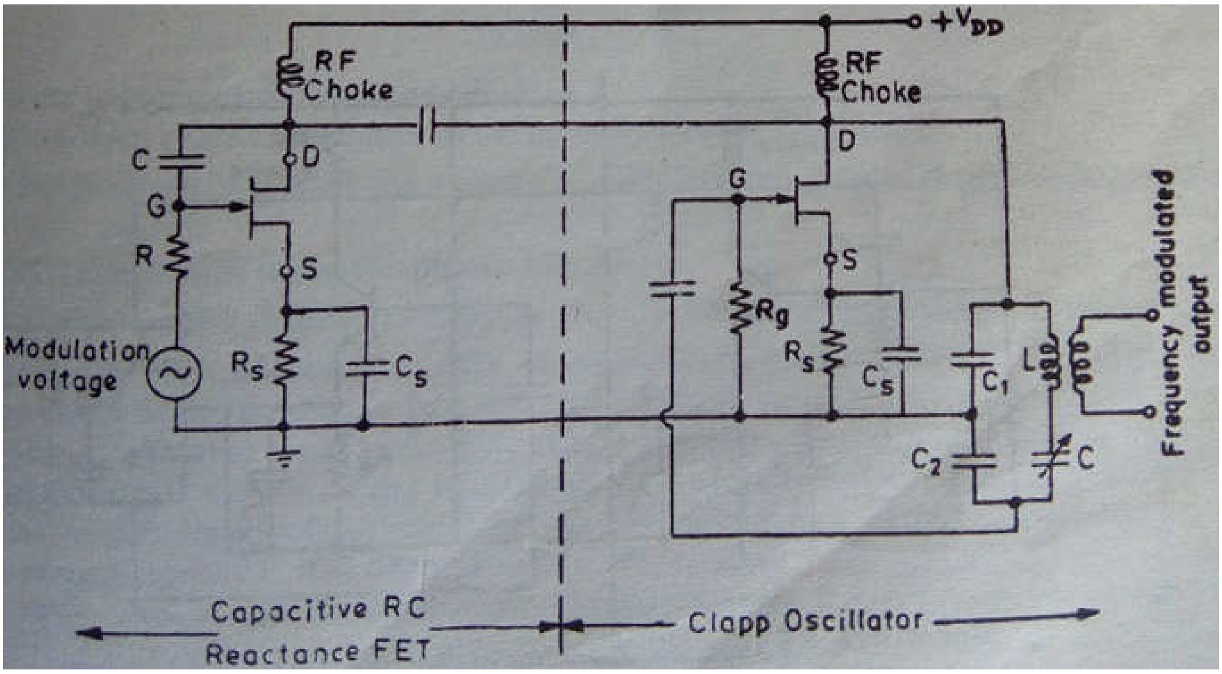


Figure2.7. Capacitive RC reactance FET modulator.

Furthermore, each of such oscillator must be followed by a buffer amplifier to avoid direct loading of the master oscillator. The RF chokes in the drain circuits are used to isolate various points of the circuit for alternating current while simultaneously providing A.C paths. The modulating voltage is applied at the gate of the FET by keeping it in series with R. By keeping C and R constant and by varying the g_m, in accordance with the modulating voltage, the equivalent capacitance C_e offered by the reactance FET may be varied and hence frequency modulation may be produced.The analysis for deriving expression for the modulation index in the capacitive RC reactance FET modulated oscillator is exactly the same for capacitive RC reactance tube modulated oscillator.

3.5. Reactance Modulator Using BJT

This is the type of modulator where a BJTis operated actively in conjunction with R-C or phase splitting circuit while making a zero inductive or capacitive reactance between its collector and emitter terminals. In similarity with FET, a BJT so used may be in any of the four circuit configurations as shown in Table 2.2. Further in the case of BJT also, the most popular circuit arrangement is R-C capacitive.

In this arrangement,Z_1(or Z_{bc}) is a capacitor C with Z_2 (between base and emitter) is an inductor.Further at the operating frequency $X_c >> R$, then collector and emitter terminals of the BJT offers a capacitanceC_e. The modulating voltage V_m is applied in with the resistor $R$of $R - C$phase splitting network. Then as the modulating voltage varies, the $C_e$varies. Figure 2.8 gives the basic circuit of capacitive RC reactance transistor modulator. The effective capacitance C_e offered by collector-to-emitter circuit of reactance BJT comes in shunt with the tuned circuit of the master oscillator. Figure2.8 shows a Clapp oscillator. This oscillator circuit is the one most popularly used although other circuits such as Hartley, Colpitt or tuned collector oscillator may also be used.

Further, the oscillator must be followed by a buffer amplifier in order to isolate it from the loading effect of the fell stage. The RF chokes in the circuit are used to isolate various points of the circuit for alternating current, simultaneously providing a D.C . path.

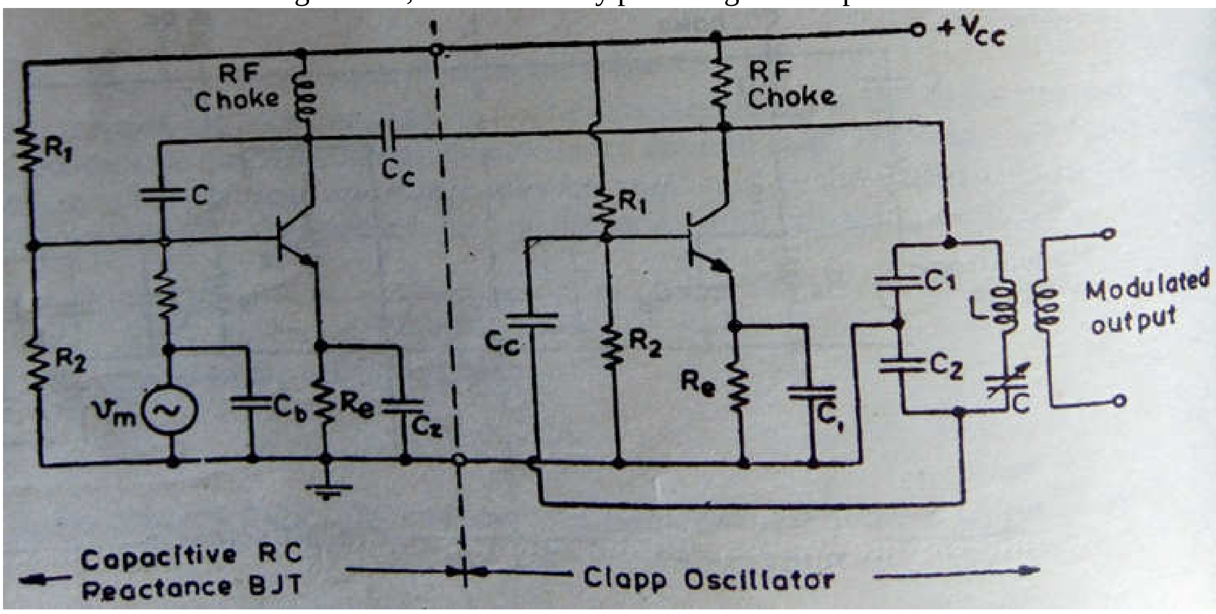


Figure2.8. Capacitive RC reactance BJT modulator.

3.6 Frequency Modulator using Varactor Diode

In this type of frequency modulator, Figure2.9 gives a complete representation of the circuit using varactor diode. It uses the junction capacitance of a reverse biased varactor diode placed across the tuned circuit of an oscillator. The reverse bias of the varactor diode varies in accordance with the modulating voltage placed in series with the bias. The capacitance, therefore, varies causing the frequency of oscillation to vary. This forms the simplest reactance modulator circuit.

However, it has the disadvantage that it uses a two terminal device making its application to be seemingly restricted. However, the circuit is frequently used for automatic frequency control and for remote tuning.

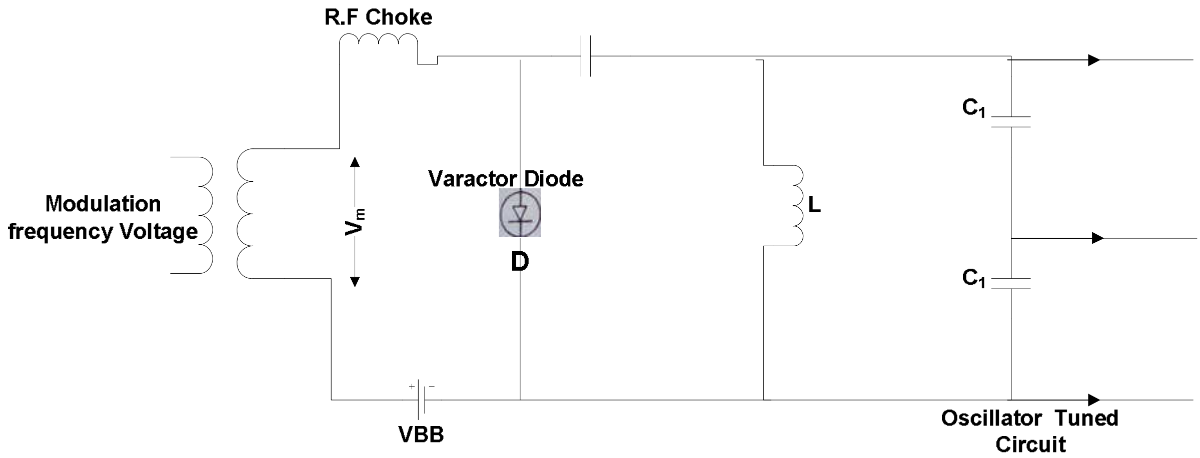


Figure2.9. Frequency modulator using varactor diode.

The entire voltage applied to the varactor diode is given by,

$$v = V_{BB} + v_m \tag{2.38}$$

Where v_{BB} is the reverse bias, andv_m, is the modulating voltage. The capacitance of the varactor junction varies as $k\sqrt{V_m}$ where k is a constant.The frequency of oscillation is then given by,

$$f = \frac{1}{2\pi\sqrt{L(C_1+Kv_m^{1/2})}} \tag{2.39}$$

Where$L$ is the Inductance of the tuned circuit coil, and$C$is the total shunt capacitance of the capacitors in the tuned circuit, then for small values of modulating voltage V_m the frequency $f$varies almost linearly with the modulating voltage V_m,.

3.7. Simple FM Transmitter using Reactance Modulator

Basically, Figure2.10gives the block diagram of a simple 96 MHz PM Transmitter using reactance modulator. In this transmitter the reactance tube (FET or transistor) frequency modulates the master oscillator of carrier frequency say 4MHz, producing frequency deviation of say 3.125 kHz. The master oscillator is followed by a buffer amplifier to avoid loading the master oscillator. Buffer amplifier is followed by a chain or frequency multipliers to raise the carrier frequency to the required frequency level.

Since the frequency multiplier follow the modulated master oscillator, the carrier frequency as well as the frequency deviation get raised by the same factor say n. The FM transmitter showed Figure2.10 uses three frequency doublers and one frequency tripler. Frequency multipliers are followed by a chain of RF power amplifiers to raise the carrier power to the desired level. The final output is then fed to the transmitting antenna.

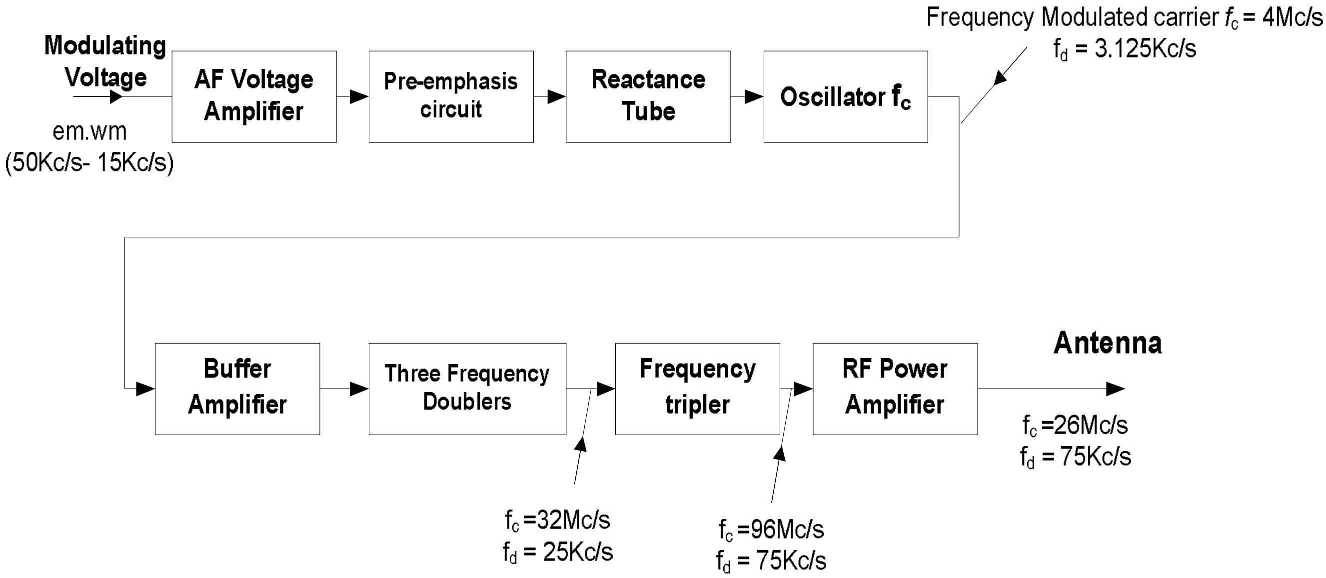


Figure 2.10. Simple FM transmitter using reactance modulator.

3.8. Frequency Drift in Reactance Modulator FM Transmitter

It is very important that we understand that the carrier frequency of reactance modulator FM transmitter may drift due to one of the follow causes namely.
i. Variation in supply voltage.
ii. Variation in temperature and humidity.
iii. Aging of electron device.

Essentially, in reactance transistor modulator as the supply voltage varies the electrode voltages of reactance transistor and master oscillator transistor vary resulting in corresponding change in the carrier frequency. Such a frequency drift may be reduced by use of stabilized power supply and use of parallel reactance transistor modulator.
Changes in temperature and humidity result in change in component values and hence change in master oscillator frequency. This trouble may be reduced to a large extent by placing the relevant components constant temperature chamber.Carrier frequency may also vary with the aging of electron device.

3.9. Stabilized Reactance Modulator FM Transmitter

It is important you know that the use of (i) stabilized power supply, (ii) parallel reactance transistor modulator and use of constant temperature chamber no doubt provide high order of frequency constancy but for best results more positive methods of frequency stabilization are needed. The CCIR regulatory requirements for FM broadcast require that the carrier frequency FM transmitter be maintained constant with ± 2 kHz of the declared value.

In AM transmitter use of Crystal oscillator as master oscillator avoids all these troubles. However in FM transmitter, crystal controlled master oscillator cannot be used since the reactance transistor is required to vary the master oscillator frequency in accordance with the

modulating Voltage. A crystal oscillator may, however, be used to provide a standard reference frequency with which either the carrier frequency or a sub-harmonic of the carrier frequency may be compared. The frequency difference then indicates the drift in carrier frequency and this frequency is used to operate control circuits which work in such a way as to reduce this deviation.
The following methods of frequency stabilization making use of crystal oscillator are popularly used:

i. Frequency stabilization using Automatic Frequency Control (AFC method).
ii. Frequency stabilization using Balanced Phase Detector (also called the Federal Telecommunications Laboratory method).

3.10. Reactance Modulator FM Transmitter using AFC Frequency Stabilization.
This is also method of Frequency stabilization is shown inFigure2.11. It shows the basic arrangement. Here, the master oscillator which the reactance modulator operates cannot be a crystal controlled oscillator. But it may be made to have the stability of crystal oscillator through use of an automatic frequency control (AFC) system. The crystal oscillator frequency differs from the centre carrier frequency of the master oscillator by a fixed value say 500 kHz. Thus, in the transmitter of Figure. 2.11, the master oscillator centre carrier frequency is 4MHz, while crystal oscillator frequency is 3.5 MHz the difference being 500 kHz. This difference frequency (or Intermediate Frequency) is amplified in an I.F. amplifier and then fed to a discriminator designed for operation at 500 kHz.

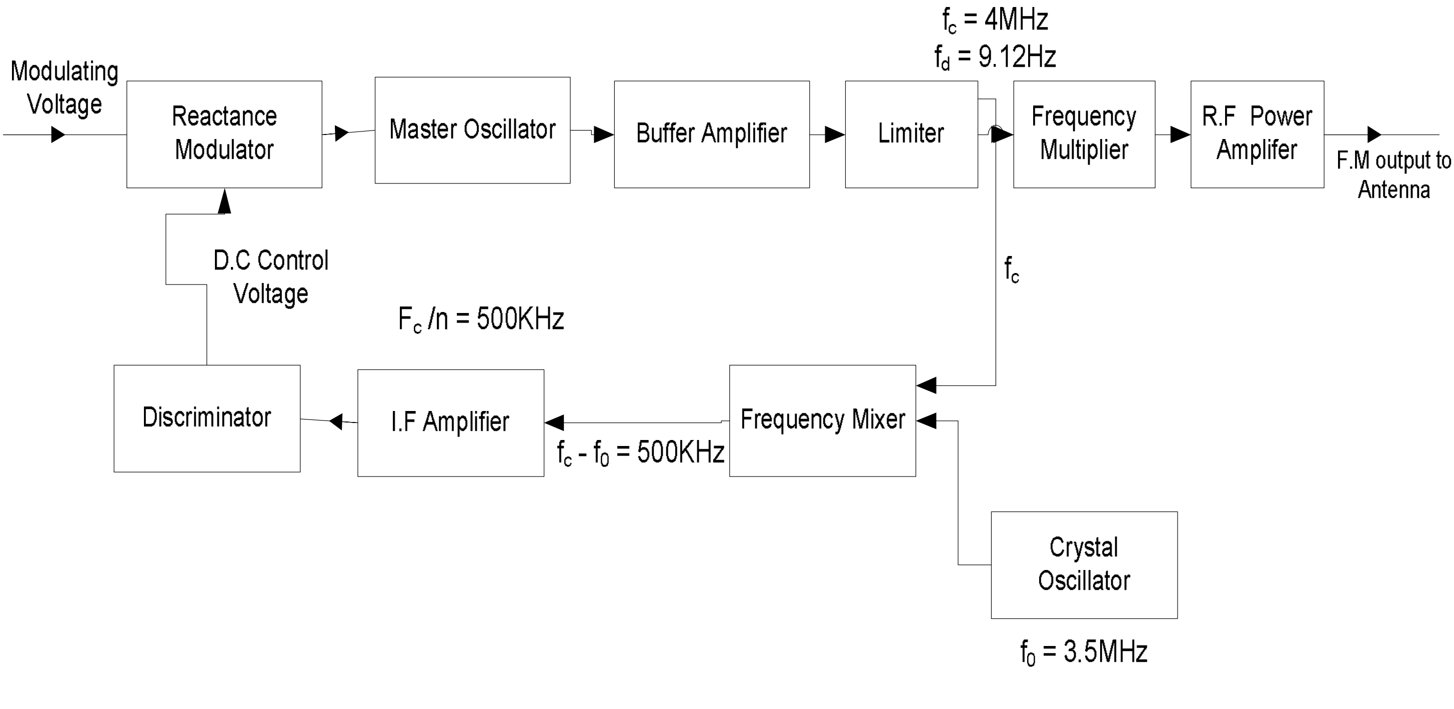


Figure2.11. Reactance modulator F.M. transmitter using AFC frequency stabilization.

This discriminator gives at its output a D.C . voltage which is positive or negative depending upon whether actual input frequency is greater or less than the prescribed frequency (500-kHz in this case) and the magnitude of this D.C . voltage depends upon the shift Δf of the incoming frequency from the prescribed value.

This output voltage of the discriminator is fed to the base of the reactance BJT in such polarity that the resulting change in the oscillator frequency tends to cancel the frequency drift and thus provide correct prescribed frequency voltage to the discriminator resulting in zero D.C control voltage at its output any incidental frequency drift is thus considerably reduced.

Evidently, for frequency stabilization, a control voltage is essential and this control voltage results only when the input centre frequency to the discriminator differs from the prescribed value i.e. when a certain drift, how so small exists. Thus, perfect frequency stability is not possible. However, the residual frequency variation is well within the permissible limits of ± 2 kHz.

The main circuit of the reactance modulator FM transmitter is conventional and consists of the reactance modulator, modulated master oscillator, buffer amplifier to isolate the master oscillator, limiter to remove any amplitude variations, frequency multipliers to raise the master oscillator frequency to the desired high value of the carrier and the RF amplifier to raise the master oscillator, limiter to remove any amplitude variations, frequency multipliers to raise the master oscillator frequency to the desired high value of the carrier and the RF amplifier to raise the power level of the carrier voltage to the desired value. The final frequency modulated voltage may then be fed to the transmitting antenna.

In the circuit of Figure2.11, the D.C control voltage is fed to the reactance modulator. Alternatively, this D.C control voltage may be fed to a varactor diode connected across the oscillator tank and this varactor diode may then be used exclusively for the AFC. In yet another system, this D.C control voltage may be amplifier in a D.C voltage amplifier and fed to a servo motor which is connected to a trimmer capacitor in the oscillator circuit. The D.C control voltage then moves the motor and alters the setting of the trimmer capacitor and hence alters the oscillator frequency by such amount as to bring it to the correct value.

3.11. Need for Frequency Mixing

Often, most people want to know the reason for heterodyning or frequency mixing in a communication system. Now, if we can stabilize the frequency of the master oscillator directly without mixing it with the output of a crystal oscillator, the resulting stabilization circuit would be much simpler. But, then the quality suffers. This is because the stability of the entire circuit depends basically on the stability of the discriminator. If the frequency of the discriminator drifts, the resulting output frequency of the entire system also drifts equally. Since the discriminator is essentially a passive network and its frequency stability is higher than that of master oscillator by a factor of typically 3:1. A properly designed L-C oscillator may have drift within 5 parts in 10,000 i.e. about 2kHz at 4MHz. Hence direct stabilization without use of crystal oscillator would result in drift within about 700 Hz at best in this case. On frequency multiplication the resulting drift will increase to 700 x 24 Hz = 16.8 kHz, a very high value.

Consider now the situation depicted in Figure 2.11 using a crystal oscillator. In this case, the discriminator has centre frequency of only 500 kHz instead of 4 MHz i.e. reduced by a factor of 20. Hence in the case, the actual drift is about one-twentieth of that in the direct stabilization i.e. only about 35 Hz.

3.12. Federal Telecommunication Laboratory Frequency Stabilization Method

Figure2.12gives the block diagram of a reactance modulator FM transmitter using federal telecommunication laboratory method of frequency stabilization. In this case, the crystal oscillator operates at the master oscillator frequency and some sub-harmonics (say 8^{th}) of the master oscillator and the crystal oscillator frequencies are fed to a balanced phase detector which compares the phases of the two voltages of almost the same frequency. The output D.C voltage of this phase detector is proportional to the phase difference between the two oscillator outputs. This D.C voltage of this phase detector is proportional to the phase difference between the two oscillator outputs. The D.C voltage is fed to the base of the reactance modulator transistor and it changes the master oscillator frequency in such a way as to make it coincide with the crystal oscillator frequency. This system also does not yield perfect frequency stability since a controlling voltage is essential. However, it maintains the carrier central frequency constant well within ± 1 kHz which is less than the permissible variation of±2kHz.

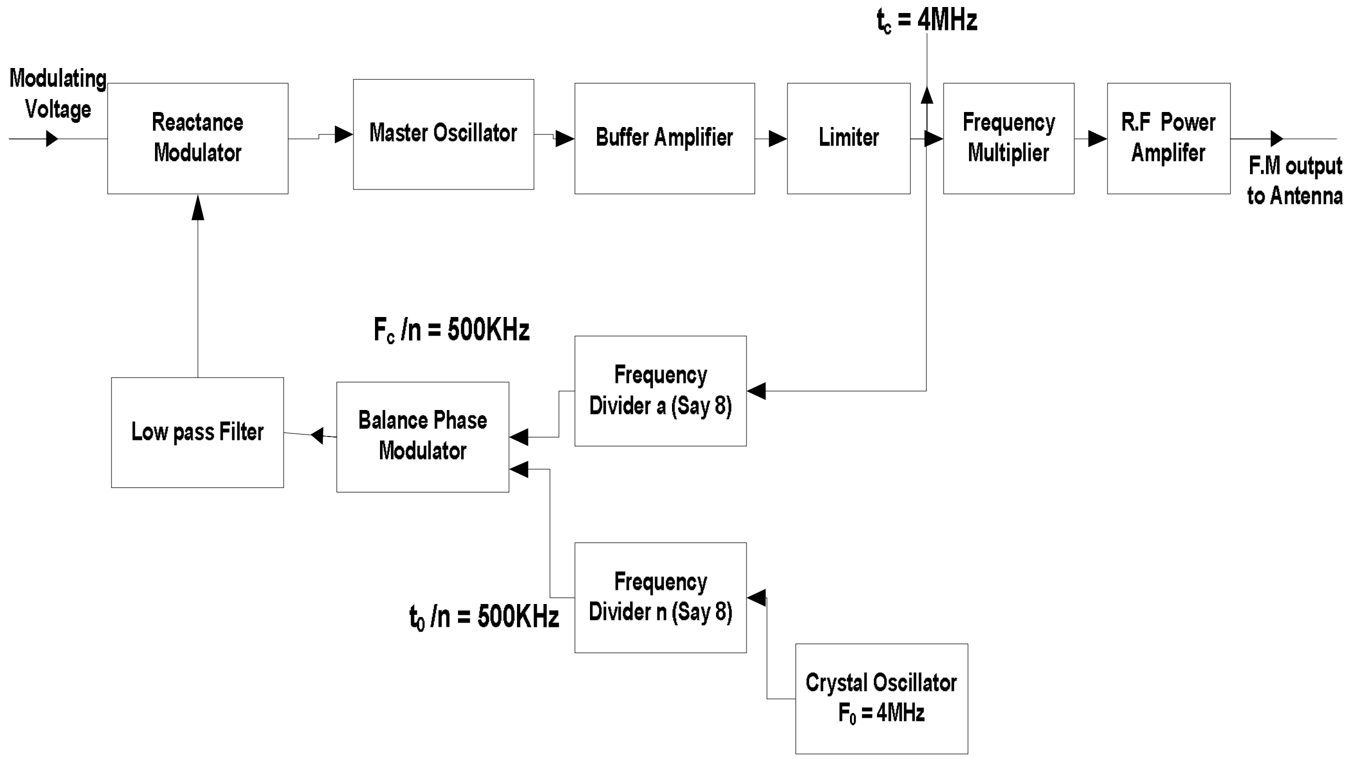


Figure2.12.Reactance modulator FM transmitter using federal Telecommunication Laboratory method of frequency stabilization.

3.13. Armstrong Method of Frequency Modulation

The Armstrong method generates a double sideband suppressed carrier signal, phase shifts this signal, and then reinserts the carrier to produce a frequency modulated signal.A balanced modulator mixes the audio signal and the radio frequency carrier, but suppresses the carrier, leaving only the sidebands. The output from the balanced modulator is a double sideband suppressed carrier signal and it contains all the information that the AM signal has, but without the carrier. It is possible to generate an AM signal by taking the output from the balanced modulator and reinserting the carrier.In the Armstrong method, the audio signal and the radio frequency carrier signal are applied to the balanced modulator to generate a double sideband suppressed carrier signal. The phase of this output signal is then shifted 90 degrees with respect to the original carrier. The balanced modulator output can either lead or lag the carrier's phase. The double sideband signal and the original carrier signal are then applied to the mixer, and the original carrier 90 degrees out of phase is reinserted. The output from the mixer is a frequency modulated signal.

This method of frequency modulations consists (i) pre-distorting the modulating voltage in an integrator to make its amplitude vary inversely with its frequency and (ii) making this modified modulating voltage to phase modulate the carrier. Thus, consider the expressions for frequency modulated voltage and phase modulated voltage already derived (Module 2, Unit 2) and reproduced below.

For frequency modulated voltage: $u = V_c \sin\left[\omega_c t - \frac{k_f V_m}{\omega_m} cos\omega_m t\right]$ (2.40)

For phase modulated voltage: $v = V_c \sin\left[\omega_c t + K_p sin\omega_m t\right]$ (2.41)

Where unmodulated carrier voltage is given by,

$v = V_c \sin\omega_c t$ (2.42)

and the modulating voltage is given by, $v = V_m \sin\omega_m t$ (2.43)

Obviously then frequency modulated voltage as given by Equ. (2.40/7.40) may be obtained by actually using phase modulation and pre- distorting the modulating voltage such that its amplitude varies inversely with its carrier. Now, from Figure 2.14

$$\frac{V_0}{V_i} = \frac{[R_2/j\,\omega C]/\left[R_2+\frac{1}{j\omega C}\right]}{R_1+[R_2/j\,\omega C]\ /\ \left[R_2+\frac{1}{j\omega C}\right]} = \frac{1}{\left(1+\frac{R_1}{R_2}+j\,\omega\,CR_1\right)} \quad (2.44)$$

If $R_2 >> R_1$ then $\left(1 + \frac{R_1}{R_2}\right) \approx 1$.

Further if frequency ω is high enough, $\omega\,CR_1 \gg 1$, Hence, the Equ. (2.44) may be written as

$$\frac{V_0}{V_1} \approx \frac{1}{\omega\,CR_1} \quad (2.45)$$

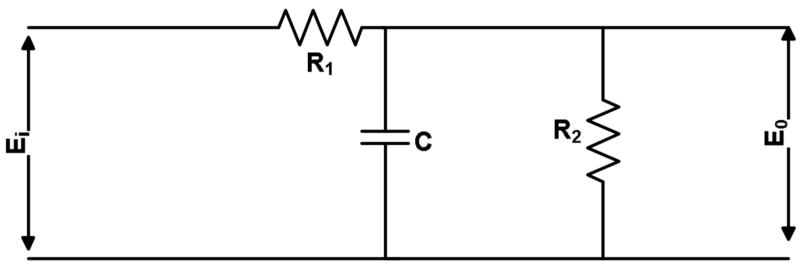


Figure2.13. Basic pre-distorted circuit

This voltage V_0 then phase modulates the carrier to produce frequency modulation. It is pertinent to note that in Armstrong modulation method, phase modulation is produced by special amplitude modulation method. The modulating voltage amplitude modulates the carrier in a balanced modulator to produce the two sidebands alone. Both these sidebands are shifted in phase 900 and are added to the carrier. This result in pure phase modulation of the carrier provided that the resulting phase deviation ad is less than 0.5.

Typically, prior to some processes such as transmission over cable, or recording to phonograph record or tape, the input frequency range most susceptible to noise is boosted. This is referred to as pre-emphasis i.e., the process the signal will undergo. Later, when the signal is received, or retrieved from recording, the reverse transformation is applied (i.e., de-emphasis) so that the output accurately reproduces the original input. Any noise added by transmission or record/playback, to the frequency range previously boosted, is now attenuated in the de-emphasis stage. The high-frequency signal components are emphasized to produce a more equal modulation index for the transmitted frequency spectrum, and therefore a better signal-to-noise ratio for the entire frequency range. Emphasis is commonly used in FM broadcasting. These are discussed below.

3.14. Pre-Emphasis and De-Emphasis

We have discussed these concepts previously; however, we shall still revisit it here to refresh previous understanding. In processing electronic audio signals, pre-emphasis refers to a system process designed to increase (within a frequency band) the magnitude of some (usually higher) frequencies with respect to the magnitude of other (usually lower) frequencies in order to improve the overall signal-to-noise ratio by minimizing the adverse effects of such phenomena as attenuation distortion or saturation of recording media in subsequent parts of the system. The mirror operation is called de-emphasis, and the system as a whole is called emphasis.Pre-emphasis is achieved with a pre-emphasis network which is essentially a calibrated filter. The frequency response is decided by special time constants. The frequency can be calculated from that value.Pre-emphasis is commonly used in telecommunications, digital audio recording, record cutting, in FM broadcasting transmissions, and in displaying the spectrograms of speech signals.

Now, in the modulating voltage, high frequency components have small amplitudes and hence these components produce frequency deviation far less than the maximum permitted value of 75 kHz. Consequently the signal strength at these high modulating frequencies is very low relative to the transistor and circuit noises which are uniformly distributed over the entire spectrum occupied by the channel. To improve signal/noise ratio at these high modulating frequencies, pre-emphasis circuits are used to emphasize the high frequency components prior to modulation. The F.M. receiver then has the corresponding de-emphasis to restore the relative amplitudes of all modulating frequencies. The pre-emphasis circuit generally follows the following relation:

$$\frac{V_0}{V_1} = \frac{1}{\sqrt{1+\left(\frac{\omega_1}{\omega}\right)^2}} = \frac{1}{\sqrt{1+\left(\frac{f_1}{f}\right)^2}} \qquad (2.52)$$

The usual value of $1/\omega_1$ is equal to 50 µS or f_1 = 3100 Hz. Ratio $\frac{V_0}{V_1}$is almost unity for frequency greater than 31,000 Hz, is equal $1\sqrt{2}$ for frequency f= 3100 Hz and equal to 1/10 for f= 310 Hz. Thus the high frequency terms are relatively emphasized.

Figure2.16 shows a plot of ratio $1/\sqrt{1+(f_1/f)^2}$ for different values of f and for values off_1 equal to 1600 Hz and 3100 Hz corresponding to time constant of pre-emphasis 100 µs 75 and 50 µs, respectively.

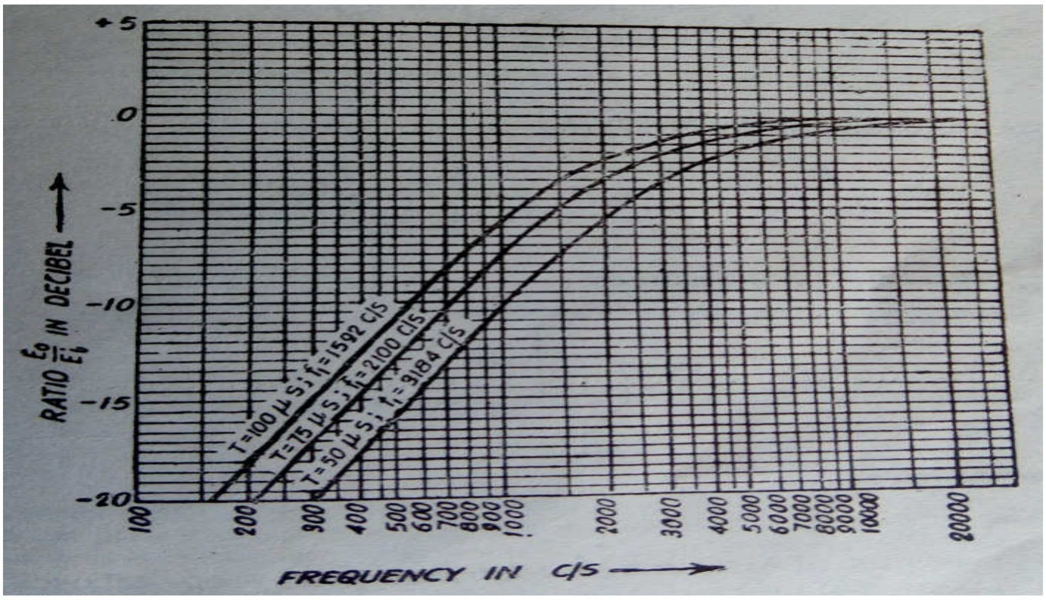
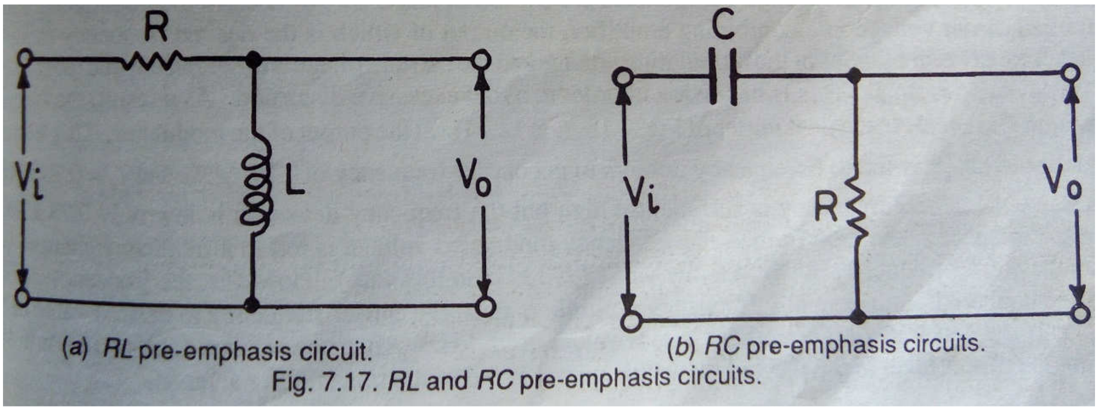


Figure 2.16. Response of pre-emphasis circuit.

This pre-emphasis circuit may use either RL or RC circuit as shown in Figure 2.17.



(a) RL pre-emphasis circuit. (b) RC pre-emphasis circuits.

Figure 2.17. RL and RC Pre-Emphasis circuits.

In RL pre-emphasis circuit of Figure 2.17a, the transfer ratio is given by

$$\frac{V_0}{V_1} = \frac{j\,\omega L}{R + j\,\omega L} = \frac{1}{1+\frac{R}{j\,\omega L}} = \frac{1}{\sqrt{1+\left(\frac{R}{\omega L}\right)^2}} tan^{-1}\frac{R}{\omega L} \qquad (2.53)$$

If reference frequency $\omega_1 = {}^{R}/_{L}$ then Equ.(2.53)becomes,

$$\frac{V_0}{V_1} = \frac{1}{\sqrt{1+\left(\frac{\omega_1}{\omega L}\right)^2}} tan^{-1}\frac{\omega_1}{\omega} \quad (2.54)$$

In the RC, pre-emphasis circuit of Figure2.17b,

$$\frac{V_0}{V_i} = \frac{R}{R+\frac{1}{j\,\omega C}} = \frac{1}{1+\frac{1}{j\,\omega C}} = \frac{1}{\sqrt{1+\left(\frac{1}{\omega CR}\right)^2}} tan^{-1}\frac{1}{\omega CR} \quad (2.55)$$

If reference, $\omega_1 = \frac{1}{CR}$,then Equ. (2.55) becomes

$$\frac{V_0}{V_i} = \frac{1}{1+\frac{\omega_1}{\omega}} = \frac{1}{\sqrt{1+\left(\frac{\omega_1}{\omega}\right)^2}} tan^{-1}\frac{\omega_1}{\omega} \quad (2.56)$$

Furthermore, de-emphasis in FM systems is the complement of pre-emphasis in the anti-noise system called emphasis. De-emphasis is a system process designed to decrease, (within a band of frequencies), the magnitude of some (usually higher) frequencies with respect to the magnitude of other (usually lower) frequencies in order to improve the overall signal-to-noise ratio by minimizing the adverse effects of such phenomena as attenuation distortion or saturation of recording media in subsequent parts of the system.

In summary,the following are valid for Pre-emphasis and De-emphasis circuits viz:

1. At the transmitter the modulating signal is passing through a simple network which amplifies the high frequency component more the low-frequency component. The simplest form of such circuit is a simple high pass filter.

2. The pre-emphasis circuit increases the energy of the higher content of the higher-frequency signals so that will tend to become stronger than the high frequency noise component. This improves the signal-to-noise ratio.

3. To return the frequency response to its normal level, a de-emphasis circuit is used at the receiver. This is a simple low-pass filter

4. The de-emphasis circuit provides a normal frequency response.

5. The combined effect of pre-emphasis and de-emphasis is to increase the high-frequency components during the transmission so that they will be stronger and not masked by noise.

3.15. Armstrong F.M. Transmitter

Figure 2.18a gives the block diagram of Armstrong F.M. transmitter. A crystal oscillator is used to provide the stable carrier frequency. A portion of the carrier voltage is fed to a suppressed carrier balanced modulator to which is also fed the pre-emphasized pre-distorted and amplified modulating signal. The modulating voltage amplitude is then proportional V_m/f_m, where V_m is the original modulating voltage amplitude and f_m is its frequency. The resulting sideband components are shifted in phase by 90° and are then combined with the amplified carrier voltage in a combining amplifier, the output of which is the desired frequency modulated voltage. Relative amplitudes of the modulating voltage and the carrier voltage are so adjusted that the maximum phase deviation is small. This is necessary in order to avoid excessive distortion. As a result, the frequency deviation f_d is small.

In the transmitter of Figure 2.18, f_d is 12.2 Hz at the output of the modulator. This frequency modulated voltage is fed to 6 frequency doublers to get carrier frequency of 12.8 MHz and f_d = 0.78 kHz. At this stage, the carrier frequency is sufficiently high but the frequency deviation is low only 0.78 kHz. To increase frequency deviation further, the frequency modulated voltage is fed to a frequency changer which shifts the carrier frequency down to a low value (993 kHz in this case). However, the frequency deviation remains unaltered. Subsequently frequency multipliers raise the carrier frequency to desired value of 95.3 MHz and frequency deviation to the standard value, of 75 kHz. After power amplification in a few stages raising the power level to the desired value, the output is fed to the transmitting antenna. Figure2.18(7.18b) shows the block diagram of indirect method of generating a wide band FM signal.

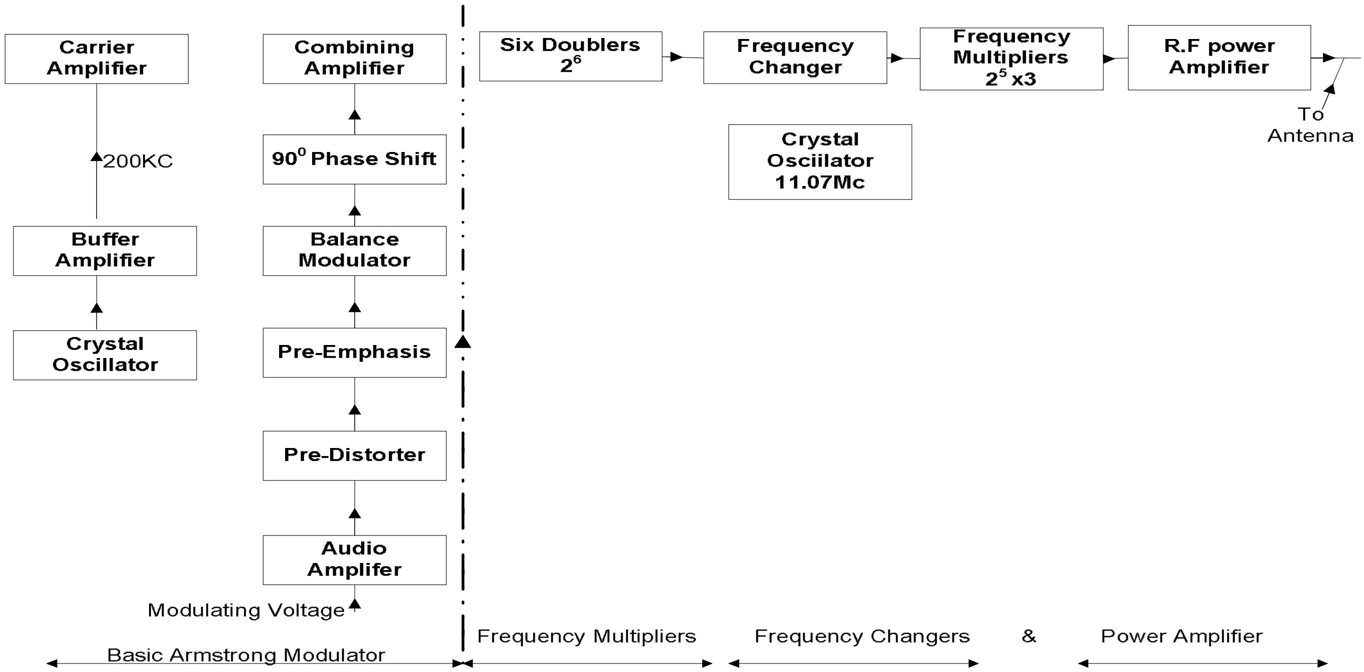


Figure 2.18a. Block diagram of Armstrong FM Transmitter.

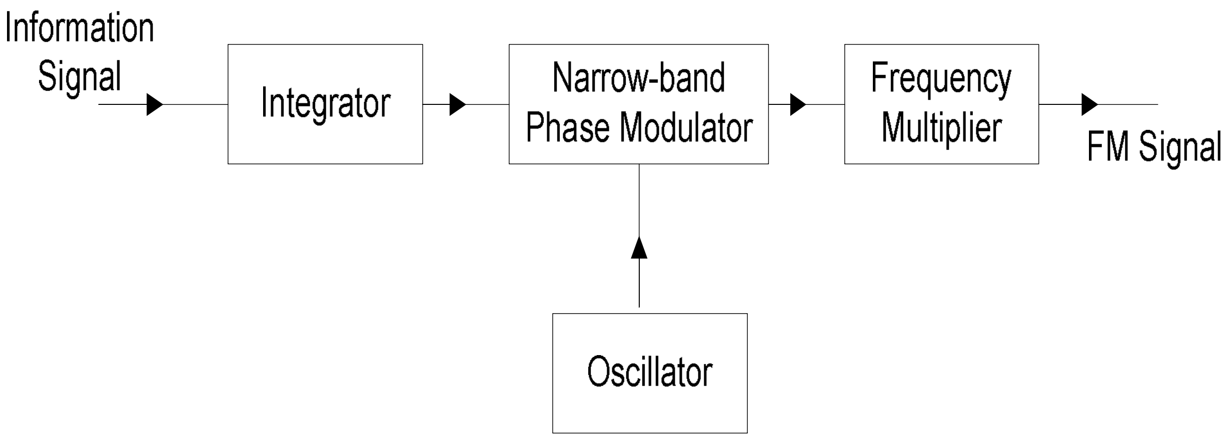


Figure 2.18b. Block diagram of Armstrong FM Transmitter.

4.0 Conclusion

This unit has explained the concept of frequency modulation in a transmitter system while distinguishing between the direct and indirect modulation techniques. We have presented mathematical expressions on modulation index in an FM system. Frequency stabilization in FM and frequency Drift in Reactance Modulator FM Transmitter are discussed. We have understood the need for Frequency Mixing. Indirect methods such as Armstrong frequency modulation as well as the role of pre-emphasis and de-emphasis in FM broadcasting are all discussed.

5.0 Summary

There are two major methods of FM Generation viz: the direct and indirect methods. The two types of direct methods include reactance modulator and varactor modulator while Armstrong method is the major type of indirect modulator.

- **In Reactance Tube Modulation this** method of frequency modulation makes use of a reactance tube, i.e. an electron tube which offers between its plate and cathode terminals, a reactance varying in accordance with the modulating voltage amplified between its control grid and cathode. This reactance tube is placed in shunt with the tuned circuit of an oscillator.
- **Capacitive Reactance Tube.** The tube offers a capacitive reactance between its plate and cathode terminals. This may be realized by a pentode with a capacitor C connected between anode and control grid and a resistor R between control grid and cathode. Further $R<<1/\omega C$. Then reactance tube behaves as a capacitance of value g_mCR.
- **Inductive Reactance Tube**. The tube offers a inductive reactance between its plate and cathode terminals. It may be realized by a pentode with a capacitor C connected between control grid and cathode and a resistor R between plate control grid. Further we make $R<<1/\omega C$. Then reactance tube behaves as an inductor of value CR/g_m.
- **Reactance FET**. An **FET** may be made to offer between its drain and source terminals, a capacitive or inductive reactance which varies in accordance with the modulating voltage applied between its gate and source. This reactance **FET** may be connected in shunt with the tuned circuit of an oscillator resulting in frequency modulation.
- **RC Capacitive Reactance FET**. The FET offers a capacitive reactance between its drain and source terminals if (i) a capacitor C is connected between its drain and the gate (ii) a resistor R is connected between its gate and source and (iii) $X_c>>R$. Then the capacitance offered equals $C_e = g_mCR$.
- **RL Capacitive Reactance FET**. An FET offers a capacitive reactance between its drain and source terminals if (i) a capacitor R is connected between the drain and the gate (ii) an inductor L is connected between its gate and source and (iii) $R>>X_L$. Then the capacitance offered equals $C_e = g_mL/R$.
- **RC Inductive Reactance FET**. An FET offers an inductive reactance between its drain and source terminals if (i) a resistor R is connected between drain and the gate (ii) a capacitor C is connected between gate and (iii) $R>>X_L$. Then the inductance offered equals $L_e = RC/g_m$.
- **RL Inductive Reactance FET**. An FET offers an Inductive reactance between its drain and source terminals if (i) an inductor L is connected between its drain and gate (ii) a resistor R is connected between gate and source and (iii) $X_L>>R$. Then the inductance offered equals $L_e = L/(g_mR)$.

- **RC Capacitive Reactance FET Modulator.** It uses an RC capacitive reactance FET placed across the tuned circuit of the oscillator to be frequently modulated. The modulating voltage is connected in series with the resistor R between gate and source. The oscillators popularly used are Hartley, Colpitts and Clapp.
- **Reactance BJT**. A BJT in similarly with FET offers a capacitive or inductive reactance between collector and emitter terminals by placing phase splitting network Z_1-Z_2 and keeping $Z_1 >> Z_2$. Thus for RC capacitive reactance BJT, a capacitor C is placed between collector and base and a resistor R is placed between base and emitter. Further we make $X_c >> R$.
- **RC Reactance Modulator using BJT**. It uses and RC capacitive reactance BJT placed across the tank circuit of the oscillator to be frequently modulated. The modulating voltage is connected in series with the resistor R between the base and the emitter. The oscillator to be frequency modulated is typically a Hartley, Colpitts and Clapp oscillator.
- **Frequency Modulation using Varactor Diode**. Frequency modulation may be achieved on using the variable junction capacitance of a reverse biased varactor diode placed across the tuned circuit of an oscillator.
- **FM Transmitter using Reactance Modulator**. In this transmitter, the reactance **FET** or **BJT** frequency modulates the master oscillator which is followed by a buffer amplifier, frequency multipliers and RF amplifier.
- **Frequency Drift in Reactance Modulator Transmitter.** Frequency drift may result due to (i) variations in supply voltage (ii) variations in temperature and humidity and (iiI) aging of electron device.
- **Reactance Modulator FM Transmitter using AFC Frequency Stabilization**. In this method, frequency of master oscillator (say 4MHz) is compared with the frequency of crystal oscillator (say 3.5 MHz) in a frequency mixer to give standard difference frequency of 500 kHz. This is amplified and fed to a discriminator. The discriminator gives at its output a dc voltage which is positive or negative depending upon whether actual input frequency is greater or less than the prescribed frequency and the magnitude of this D.C . voltage depends on the frequency shift Δf. This D.C voltage from discriminator fed to the base of the reactance BJT is in such a polarity that the resulting change in the oscillator frequency tends to cancel the frequency drift.
- **Armstrong Method of Frequency Modulation.** This method of frequency modulation consists in (i) pre-distorting the modulating voltage in an integrator to make its amplitude vary inversely with its frequency and (ii) making this modified modulating voltage to phase modulate the carrier.
- **Pre-emphasis**. It consists in emphasizing, i.e. providing higher gain, to high frequency components in the modulation signal prior to modulation. Typical pre-emphasis circuit consists in using a series capacitor and shunt resistor. Typical time constant is 50µs.

6.0 Tutor-Marked Assignment (TMA)

1. Discuss the various methods of Frequency modulation. Draw the circuit of a capacitive reactance tube modulator and derive expression for the modulation index
2. (a) Draw the circuit of an inductive reactance tube using R-C network and explain how the tube behaves as a inductance? (b) What is the expression for effective inductance offered by this reactance tube ?
2. Derive expression for the modulation index of inductive reactive tube modulator using R-C phase-shifting network.
3. a)What are the reasons for distortion in reactance tube modulator? What means are adopted to keep this distortion low? (b) Draw the basic circuit of a reactance FET and describes how it behaves as a reactance. Derive expression for the output admittance between drain and source.
4. a) Draw the circuit and explain the working of a frequency modulator using a varactor diode. Draw the circuit and explain the working of reactance modulator FM transmitter using **AFC** frequency stabilization.
5. (a) Give the principle of Armstrong method of frequency modulation. (b) What are the sources of distortion in' this method and how can this distortion be kept low.
6. (a) Define by pre-emphasis and why is it necessary? (b) Draw an RC pre-emphasis circuit and obtain expression for ratio of output voltage to input voltage, (c) Draw the block diagram of Armstrong FM transmitter and describe its working.

7.0 References/Further Readings

1. Seely S. (1958). 'Electron Tube Circuits', McGraw Hill Book Company, New York,
2. Terman F.E (1955): 'Electronic and Radio Engineering,' McGraw Hill800k Company, New York,.
3. Kennedy, G (1977). Electronic Communication Systems: Second Edition, McGraw Hill Kogakusha Ltd., Tokyo.
4. Ryder J.D (1978). 'Electronic Fundamentals and Applications: 5th Edition, Prentice Hall of India Ltd., New Delhi,.
5. The ARRL Handbook for Radio Communication, American Radio Relay League, 2008, p. 9.30
6. Hayward, Campbell, and Larkin (2003). Experimental Methods in RF Design, American Radio Relay League,, p. 9.49.
7. FM Signal generation (2017). URL: http://www.mmumullana.org/downloads/files/n5474524fb4134.pdf

UNIT 3 ULTRA HIGH FREQUENCY TUBES AND OSCILLATORS

1.0 Introduction

2.0 Objectives
3.0 Main Content
3.1. What is Ultra high frequency (UHF)
3.2. Ultra-High Frequency Responses
3.2.1. Lead Inductance and Internal Capacitances
3.2.2. Grid-To-Plate Capacitance Vs Input Admittance
3.2.3. Cathode Lead Inductance and Grid-Capacitance C_{gk} Vs Input Admittance
3.3. Tube Output Conductance
3.4. Transit Time Effect in Diodes
3.5. High Frequency Tube Current Flow
3.5.1. High Frequency Tube Condition of Current Flow
3.5.2. Graphical Development of Current of Flow
3.5.3. Transit Time effect in Triodes
3.6. Ultra High Frequency Tubes
3.7. Frequency Limit of Ordinary Electron Tube Oscillator.
3.7.1. Lead inductance and Inter-electrode Capacitances
3.7.2. Reduction in Tank Circuit Q at High Frequency
3.7.3. Grid Loading caused by C C_{gk} and L_k and Transit time.
3.8. Remedies for the Limitations
3.9. Typical U.H.F Tube Construction
3.9.1. Single Ended U.H.F. Tubes
3.9.2. Double Ended U.H.F Tube
3.9.3. Disk Seal Construction
3.9.4. Elimination of the Effects of Inter-electrode Capacitances
3.9.5. Grounded Grid Operation of Disk Seal Triodes
3.10. U.H.F Plate Cooling
3.11. U.H. F Filament Emission.
3.12. Oscillators in UHF Domain
3.12.1. Single Tube UHF Oscillators
3.12.2. Lighthouse Tube Oscillator.
3.12.3. Single Tube Oscillator using Open Wire Line
3.12.4. UHF Pushpull Oscillators
3.13. Parasitic Oscillation and Their Suppression
3.14. Multicavity Magnetron
3.15. Magnetron Oscillator Conventional Tube Design
3.15.1. Hull or Single-Anode Magnetron
3.15.2. Split-Anode Magnetron
3.15.3. Cavity Magnetron
3.15.4. Common Features of Magnetron
3.16Klystron
3.16.1. Oscillator Definition
3.16.2. Reflex Klystron
3.16.3. Reflex klystron Oscillatory Mechanism
3.16.4. Multi-cavity Klystron Amplifier
3.16.5. Performance of Multicavity Klystron Amplifier
3.17. Semiconductor Substitute

4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0 References/Further Readings

1.0 INTRODUCTION
The recent activity in the development of ultrahigh-frequency communication has created a demand for transmitting tubes having carrier power outputs of the order of 30 to 60 watts at frequencies as high as 150 megacycles and in some instances as high as 250 megacycles. In order to clarify the tube-design requirements it is well to enumerate the more common uses to which tubes are applied in ultra-high-frequency communication. These uses include police, aviation, marine, television, amateur, and point-to-point communication and employ continuous wave transmission, amplitude modulation, or frequency modulation. A survey of these uses yields some interesting requirements which can be divided in two classes: (1) those imposed by service conditions; and (2) those imposed by the frequency spectrum in which the tubes are operated. Conventional tubes intended for operation at very high frequency depart for shapes, connections and electrode structures from the traditional ones. Until World War II (WWII) the research for vacuum tubes capable of operating at very high frequencies was essentially pushed by communication markets. Military carefully watched at new high frequency tubes usable for radio localization sets, still experimental in several countries and later known as radars. Anyway, we must consider that with few exceptions before WWII, the early advanced television transmitters used frequencies just above the limit of short waves. In 1939 B.B.C,an experimental high-quality transmission at 45MHz commence while RCA operated its vision antenna on the Empire State Building at some 54MHz. This unit will look at the requirements of tubes suitable for operation in a frequency spectrum. Also, oscillators are discussed also.

2.0 Objectives
After going through this unit, you should be able to
- ✓ explain the meaning of Ultra high frequency (UHF)
- ✓ Understand Ultra High Frequency Tubes
- ✓ Explain Parasitic Oscillation and suppression
- ✓ Understand Multicavity Magnetron and the various types
- ✓ Distinguish between semiconductor and tube oscillators

3.0 Main Content

3.1. What is Ultra high frequency (**UHF**)?

One might ask, what is UHF? UHF is the ITU designation for radio frequencies in the range between 300 megahertz (MHz) and 3gigahertz(GHz), also known as the decimeter band as the wavelengths range from one meter to one decimeter. Radio waves with frequencies above the UHF band fall into the SHF (super-high frequency) or microwave frequency range. Lower frequency signals fall into the VHF (very high frequency) or lower bands. UHF radio waves propagate mainly by line of sight; they are blocked by hills and large buildings although the transmission through building walls is strong enough for indoor reception. They are used for television broadcasting, cell phones, satellite communication including GPS, personal radio services including Wi-Fi and Bluetooth, walkie-talkies, cordless phones, and numerous other applications.

At low frequency, $i.e$ up to about 300 KHz (for the purpose of analysis), an electron tube may be represented by a simple a.c. equivalent circuit. Thus, an electron tube with sinusoidal input voltage E_{gk}may be represented (i) by a voltage generator of open circuit voltage $\mu\, E_{gk}$and internal series resistance $r_{p.}$ Either of these equivalent circuit serves to represent truly the electron tube and explain all the observe phenomena. However, as the input signal frequency increase to a few Mega-Hartz, inter-electrode capacitance becomes significant. Also the stray wiring capacitances come in shunt with the inter-electrode capacitance. These capacitance are then required to be included in the equivalent circuit representing the electron tube at such high frequencies. When the frequencies of electron exceed about 30MHZ, the following two quantities assume significance (i) self and mutual inductance of electrode leads and (ii) electron transit time from cathode to mode. Consequently, at ultra-high frequencies, we have to take into account the following three quantities in representing the electron tube truly: (i) inter-electrode capacitances, (ii) lead inductance and (iii) electron transit time. We shall study the elements with respect to ultra high frequency tubes below.

3.2.Ultra-High Frequency Responses

3.2.1. Lead Inductance and Internal Capacitances

Lead inductance as well as inter-electrodecapacitance is referred to as the distributed elements.But for the purpose of analysis, we may consider these as lumped quantities.Figure3.1 shows a triode along with the three inter-electrode capacitancenamelyC_k, C_{pk} and C_{gp} and three leadinductance namely L_k, L_p and L_g. Thesecapacitance are small, being the order of 5 $\mu\mu$ F for receiving tubes.

We first consider the effect of grid-to-plate capacitance C_{gp} alone on the input admittance of the triode; subsequently we will consider the effect of cathode lead inductance L_k and grid-ti-cathode capacitance C_{gk} on the admittance. Hence, the effect of Input admittance on the variable of Figure 3.1 shall be discussed below.

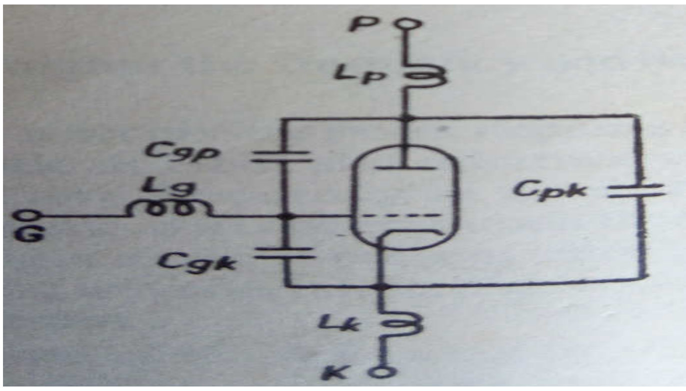


Figure3.1. Triode along with lead inductances and inter-electrode capacitances

3.2.2. Grid-To-Plate Capacitance Vs Input Admittance

When you look at Figure 3.1, out of six reactive elements, the one which is of significance even at low frequencies is the grid-to-plate capacitance C_{gp}. The input admittance caused by its presence is considered here.

The assumptions made are(i) C_{gp} is a lumped capacitance and (ii) lead inductance are negligible.Figure3.2 (a) show the circuit of a triode amplifier after inclusion of admittance Y_{gp} of C_{gp} Figure3.2(b) show the A.C equivalent circuit.By application of Millman theorem between points *P* and *K*, we get

$$E_o = \frac{E_i Y_{gp} - \mu\, E_i\, E_i\, Y_p}{Y_{gp} + Y_p + Y_t} \qquad (3.1)$$

Where, $Y_p = \frac{1}{r_p}$ (3.2)

But, Millman's theoremor the parallel generator theorem) is a method to simplify the solution of a circuit. Specifically, Millman's theorem is used to compute the voltage at the ends of a circuit made up of only branches in parallel.

Also, $I_i = Y_{gp}\,(E_i - E_O)$

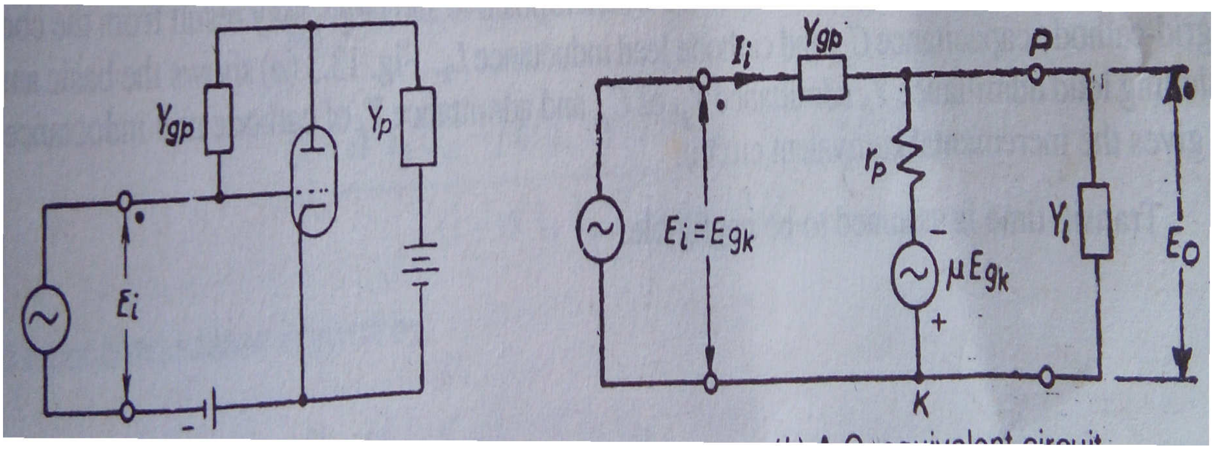


Figure3.2. Triode amplifier including $C_{gp.}$

Substituting the value of E_o from Equ. (3.1) into Equ. (3.2), we get

$$I_i = Y_{gp}\left[E_i - E_i \frac{\mathrm{Y_{gp}}}{\mathrm{Y_{gp}}} \frac{+}{+} \frac{\mu\, \mathrm{Y_p}}{Y_p + Y_1}\right] = Y_{gp}.E_i \left[\frac{\mathrm{Y_1 + Y_p + g}}{Y_{gp} + Y_p + Y_t}\right] \qquad (3.3)$$

Hence, input admittance is given by,

$$Y_i = \frac{1}{E_i} = Y_{gp} - \frac{Y_t + Y_p + g_m}{Y_t + Y_p + Y_{gp}} \qquad (3.4)$$

$$= j\omega\, C_{gp} \frac{Y_t + Y_P + g_m}{Y_t + Y_p + j\,\omega\, C_{gp}} \qquad (3.5)$$

From Equ. (3.5), we find out that Y_t becomes infinite if $(Y_t + Y_p) = j\,\omega\, C_{gp}$. For values of$(Y_t + Y_p$, near this value, the input conductance is large and positive depending upon whether

$$\left(Y_p + Y_t\right) < Y_{gp} or (Y_p + Y_t) > Y_{gp.}$$

Thus, we see that grid-to-plate capacitance may produce an input conductance which is relatively large and either positive or negative. Oscillations are likely to occur if the conductance is negative. This depends upon the magnitude of associated positive conductance. This effect is quite pronounced even at low frequencies. It was the primary factors in the development of screen grid greatly reduce the direct grid-plate capacitance.

If Y_{gp} is neglected in comparison with $(Y_p + Y_i)$ Equ. (3.5) reduces to the following:

$$Y_i = j\,\omega\, C_{gp} \left[1 + \frac{g_m}{Y_t + Y_p}\right] = j\,\omega C_{gp} \left[1 - \frac{\mu\, R_t}{R_p + R_t}\right] \qquad (3.6)$$

Where $R_i = \frac{1}{Y_t}$

Or $Y_i = j\,\omega\, C_{gp}[1 - A]$ (3.7)

Where A is the Voltage gain of the amplifier neglecting C_{gp}.

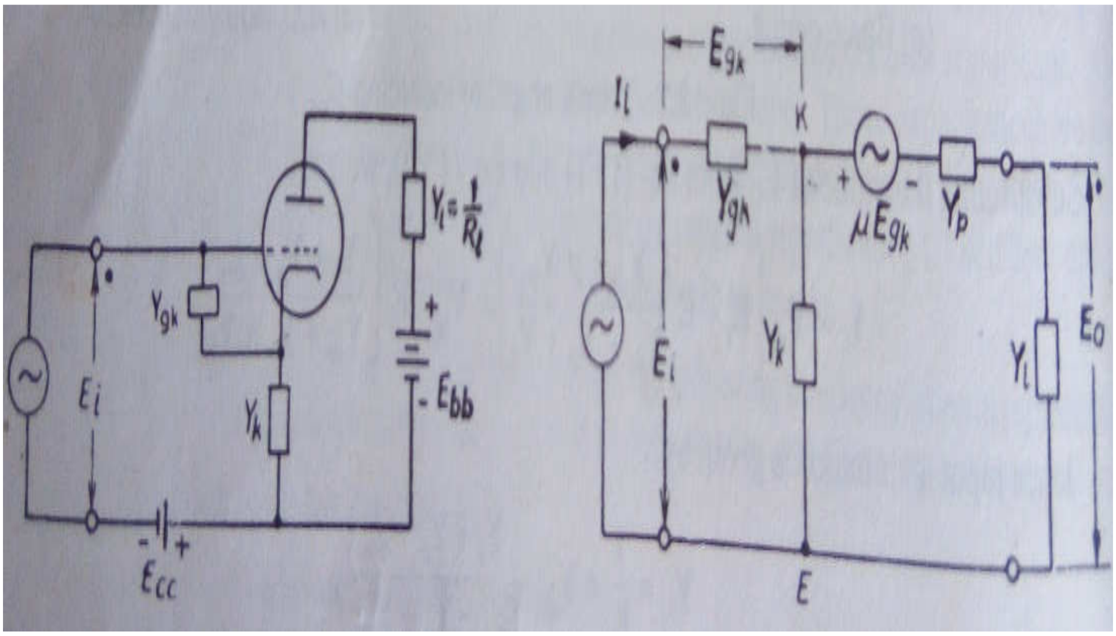
To this input admittance, we may add $j\,\omega\, C_{gp}$ to get total input admittance due to both C_{gk} andC_{gp}.

Thus,$Y_i = j\omega C_{gk} + \omega C_{gp}(1 - A)$ (3.8)

This equation has already been derived earlier.

3.2.3. Cathode Lead Inductance and Grid-Capacitance C_{gk} Vs Input Admittance

In modern tetrodes and pentodes with respect to tube technology, the input admittance due to C_{gp} is negligible or small even at the highest frequency. However, an input conductance which is positive and large may result from the combined action of grid-cathode capacitance C_{gk} and cathode lead inductance L_k Figure3.3(a) show the basic amplifier circuit including load admittance Y_t, admittance Y_{gk} of C_{gk} and admittance Y_k of cathode and inductance L_k. Figure3.3(b) gives the incremental equivalent circuit.Transit time assumed to be negligible.



(a) Basic circuit. (b) Incremental equivalent circuit

Figure3.3. Basic circuit and incremental equivalent circuit of vacuum tube amplifier having included $\boldsymbol{C_{gk}}$ and L_k

Application of Millman theorem between point k and E yields.

$$\left(E_i - E_{gk}\right) = \frac{E_1\, Y_{gk} + \mu\, E_{gk} Y_{pt}}{Y_{gk} + Y_k + Y_{pt}} \quad (3.9)$$

Where, $Y_{pt} = \frac{1}{r_p + R_t}$

But, $Y_{gk}\, E_{gk} = I_i$ (3.10)

On substituting value of E_{gk} from Equ. (3.10) into Equ. (3.9), we get

$$E_i - \frac{I}{Y_{gk}} = \frac{E_i\, Y_{gk} + \mu\, Y_{pt} \frac{I_i}{Y_{gk}}}{Y_{pt} + Y_k + Y_{gk}} \quad (3.11)$$

Or $E_i\left[Y_{pt} + Y_k + Y_{gk} - Y_{gk}\right] = \frac{I_i}{Y_{gk}}\left[Y_{pt} + Y_k + Y_{gk}\right] + \mu Y_{pt} \frac{I_i}{Y_{gk}}$

Or $Y_i = \frac{I_i}{E_i} = \frac{Y_{gk}[Y_{pt} + Y_k}{[Y_k + Y_{gk} + (\mu+1) Y_{pt}}$ (3.12)

But$Y_{pt} = \frac{1}{r_p + R_t} = \frac{1}{r_{pt}}$

Also, $(\mu + 1)Y_{pt} = \mu \frac{1}{r_p} = g_m$

Hence, $Y_i = \frac{j\omega\, C_{gk}\left[\frac{1}{r_p} + \frac{1}{j\omega L_k}\right]}{\frac{1}{j\omega\, L_k} + j\omega\, C_{gk} + g_m} = \frac{j\omega\, C_k\left[1 + \frac{j\omega\, L_k}{r_p}\right]}{1 + j^2\, \omega^2\, L_k\, C_{gk} + g_m\, j\omega\, L_k}$

$$= \frac{\left[1 - \omega^2\, L_k\, C_{gk} - j\omega g_m L_k\right]\left[j\omega C_k - \omega^2 g_m\, C_k\, \frac{1}{r_p}\right]}{(1 - \omega^2 L_k C_{gk})^2}$$

Hence input conductance is given by,

$$G_i = \frac{(1 - \omega^2 L_k C_{gk})\left(-\omega^2 L_k\, C_{kg} \frac{1}{r_p}\right) + \omega^2 g_m L_k C_{gk}}{(1 + \omega^2\, L_k\, C_{gk})^2 + \omega^2\, g_m^2\, L_{k^2}}$$

$$G_i = \frac{\omega^2 L_k C_{gk}\left[-\frac{1}{r_p} + \omega^2\, L_k\, C_{gk}\, \frac{1}{r_p} + g_m\right]}{(-\, \omega^2\, L_k\;\; C_k)^2 + \omega^2 g_m^2\, L_k} \qquad (3.13)$$

Even for the frequencies in ultra frequency range $\omega^2\, L_k\, C_{k.}\, \frac{1}{r_p} \ll (g_m - 1/r_p)$ so that numerator in Equ. (3.13) Simple reduces to $\omega^2 L_k\, C_{gk}\left(g_m - 1/r_p\right)$. Also the denominator in Equ. (3.13) reduces to unity.

Hence, $G_i = \omega^2 L_k\, C_{gk}\left(g_m - \frac{1}{r_p}\right)$ (3.14)

In the case of pentode, $\frac{1}{r_p} \ll\; g_m$ so that,

$G_i \approx \omega^2 L_k C_{gk} g_m$ (3.15)

Equ.(3.15) show that input conductance $G_i$varies directly as the square of theoperating frequencyω. Hence, evenmoderate extension of the operatingfrequency may greatly increase theinput conductance.

Figure3.4 givesthe greatly increase the inputconductance (in $\mu\, \mho$) as caused by C_{gk} and plotted against frequency in MHz for a typical high frequency pentode. Log scale is used for both G_i and the frequency range, the transit time effects are negligible. The term input loading is often used to indicate the input conductance since the conductance from grid to cathode serves to load the circuit which feed the grid.

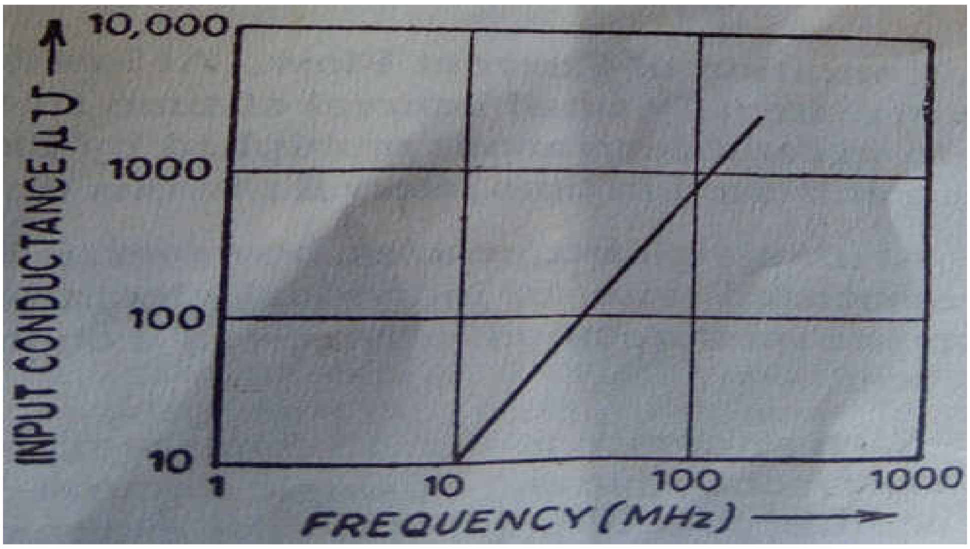


Figure 3.4. Input conductance Vs frequency curve of a typical High frequency pentode.

3.3. Tube Output Conductance

By analysis similar to that following for the input conductance, it may be shown that the output conductance of triode and pentodes is increase due to presence of lead inductance. However, this effect is much less important that of input conductance of often ignored.It is found that the multi-grid tubes are, in general more subject to lead inductance and inter-electrode capacitances that do the triodes. The upper frequency for satisfactory operation of the multi-grid tubes is set by the lead effect, whereas the upper limit for triodes is set by the transit time.

3.4. Transit Time Effect in Diodes

Transit time in tubes of ordinary size and operating at normal potentials is of the order of 10^{-9} second. This period is short compared with the periodic time at lower frequencies and we may consider that the electron reaches the plate immediately if it has been emitted from cathode. At ultra-high frequencies however, transit time is comparable with the periodic time. We therefore, study the manner in which the transmit time effects the operation of an electron tube.

For the purpose of analyzing the effect of transit time, let us that a plane parallel plate diode, with a spacing d between electrodes small compared with the size of the electrodes. The electric field intensity between plates is the uniform. Figure 3.5 shows the arrangement whereby the electric field intensity is given by

$$E = \frac{V}{d} \tag{3.16}$$

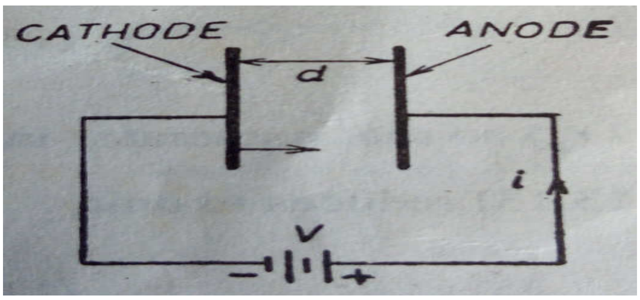


Figure 3.5. Parallel plate diode.

We assume that *(i)* the initial velocity of emission is zero and *(ii)* contact potential is zero.Then the velocity of an electron on reaching the anode is given by

$$v_p = \sqrt{\frac{2g.V}{E}} \qquad (3.17)$$

$$= 5.94\ x\ 10^6\sqrt{V}\ meters/sec \qquad (3.18)$$

Where V is in volts.
Since the acceleration is constant, the average velocity $\bar{v}$ is half v_p.

Hence, transmit time, $t = \frac{d}{\bar{v}} = \frac{d}{2.97\sqrt{V}}\ second$ (3.19)

3.5. High Frequency Tube Current Flow

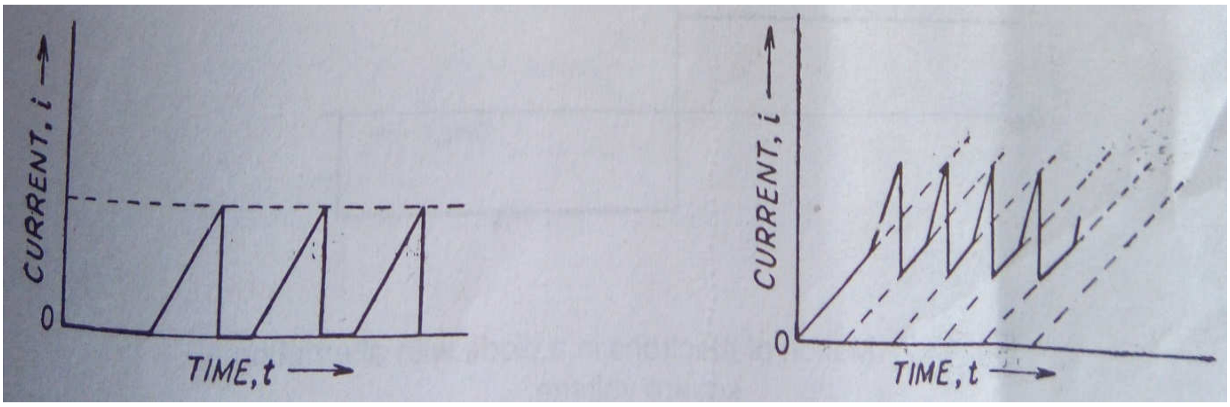
3.5.1. High Frequency Tube Condition of Current Flow

We must pay attention to appreciate the points here. At low frequencies, a large number of electrons move from cathode to anode in extremely small interval of time so that the charge transferred equals the number of electrons that move from cathode to anode multiplied by the charge of an electron. Electric current is then the ratio of the charge transferred in time interval*t* to this time interval *t*.
Now, at ultra-high frequencies, however, it is necessary to know over what interval of time does the current flow associated with an individual electron motion exists. This result may be developed directly by an appropriate interpretation of Maxwell's equations. A development based on the conservation of energy is given here.
With reference to Figure3.5 instantaneous power supplied by the external circuit equals $I\ V$, where i is the instantaneous current. A single electron existing in the cathode anode space is accelerated by the electric field and power supplied by the external circuit is converted into kinetic energy by this process. The mechanical power is given by the product $F_{v,}$ where F is the mechanical force acting and v is the electron velocity.
But, force $F = qE$
Hence,

$$V_i = qEv \qquad (3.20)$$

$$i = \frac{qEv}{V} = \frac{qv}{d} \qquad (3.21)$$

Equ. (3.21) shows that a current that a current flows during the entire interval when the electron is in motion and that the current is proportional to the instantaneous velocity of the electron.
Figure3.6 shows the current flow in the external circuit for two different rates at which electrons are freed from the cathode. In Figure3.6a, the emission rate is slow that one electron reaches that plate before the next electron leaves the cathode. In Figure3.6(b), the emission rate is such that two electrons exist in the cathode-anode space at any time. The sharp corners in these waveforms result in the shot noise.



(a) Slow rate (b) Fast rate

Figure3.6. Variation of current with time in a diode.

Although the above study is made for parallel plane plate configuration, the results apply for cylindrical and other structure. In actual practice, in any electron tube, a large number of electrons are always present simultaneously and the current never falls to zero. The net electronic currents is then the sum of currents contributed by individual electrons as given by $i = q\,v\,/\,d$. To this current must be added the displacement current. The total current is then given by the expression,

$$I = \left(p\bar{v} + \frac{\partial\bar{D}}{\partial t}\right)A \qquad (3.22)$$

Where $p\bar{v}$ is the average of the product of charge density and velocity in the space

$\partial\bar{D}/\partial t$ is the average displacement current density

D is the average electric flux density

and A is the area of the plates.

The above equation assumes the following:

i. Current i is the same everywhere in the external circuit;

ii. The electric field is uniform throughout the space between the plates.

3.5.2.Graphical Development of Current of Flow

We use Figure 3.7 to qualitatively study the transmit time effects. In this case, a square wave voltage is applied to a diode using parallel plane plates and the motion of individual electrons is studied, i.e. effects of space charge are neglected.We know that under uniform acceleration the distance travelled by a body is proportional to the square of the elapsed time. Accordingly the distance curves of Figure 3.7 are parabolic.

Similarly, the kinetic energy of each electron is directly proportional to the distance travelled. It is readily seen that electrons which leave the cathode just before the reversal of voltage are not able to reach the anode and return to the cathode. Since the accelerating force is equal is both halves of the wave but opposite in directions, the distance travelled by the electron at the time of the reversal is exactly double before the electron turns back towards the cathode. Hence, the electrons which are already more than half way to the plate at the time of the reversal of the voltage reach the plate. The electrons which are less than half way to the anode are reversed and returned to the cathode.

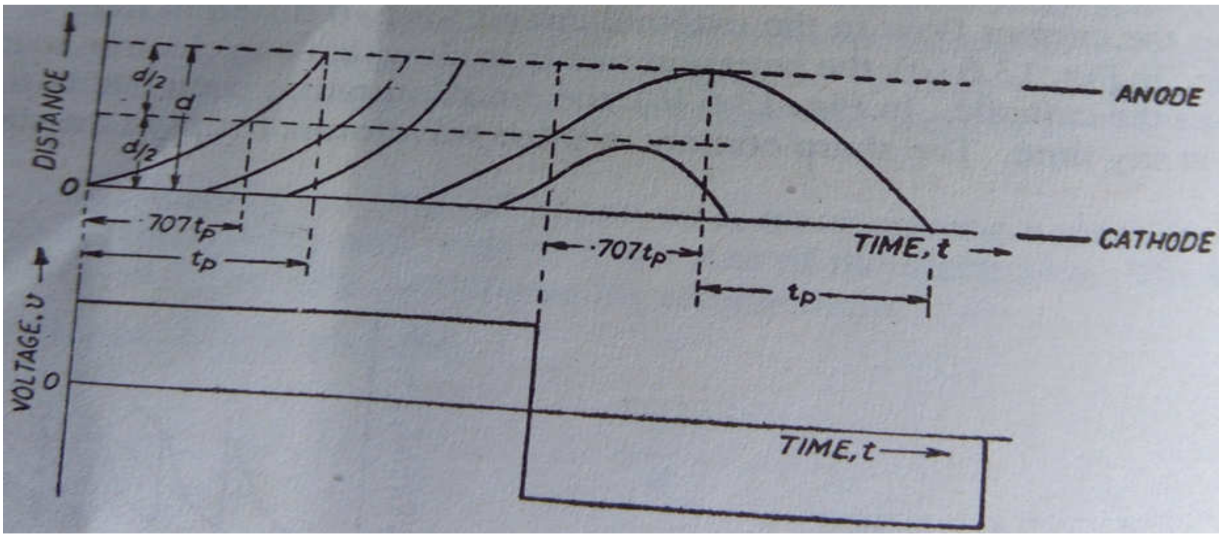


Figure 3.7. Motion of electrons in a diode with alternating square voltage.

A current wave corresponding to Figure3.7 may be deduced assuming that a large number of electrons are emitted at a uniform rate and that the space charge is negligible. The electronic current may then be competed considering the number of electrons in motion and there average velocity.

Figure3.8shows this electronic current as a function of time. Also shown is the variation of velocity of electrons emitted at different times during the positive half of the displacement current in the case of square wave.

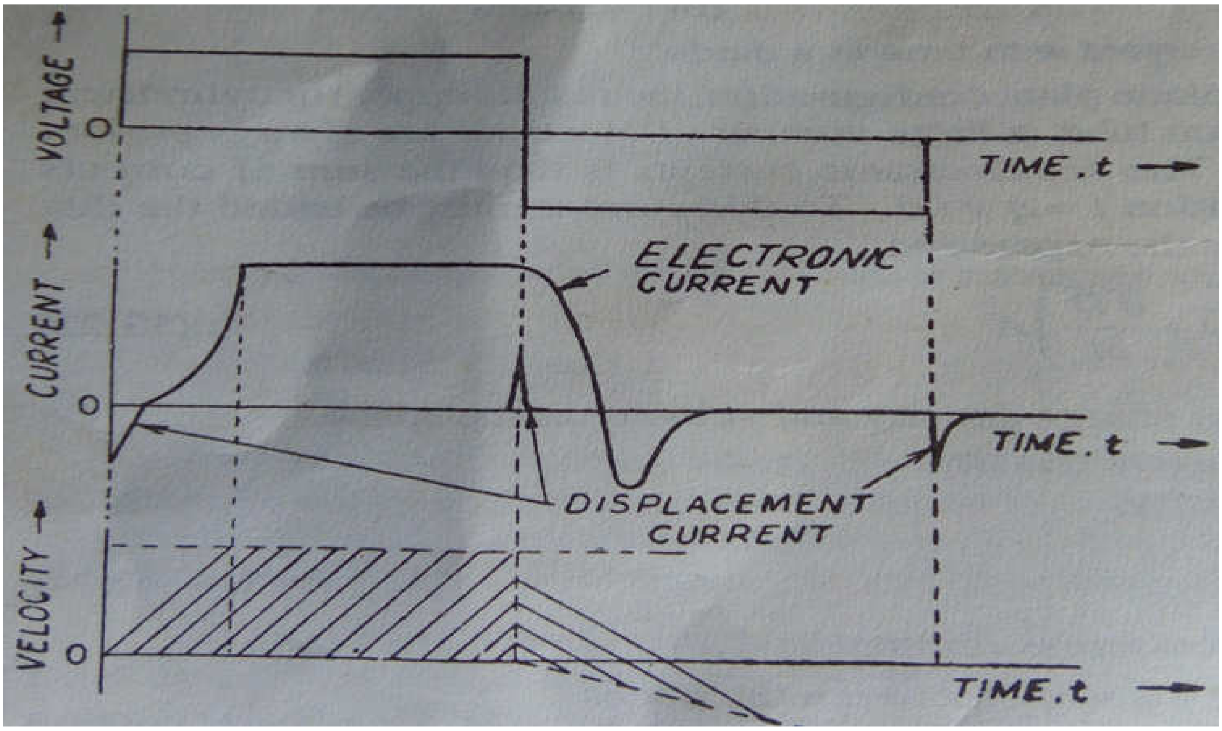


Figure 3.8. Current waveform in diode with square wave voltage

To the electronic current added the displacement current in the case of square wave voltage applied between anode and cathode, the displacement current takes the form of sharp impulses at the instants of voltage reversal.The negative electronic current which flows during part of each cycle result from electrons which fail to reach the anode and return to the cathode with considerable velocity after a relatively long time interval.

Figure 3.9 shows the time variations of distance travelled, by individual electron and current in a diode with applied square wave voltage and with large transit time. From the figure, we see the current waveform may be different from the applied voltage waveform in a diode with transit time constituting a large portion of the periodic time.

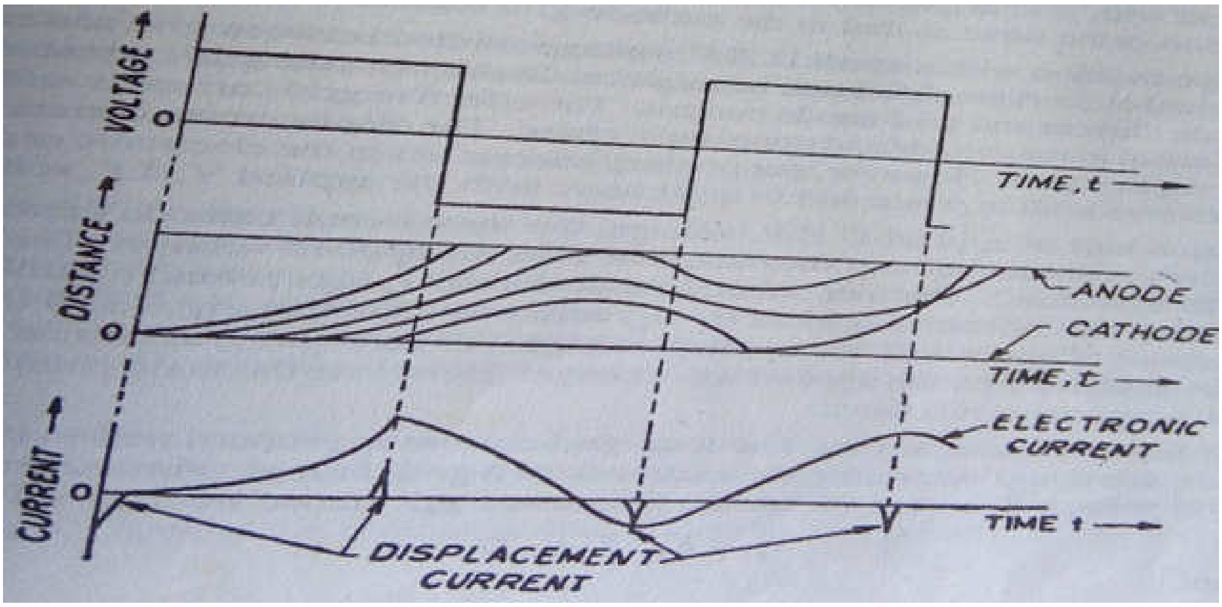
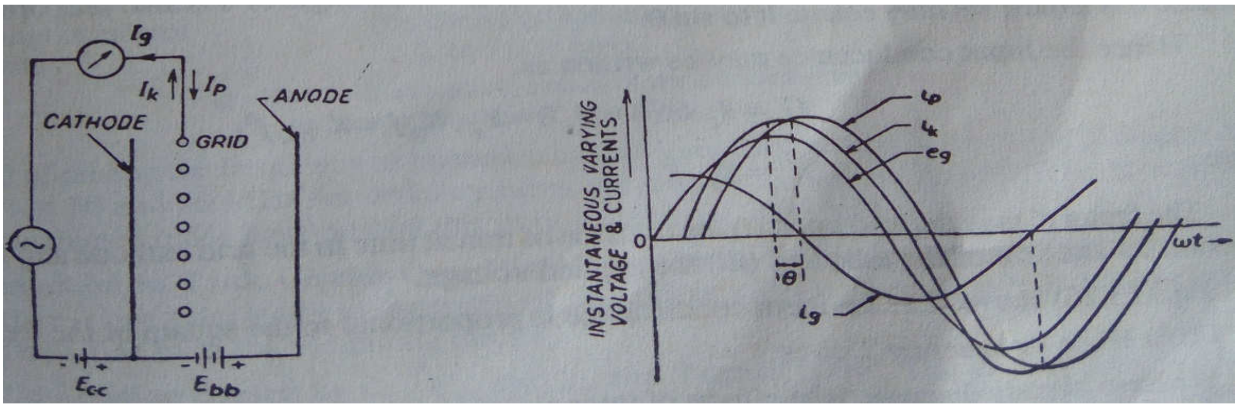


Figure 3.9. Current in diode with transit time.

Figure 3.9 clearly shows the effect of transit time on the current waveform. But this study suffers from the following three limitations (i) in actual practice voltage waveforms other than square waveform are usually applied (ii) in this study the effect of space charge is neglected (ii) no d.c bias is assumed.

3.5.3. Transit Time effect in Triodes

From the previous explanations, we have seen that the current density in the space between two parallels plate is uniform and is given by Figure 3.9. With a constant d.c applied voltage, current$I$ is constant and $\partial \overline{D}\ \partial t = o.$ the product$\overline{pv}$ is uniform over the volume. Thus, region of high density near the cathode are of low velocity. Conversely region of high velocity near the anode has low charge density. From 3.10a) we see a Triode with applied U.H.F. voltage having directions that indicate electron flow, while b) shows the waveforms of voltage and grid current I_g. In effect, Figure 3.10 sows the transit time effect in a diode.



(a) Triode with applied U.H.F. voltage (b) Waveforms of voltage

Figure 3.10. Transit time effect in a diode.

Now, we consider a plane paralleled plate triode shown in Figure3.10a, the structure may be assumed to form an effective equipotential plan the current from cathode to anode may be considered to be constituted by the following two components: (i) a current equal to $(\rho\bar{v}+ \partial\bar{D}/\partial t)$ flowing in the cathode-grid space and (ii) a current of the same from flowing in the grid-anode space. Under D.C conditions,$\partial D/\partial t$ is equal to zero and with grid negatively biased, the net current to the grid is zero. Hence the product ρv in the grid-anode region is the same as the cathode-grid region.

Next, consider the condition with a small U.H.F voltage applied between grid and cathode. The bias voltage is such that a typical class A amplification take place. In this also, the grid current may be considered as the sum of grid cathode current and grid-anode current. The effective grid current is zero only if these two current components are equal in magnitude and opposite in phase. The displacement current may be the low frequency capacitance. However, our primary interest lies in then electronic or conduction current which may have component both in phase and in quadrature with the applied V.H.F. voltage.

During the positive half of applied U.H.F voltage, the increment cathode current i_k also increases with time but reaches its maximum value somewhat later than the applied voltage. This result because of finite current i_g lags behind by still greater angle.

The grid current i_g constituting the difference current (i_k-i_g) varies as shown in Figure3.10(b). The grid current becomes more and more out of phase with the applied A.C voltage. However, the main component of current i_g is seen to have as appreciable component in phase with the applied a.c voltage. However, the main component of current i_g is such as would flow in a capacitive reactance.

3.6. Ultra High Frequency Tubes

So far, we have already discussed the effect of inter-electrode capacitance, lead inductance and transit time on the behavior of an electron tube on ultra high frequencies. These various factor limit the maximum frequency that may be guaranteed in a vacuum tube oscillator and maximum frequency that may be useful amplified. We have already discussed the Ultra-High Frequency

Responses but we shall now focus on the Electron tube oscillator. We must understand the factors of resource constraints in Section 3.7.

3.7.Frequency Limit of Ordinary Electron Tube Oscillator.

The Three factors that limit the range of frequency of electron tube oscillators include:

i. Lead inductance and inter-electrode capacitances.

ii. Reduction in tank circuit Q.

iii. Grid loading caused by $C_{gp,} L_k and$transit time.

These are discussed below.

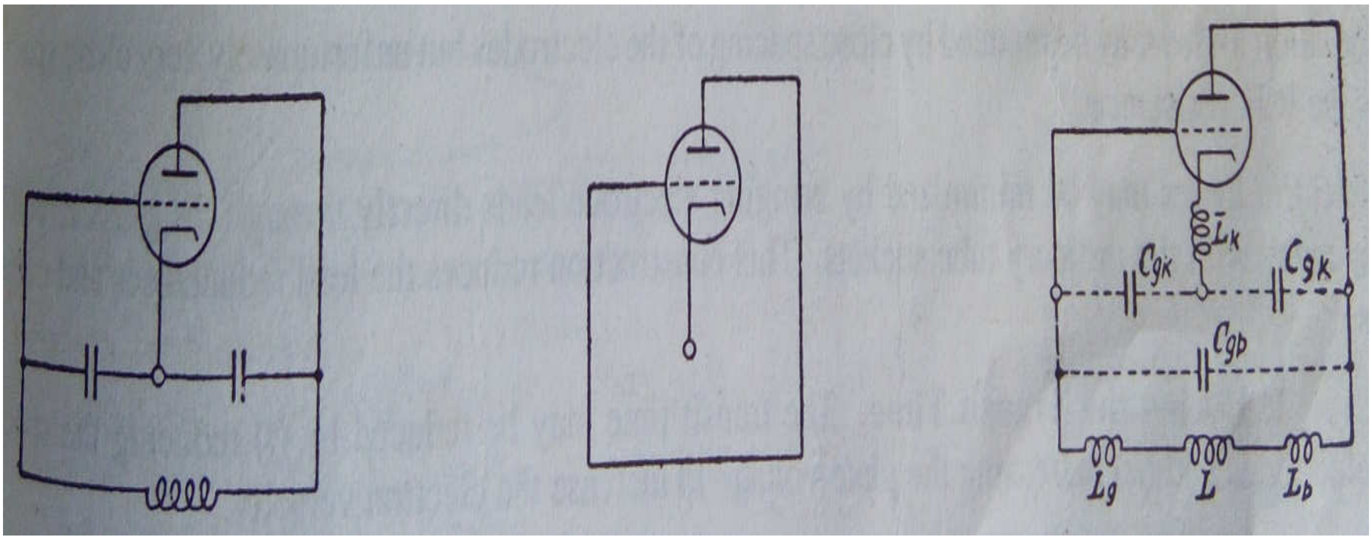
3.7.1. Lead inductance and Inter-electrode Capacitances

The frequency of oscillator is determined primarily by the constants in the tuned circuit. This frequency, neglecting the loss in the oscillating circuit is given by.

$$f = \frac{1}{2\pi\sqrt{LC}} \qquad (3.27)$$

A typical oscillator circuit is shown in Figure3.11 (a) in which only essential R.F components are indicated. In order to increase the frequency of oscillation, the magnitude of inductance and capacitance in the tuned circuit must be reduced. This reduction may be carried to the point indicated in Figure3.11 (b), where the physical lumped capacitance have been replace by apparently open circuit and the inductor by ordinary metallic conductor. The diagram hardly appears as a complete circuit.

However, when the inter-electrode capacitance and lead inductance have been inserted, the diagram so formed is shown in Figure3.11(c).In the limit, the minimum values of L and C in all oscillating circuit are the inductances that are present in the plate, grid and cathode leads and the capacitance that exit between the tube lead and electrode. Thus, the extent to which the constant of the oscillating circuit can be reduced imposes one restriction upon the upper operating frequency limit and is ordinarily determined by the geometry of the tube and associated circuit.



(a) Conventional oscillator Circuits (b) Circuit on reducing L and C to zero (c) Equivalent circuit

Figure 3.11. Equivalent circuit of an oscillator on reducing physical L and C to zero

The method of raising the frequency of operation by changing the constant in the tuned circuit produces upper limit between 10 and 60 MHz for ordinary tubes but reaches as high as 700 MHz for special tubes. Other specially constructed tubes have pushed this limit up to about 3000 MHz.

3.7.2. Reduction in Tank Circuit Q at High Frequency

This lowering of circuit Q with the increase of frequency is due principally to increase (i) skin effect, (ii) dielectric losses in the tube base and envelope and (iii) radiation from the oscillating circuit.

3.7.3. Grid Loading Caused by C C_{gk} and L_k and Transit time.

The grid-to-plate capacitance C_{gp}may cause the input conductance of the tube to be undesirable large. Furthermore this conductance may be either positive or negative, depending principally upon the impendence in the plate circuit. The cathode lead inductance L_k in the conjunction with C_{gk} has the effect of increasing the input and output conductance of the tube. The input conductance is generally more important and is given by,

$$G_i = L_k C_{gk\ \omega^2} \tag{3.28}$$

This relationship is very nearly true for pentode but is quite approximate for triodes. The transmit time cause an input conductance which increases linearly with the square of the frequency as given by the relation,

$$G_t = K_{gm}\omega^2 t \tag{3.29}$$

Where t is the transit time.

Equ. (3.29) applies for small values of transit time.

At extremely high frequencies, where the electron transit time reaches a half or even a quarter of a period, this law no longer applies.

As the input conductance increases and greater power is required to be fed to the grid. This grid loading reduces the oscillator power output to impractically small values at frequencies in the neighborhood of 40 or 50 MHz assuming operation with conventional tubes.

3.8.Remedies for the Limitations

We shall now discuss the identified remedies for the identified challenges below

i. Minimizing Tube Inter-electrode Capacitances and Lead Inductances.

The lead inductances and capacitance may be reduced by shorter length of wire for the leads. This also reduces the circuit resistance. The inter-electrode capacitance may be reduced by using smaller electrodes. The capacitance may also be increased by increasing the spacing between the elements, by unfortunately, this increase the transit time.

ii. Preventing the fall of Tank Circuit Q

The losses due to skin effect may be reduced by using low resistance large-surface conductance.Radiation losses may be reduce by close spacing of the electrodes but unfortunately very close resistance increase the R.F resistance.

Dielectric losses may be minimized by bring electrode leads directly through the glass envelop, thus doing away with the ordinary sockets. This construction reduces the lead inductance and capacitance also.

iii. Reducing the Transit Time

The transit time may be reduce by (i) reducing the spacing between electrodes and (ii) increase the plate voltage to increase the electrode velocity.There are however, limit for these modifications. Thus the decrease in electrode spacing makes the tube manufacture more difficult and in other to keep the inter-electrode capacitance from becoming excessively large, the decrease in spacing must be accomplished by a decrease in electrode surface area. This reduces the allowable plate dissipation. The increase in plate voltage is limited by the allowable plate dissipation. As the plate voltage increases the plate current increase resulting in increase plate dissipation. For prescribed maximum plate dissipation as the plate voltage is increases the corresponding plate current must be reduced by circuit arrangement. The transconductance reduces therefore.In pulsed operation, for use in Radar for instance, transit time generally reduced by using very high plate voltage during the pulse. This is possible without plate current reduction because the plate dissipation during the pulse is permitted to become thousands of times greater than the average allowable dissipation due to the inherent thermal of the plate structure. Thus tubes may run at higher frequencies when pulse then when operated C.W

For pulsed operation, plate voltages of the order of ten thousand volts are used in order to obtain high peak power and to reduce transit time. These voltages exert considerable influence upon the mechanical features of the tube designs. The glass seals between electrodes are separated as much as possible in other to reduce the voltage gradient in the glass and on its surface. Further sharp coroners on high voltage conductors are avoided in order to reduce the possibility of coroner.

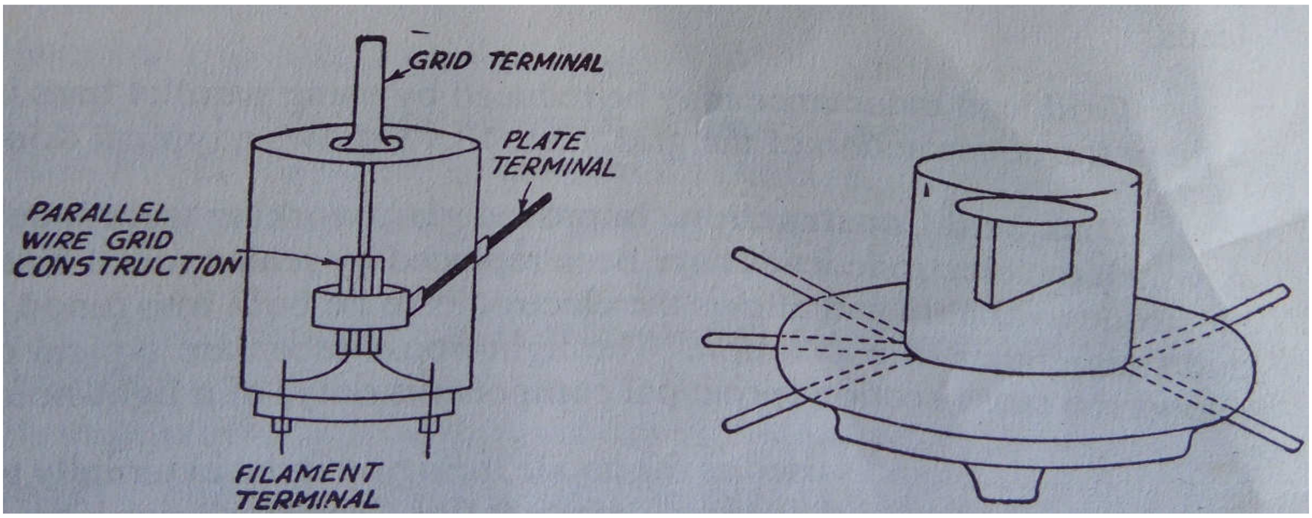
3.9. Typical U.H.F Tube Construction

We have clearly highlighted the various factors of limitation above. Now, because of the factors discussed above, special tube has been developed for use at ultra high frequencies. These construction methods may be put into the following three categorieswhich we shall be discussing here namely:

i. Single ended construction
ii. Double ended construction
iii. Disk seal construction.

3.9.1. Single Ended U.H.F. Tubes

These are the conventional tubes inwhich care has been taken to reduceleads inductions, inter-electrodecapacitance and transit time by reducing the size and spacingof the electrodes, by suitableglass-metal seals. Figure3.12 (a) Show typical single ended tube. Figure3.12(b) show an acorn tube which is also asingle ended tube but the electrodeterminal are brought out of envelop around the periphery of the tube envelop. Grid and plate terminals are brought out on side of the periphery while the filament terminals are brought out on the other side. Figure 3.14 shows a cross section of lighthouse tube.



(a) Typical single ended U.H.F. tube (b) Acorn tube

Figure3.12. Typical single ended U.H.F. tubes.

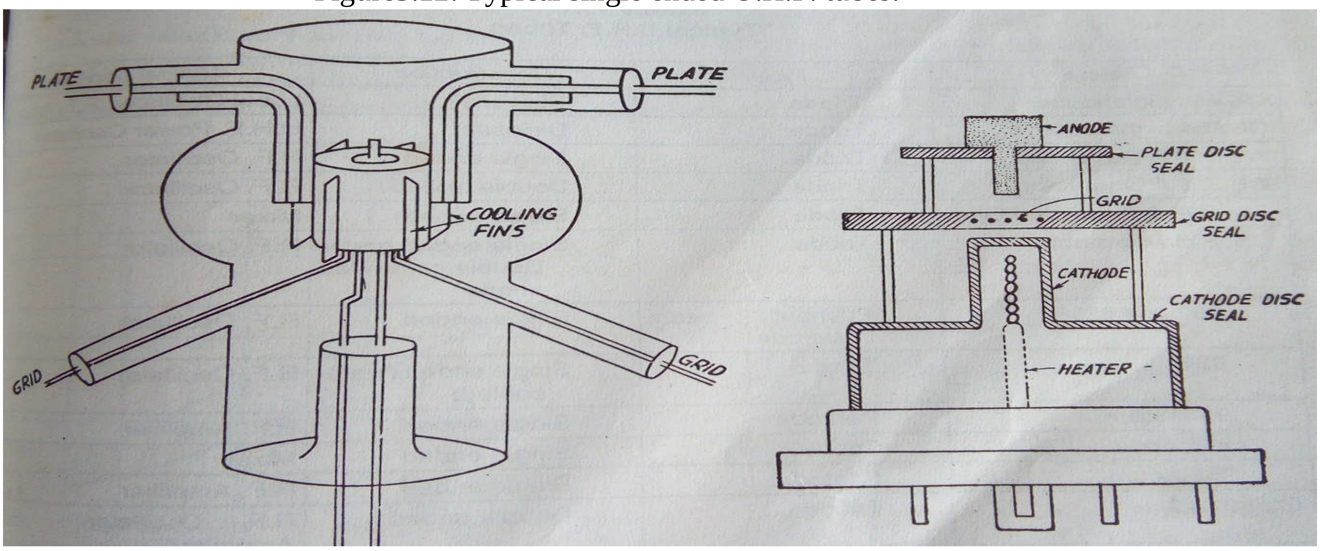


Figure3.13. Double ended U.H.F. tube.

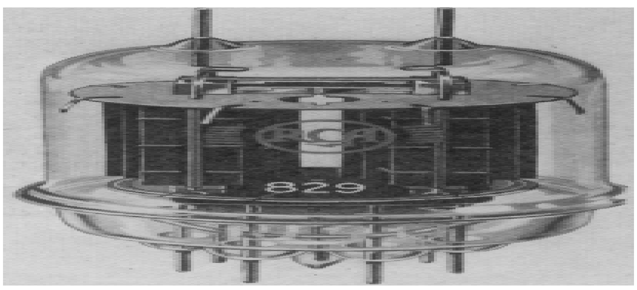


Figure 3.14. Cross section of lighthouse tube.

3.9.2. Double Ended U.H.F Tube

This is another form of UHF tube construction method. In such tube, each electrode has two separate terminal brought out individual through glass envelope. These terminals are then

connected symmetrically to the external circuit so that the lead inductances are effectively paralleled. Another technique is to arrange the lead inside the tube such that they are extensions of external resonant transmission lines. The tubes elements are then maintained at the centre of a resonant half-wave transmission lines system with shorted ends. The lumped capacitance between the electrodes is the assumed to be equally divided between the two quarter wave sections on either side of the tube elements. Such tubes will oscillate at higher frequencies because the capacitance associated with a single quarter wave construction also reduces the radiation losses and the I^2R losses in the leads.
Grid lead inductance may be reduced by using parallel bars instead of more common helically wound wire in the construction of the grid. Figure3.13 shows a typical double ended U.H.F triode.

3.9.3. Disk Seal Construction

In this case, we need to know that improved glass working techniques have permitted the construction of tubes in which the electrode leads have been replaced by sealed metal discs.

This makes the whole perimeter of the electrode available and allows the electrodes to be built into tuned circuits directly, thus yielding higher Q's and frequencies of oscillation. The light-house tubes are typical of this disk seal construction.

Figure3.14 shows the cross-section (principal components only) of a light-house triode.A solid block is used as the anode, bringing it out eternally to the tube proper by disk seal construction and then provides it with thermal radiating fins. The small working surface area of the anode near the grid and cathode reduces the tube capacitances but the large overall are plus the eternal exposure to air increases the heat dissipating ability of tube.Extremely close spacing of the electrodes can be achieved by the parallel plane construction. The disk seal tubes of this type have been built with grid cathode spacings as small as 1mil. Table 3.1 gives the list of some of the representative UHF tubes. It also gives for each tube the type of construction and typical services.

Table 3.1: Typical U.H.F Tubes

Name	**Type**	**Construction**	**Typical Service**
GL 446 Lighthouse	Triode	Disk seal	Local Oscillator, R.F. Amp.
GL 646 Lighthouse	Triode	Disk seal	U.H.F. Power Oscillator
WE 316 A Door Knob	Triode	Single ended	R.F. Oscillator
WE 318 A Door Knob	Triode	Double ended	R.F. Oscillator
708 A Grounded Grid	Triode	Single ended	Mixer
811 Micro-pup	Triode	Single ended grid + Double ended anode	R.F. Oscillator
832 and 8298	Pushpull beam tetrode	Single ended	R.F. Oscillator
888	Triode	Single ended (water cooled)	R.F. Oscillator
954 Acorn	Pentode	Single ended	R.F. Amplifier

955 Acorn	Triode	Single ended	Mixer
956 Acorn	Pentode	Single ended	R.F. Amplifier
GL 8012 A	Triode	Double ended	R.F. Oscillator and RF Amplifier

3.9.4. Elimination of the Effects of Inter-electrode Capacitances

We have explained the disk seal construction above. It must be made clear that disk seal value is very suitable for use with concentric line circuits which may be arranged one inside the other. In this way, the resonant circuits can be placed between cathode and grid and between grid and anode, care being taken to provide the necessary d.c insulation. This procedure eliminates the difficulty of lead inductances but it leaves the problem of tube capacitances in a different form. In a Disk Seal Tube, the effects of Inter-electrode Capacitances must be eliminated. Figure 3.15 illustrates the effect of a lump capacitance at the open end of a line in increasing the resonant wavelength.

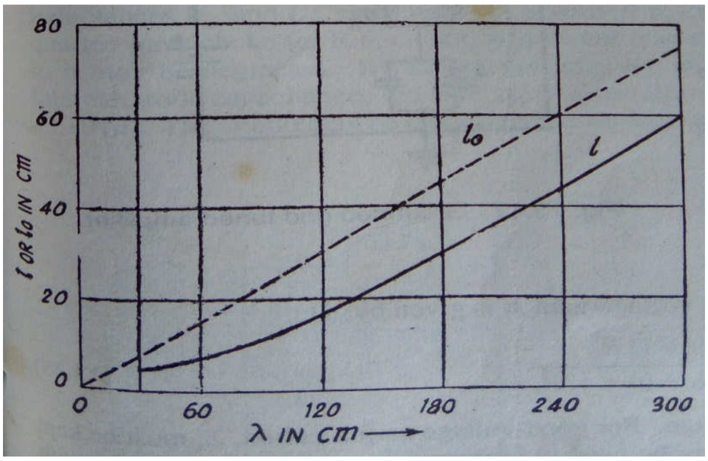


Figure 3.15. Effect of a lump capacitance at the open end of a line in increasing the resonant wavelength.

If the tube forms a shunt impedance $1/j\omega C$ at the end of a line of characteristic impedance Z_0, the input impedance Z_i of a closed length l is given by,

$$\frac{1}{Z_i} = j\omega C + \frac{1}{Z_0 \tanh \gamma l} \quad (3.30)$$

Where $\omega = 2\pi f = \frac{6\ X\ 10^{10}\pi}{\lambda}$

Let the losses in the line be negligible then, $\gamma = j\beta$ and ,

$$\tanh \gamma l = j \tan \beta l = j \tan(\frac{2\pi l}{\lambda}) \quad (3.31)$$

Hence, $$\frac{1}{Z_i} = j\left[\omega C - \frac{1}{Z_0 \tan{(2l\,/\,\lambda})}\right] \quad (3.32)$$

The line becomes resonant if Z_i is infinite, i.e. when

$$\tan\left(\frac{2\pi l}{\lambda}\right) = \frac{1}{Z_0 \omega C} \quad (3.33)$$

$$= \frac{\lambda}{6\pi \times 10^{10} \times Z_0 C} \quad (3.34)$$

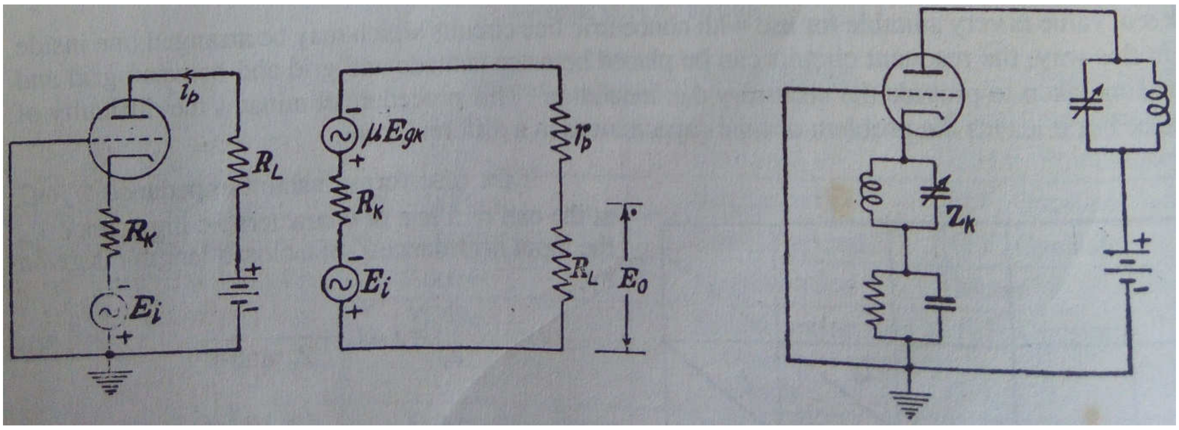
Figure 3.15 exhibits the mode of variation l with λ assuming Z_0 = 70Ω and C = 5 μμ F. This value of C is the inter-electrode capacitance of the value when placed in the circuit and is typically 5μ μF. The dotted line in Figure 3.15 shows l_0 =λ/4. The difference between l and l_0 gives the shortening caused due to C. it is clear that for the values mentioned a concentric line circuit can be used at wavelength less than 30 cm only in a resonating mode higher than quarter wave.

British and American disk seal valves use metal-to-glass seals while German valves employ a ceramic in place of a glass in order to obtain greater accuracy in electrode spacing. The ceramic used has a lower loss factor than the glass but may not have any significance except at the highest frequencies.

In large UHF triodes, it has become a standard practice to adjust the cathode relative to the grid after manufacture. This gives greater uniformity and improved performance.

3.9.5. Grounded Grid Operation of Disk Seal Triodes

Basically, the UHF triodes whether used as signal amplifiers or as oscillators are generally operated in some form of grounded grid or common-grid. Disk seal triodes are almost always used as grounded-grid circuit. Figure 3.16 (a) shows a simple grounded-grid amplifier, R_l is the load resistance and R_k is the cathode resistance. Figure 3.16 (b) shows the incremental equivalent circuit.



(a) Circuit (b) Incremental equivalent circuit (c) Grounded grid tuned amplifier

Figure3.16. Grounded-grid triode amplifier.

From the equivalent circuit of Figure3.16 (b), the voltage gain A is given by

$$A = \frac{E_0}{E_i} = \frac{(\mu+1)R_L}{r_p + R_L + (\mu+1)R_k} \tag{3.35}$$

Figure3.16c shows a typical R.F. grounded grid stage. For good voltage amplification, $Z_k$must be kept low. Coil and capacitor, forming the tuned circuit, may be used as frequencies as high as 400 MHz, with consequent ease of tuning and small bulk. At frequencies above 400 MHz, it becomes impossible to use lumped circuit elements even at the cathode. It then becomes necessary to use transmission lines in both cathode at anode circuits. The input line should then be of low characteristic impedance while the output line should be of high characteristic impedance.

3.10. U.H.F Plate Cooling

Heating effect is another concern in UHF systems. High power operation of UHF tubes is impeded by the reduction in electrode area necessary for small inter-electrode capacitances. The small electrode area and the large anode voltage necessary for the reduction of transmit time makes the problem of heat dissipation from the anode quite difficult. The heat may be removed more effectively by following means:

i. By adding cooling fins to the anode structure. This increases the thermal radiating surface without changing the tube capacitance.
ii. By allowing an blast over the anode.
iii. By water cooling of anode used for high power operation, and
iv. By using anode made of tantalum and by operating the anode red hot. This allows higher plate dissipation due to more rapid radiation of heat through operation at a higher temperature.

3.11. U.H. F Filament Emission.

Now, itgood we understand that very large filament emissions are generally needed in UHF tubes for the following two reasons:

(i) The large R.F displacement currents in the plate-to-cathode capacitances are, to a very large extent, supplied by the cathode.
(ii) In pulsed operation, extremely large conduction currents are needed over short duration of time.

Both these currents require actual cathode emission. Hence, large filaments operating at high temperatures are needed, adding to the problem of tube cooling. Oxide coated cathodes are occasionally used in lower power tubes. However, thoriated tungsten is most commonly used since it is less susceptible to positive ion damage and can safely provide the require degree of emission.The use of thorium on the filament however, results in grid contamination due to deposit on evaporation from filament. As the grid itself is generally quite hot, grid emission may result causing a D.C current to flow in the grid circuit.

3.12. Oscillators in UHF Domain

We have established various interesting perspectives in the construction of UHF tubes. We shall now look at the frequency oscillators that stimulate its processes below.

3.12.1.Single Tube UHF Oscillators

Normally, at ultra high frequencies, two significant features of oscillator need be noted: (i) because of the presence of inevitable inter-electrode capacitances. For instance, Colpitts oscillator circuit are more common than the Hartley circuits, (ii) difficulty of obtaining lumped inductances and capacitances of required small values and desired power handling capacity, necessitates the use of resonant transmission lines.

Accordingly, the physical appearance of a UHF oscillator may not seem to resemble either the Hartley or the Colpitts circuit, and its true nature gets revealed only when the circuit is reduced to its functional form.

3.12.2. Lighthouse Tube Oscillator

Recall in Unit 1, we mentioned superheterodyne receivers and its roles in communication systems. In context, the lighthouse tube oscillator may be used as a local oscillator in UHF superheterodyne receivers. Figure 3.18 gives the circuit of lighthouse tube oscillator. The R.F. equivalent circuit may be developed in the following manner. The short circuited coaxial transmission lines between cathode and grid and between grid and anode may be represented respectively by inductances L_{gk} and L_{gp} since they are assumed to be highly less than three fourths of a wavelength.

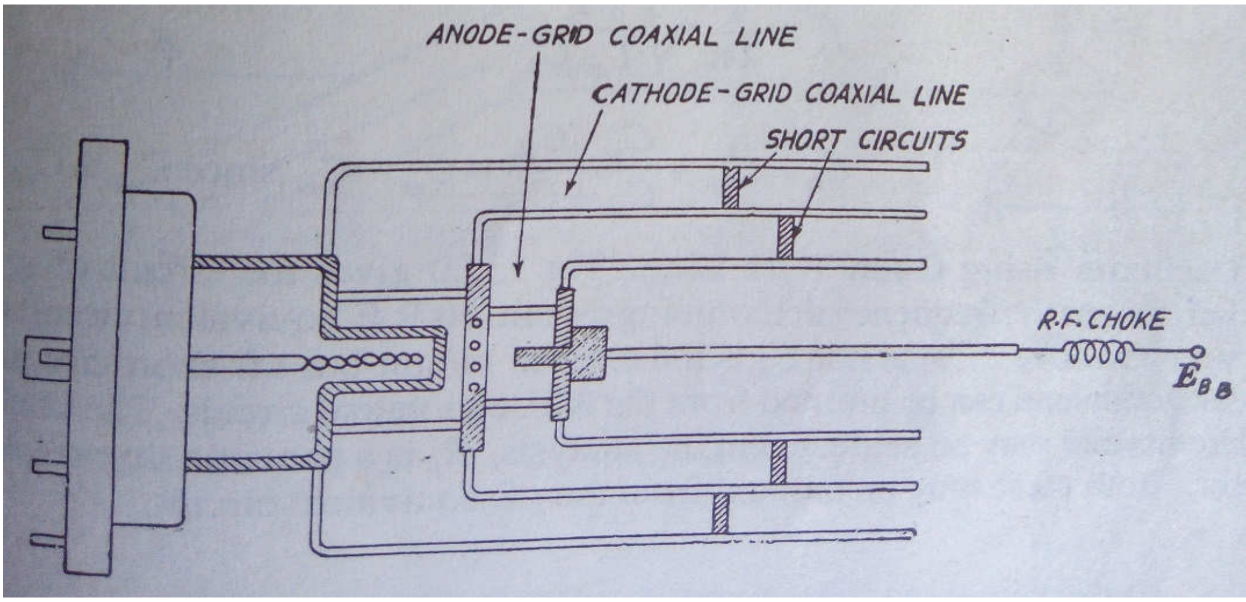
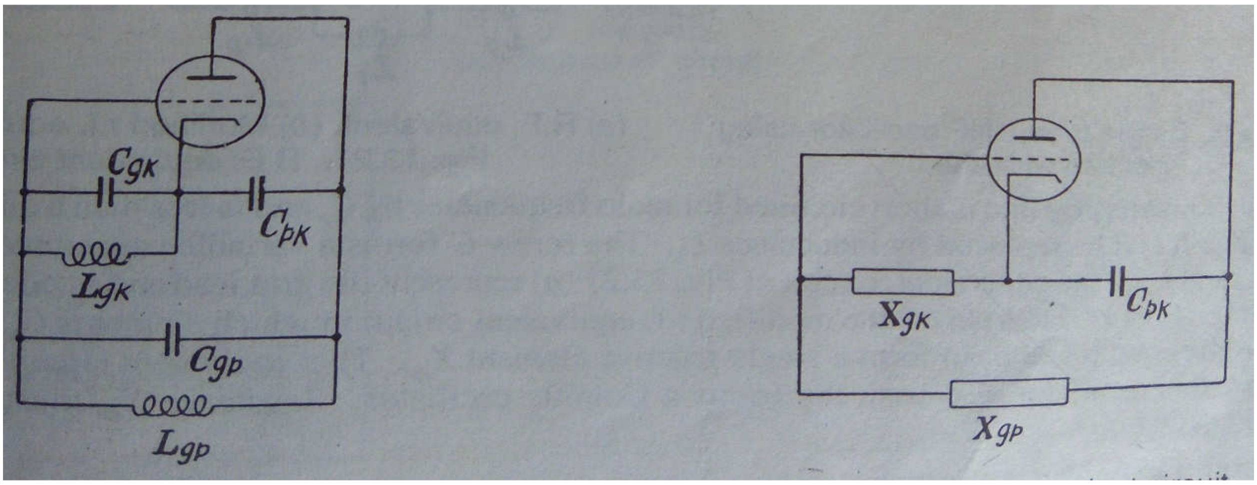


Figure3.18. Lighthouse tube oscillator

The quarter wave choke (or R.F. choke, whichever is used) is an R.F. open circuit so that all elements connected to it may be neglected. Replacing the lines by inductances as mentioned above, and including the tube inter-electrode capacitances the R.F. equivalent circuit of the light house oscillator appears as shown in Figure3.19 (a). The circuit may be rearranged by replacing the cathode circuit elements (L_{gk}, C_{gk}) and grid-to-anodecircuit element ($L_{gp,}C_{gp}$) by reactance X_{gk} and X_{gp} respectively.



(a) Equivalent circuit (b) Modified equivalent circuit

Figure3.19. Equivalent circuit of lighthouse tube oscillator.

Please, you must know that for sustained oscillation, the reactance on either side of the tap must be of the same type. Thus, the grid-to-cathode reactance X_{gk} must be capacitive. Further grid to anode reactance X_{gp} must be of opposite nature. i.e, it must be inductive in this case. The resulting final R.F. equivalent is shown in Figure3.19 (b). This obviously is a colpitts oscillator. The tuning of grid-to-anode coaxial line, $i.e.$ position of the short circuiting plunger in this line, determined the magnitude of the inductance L_{gp} which forms the principal factor controlling the tuning of the oscillator to the desired frequency.The grid-cathode line determine the amount of feedback by establishing the values of $X_{gk.}$ It has only a negligible effect upon the frequency of oscillation which is given by,

$$f_o = \frac{1}{2\pi}\sqrt{\frac{1}{L_{gp}\,C_o}} \tag{3.36}$$

Where,$C_o = C_{gp} + \frac{C_{gk}.C_{pk}}{C_{gk}+C_{pk}} \approx C_{gp} + C_{pk}\ since\ C_{gk} \gg C_{pk.}$

3.12.3. Single Tube Oscillator using Open Wire Line

Figure3.20 gives the circuit of a single-tube UHF oscillator using wire line as the frequency determining circuit. Its R.F. equivalent circuit may be developed in the same general way as above. Capacitor $C_g$is a d.c open circuit but R.F short circuit. It provides the proper bias by grid leak action and can be omitted from R.F equivalent circuit. The chokes in the cathode leads are R.F open circuit and may be neglected in the analysis. $R_p$is a parasitic suppression resistor and R_g is the grid lead resistor. Both these may be omitted from the r.f equivalent circuit.

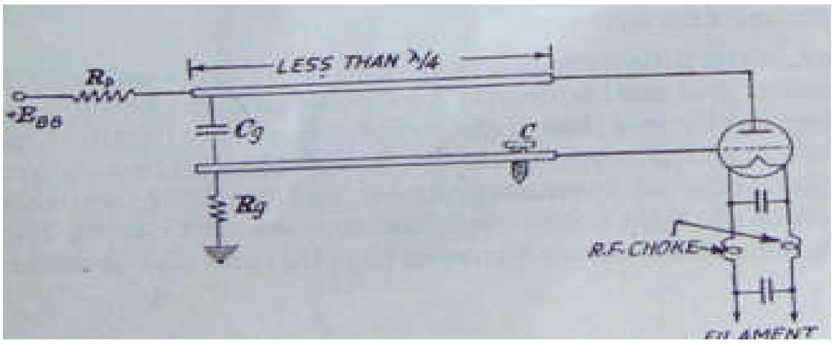


Figure3.20. Single tube UHF oscillator using (a) R.F equivalent (b) Modified R.F Parallel wire line equivalent circuit

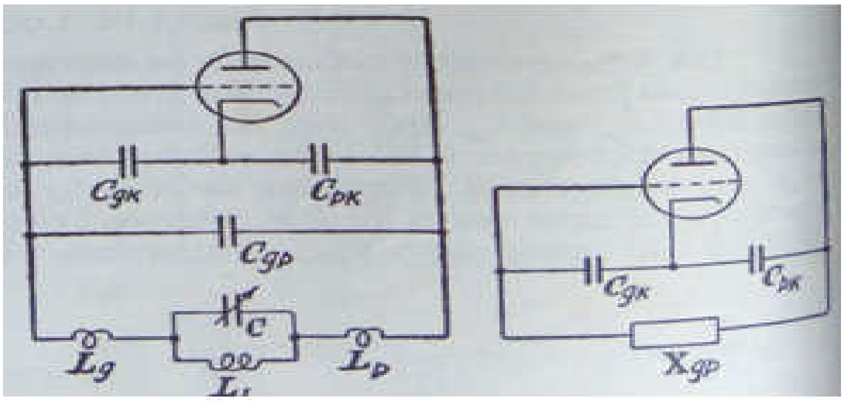


Figure3.21. (a) R.F equivalent (b) Modified R.F Parallel wire.

The transmission line is short circuited for radio frequencies by $C_g$and is less than a quarter wavelengths long. Hence, it can be replace by inductance $L_{1.}$
The screw of Figure3.21 (a) represent the grid lead and anode lead inductances respectively. Figure3.21 (b) gives the modified R.F equivalent circuit in which elements $C_{gp,}L_{g,}L_{p,}C\ and\ L_t$ have been grouped together to form a single reactive element $X_{gp.}$ This equivalent circuit of Figure3.21 (b) show that this oscillator also basically forms a colpitts oscillator. Reactance X_{gp}must be inductive for sustained oscillations.

3.12.4. UHF Pushpull Oscillators

UHF pushpull oscillate circuits have the same general form as at low frequencies except that resonant transmission lines used in place of ordinary round circuits. Furthermore, the lead inductance and inter-electrode capacitance must be included in the analysis as shown in Figure 3.22. Thus typical UHF pushpull oscillators are: (i)tuned-grid tuned-plate oscillator (TGTP), (ii) tuned-grid tuned-cathode (TGTK) and (iii) tuned-plate tuned-grid tuned cathode (TPTGTK). Each of these oscillator circuits may be reduced to simple Hartley or Colpitts oscillator.

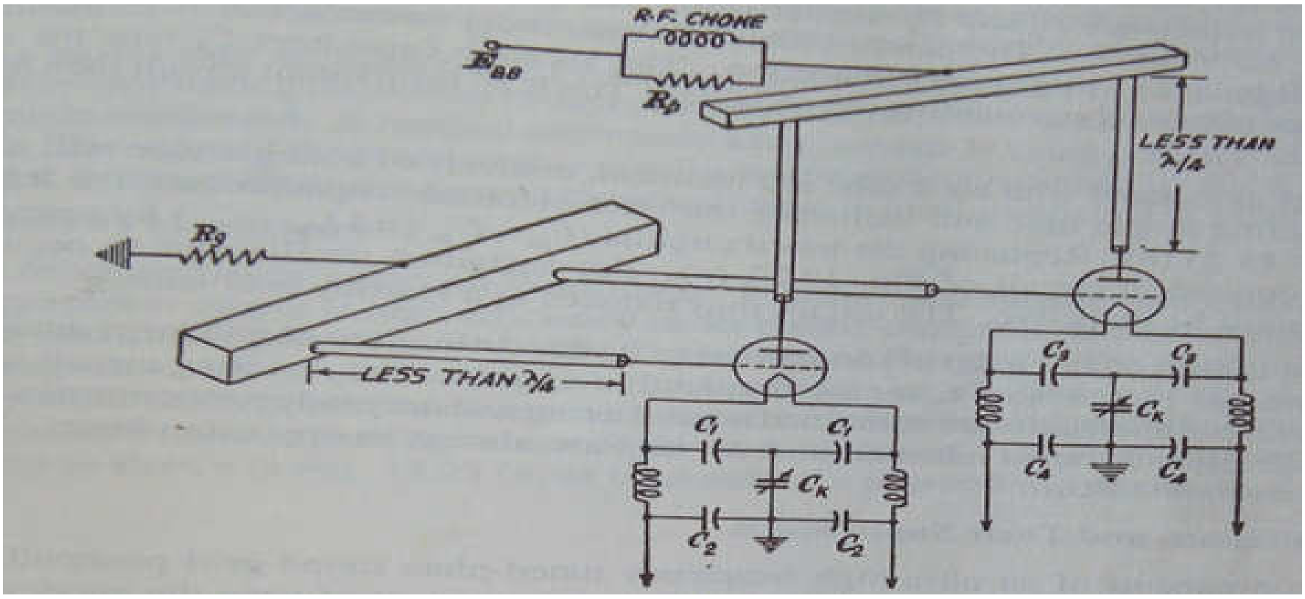
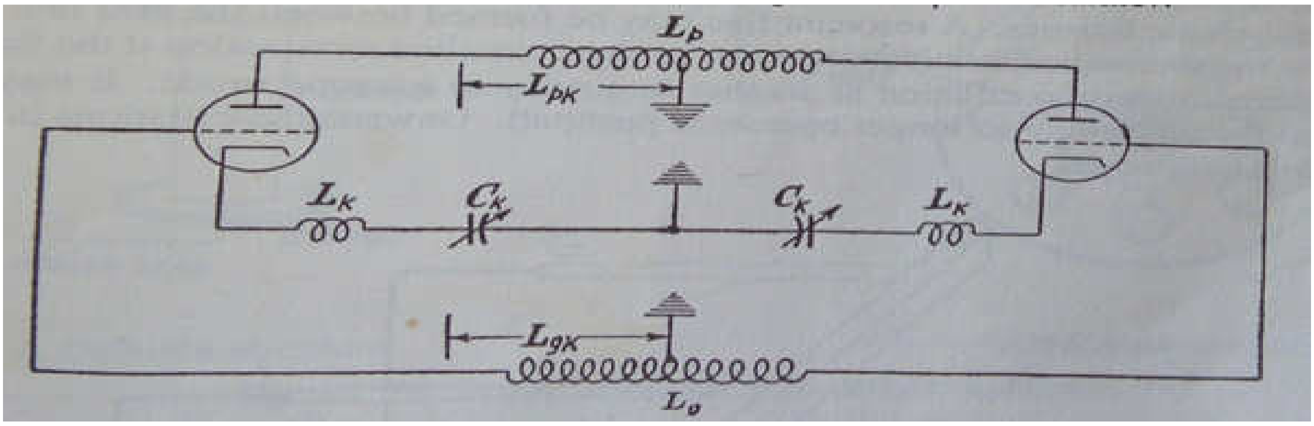
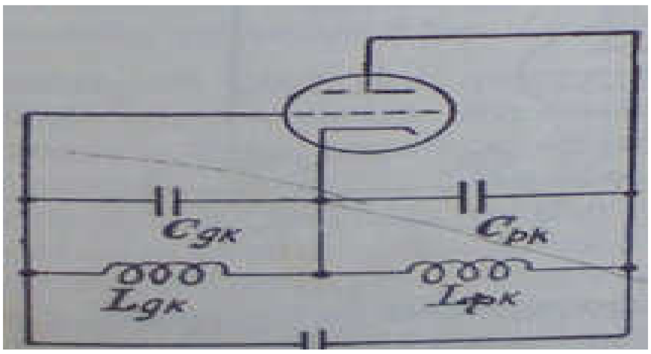
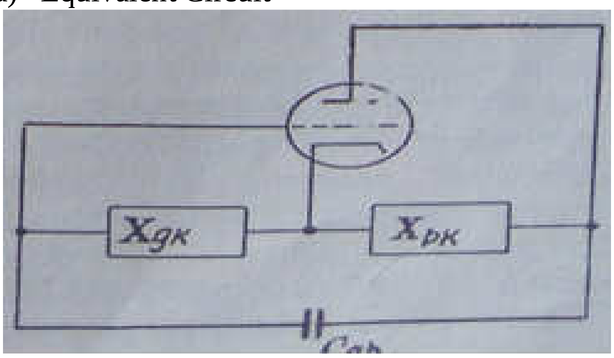


Figure 3.22. Tuned-plate Tuned-gird Pushpull Oscillator



a) Equivalent Circuit





(b) Modified equivalent circuit (c) Final equivalent circuit

Figure3.23. Equivalent circuit of TPTG pushpull oscillator.

Figure3.22 show the circuit of pushpull tuned grid oscillator. The bypass capacitors $C_{1,}C_2C_3$and C_4 appear as short circuit. The grid leak resistor R_g and parasitic suppressor R_p may both be considered as open circuits. The plat and grid lines are both less than a quarter wavelength and circuit so that they may be represented by effective inductance L_p and $L_g$respectively. The plate and grid inductance have been lumped with L_p and L_g respectively. The center point of this inductance is R.F ground potentials because of the pushpull balance-to ground connection. The variable capacitors C_k tune the cathode lead inductance to resonance grounding the cathodes. The R.F Equivalent then appears are as shown in Figure 3.23 (a). Since the circuits associated with each tube are identical, analysis of a single tube suffice for the whole circuit.

Considering with each are identical, analysis of a single tube will suffice for the whole circuit is shown in Figure 3.13(b). Replacing the tuned circuit $(L_{gk} - C_{gk}\, and\, L_{pk} - C_{pk})$ by reactance $X_{gk}\, and\, X_{pk}$ result in the final equivalent circuit of Figure3.23 (c) For sustained oscillation to occur, both these reactance $X_{gk}\, and\, X_{pk}$ must be inductive. The circuit thus reduces to a Hartley oscillator.

The small size of the electrode in UHF triodes it very difficult to obtain amounts of power even from pushpull circuits. Hence, more tubes are frequently required. The resulting circuit is called a ring oscillator because several pushpull units are combined around a ring and made to operate as a single circuit. The tubes are so arranged that adjacent tubes around the ring are always in apposite phases. Thus, any two adjacent tube form pushpull circuit.

3.13. Parasitic Oscillation and Their Suppression

Parasite oscillation refers to unwanted oscillation generated wrongly. Now, the essential components of an ultra tuned-grid pushpull are shown in Figure3.24. Since the tube operate in pushpull, at some instant of time the anodes and the grid must have the relative polarities indicated in Figure 3.24. There are standing waves on the grid and plate lines due to their resonant characteristics. A resonant line may be formed between the grid line and plate line of each tube due to the open-circuited termination or due to short circuited termination if the lines are by-passed to ground. This tends to cause oscillation at another frequency in a wrong mode. It may set up in-phase standing waves and the tubes may no longer operate in pushpull. This constitutes parasitic oscillations.

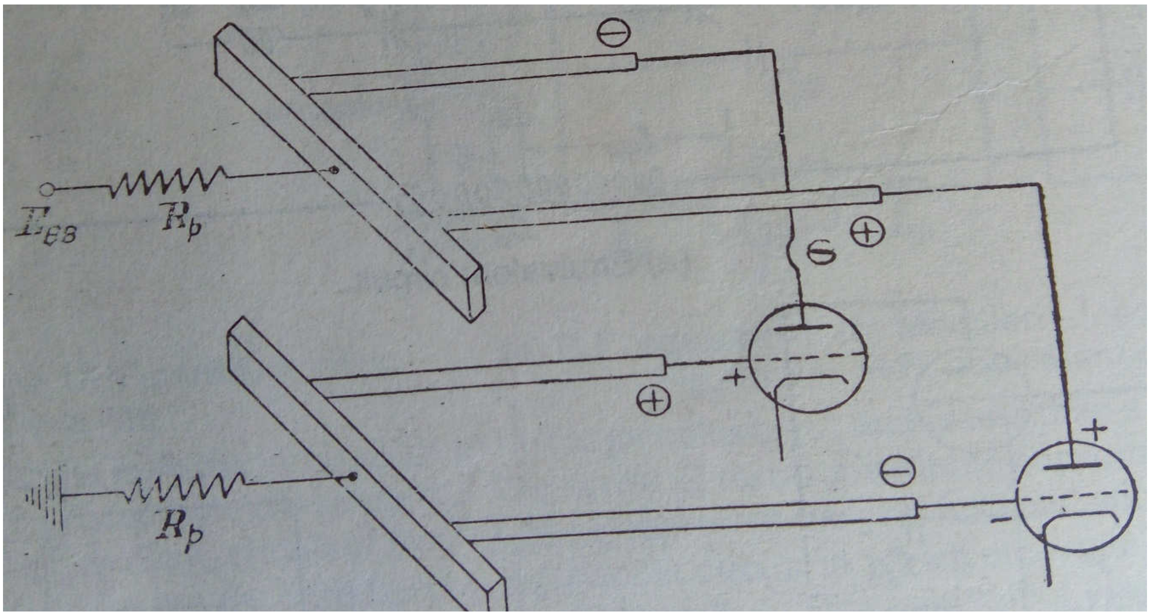


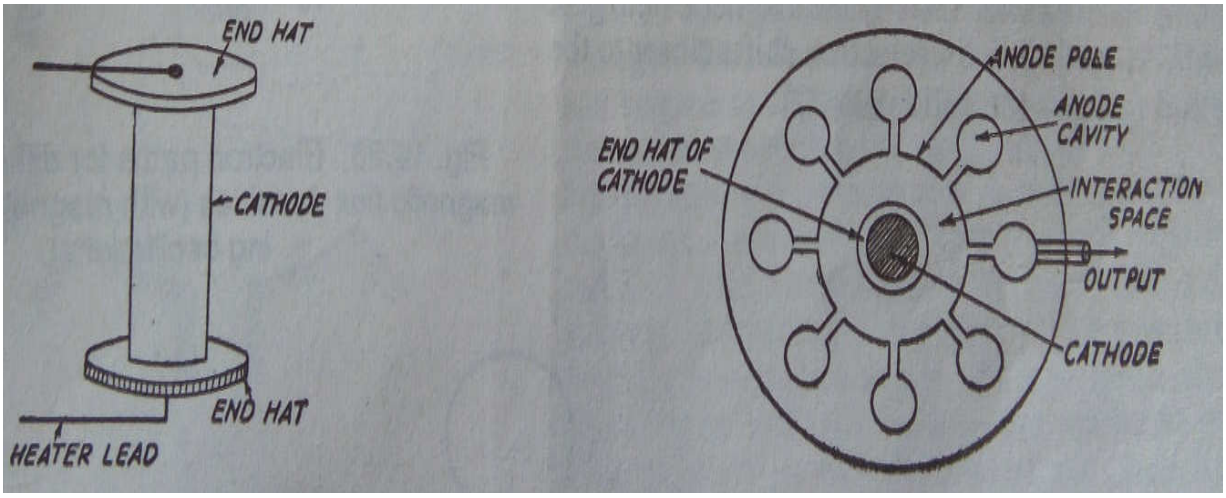
Figure3.24. TPTG pushpull susceptible to parasitic oscillations

Parasitic oscillations occur in several ways. The method given is typical and common and consist in formation of resonant line between conductors are rarely of any use since the image formed may cause parasitic to occur between the conductor and its image. Parasitic oscillations at frequencies other than of the main mode need be prevented since (i) They lower the efficiency of the system and (ii) may cause radiations at unauthorized frequencies. Further these parasitic oscillations do not necessarily respond to the tuning adjustment of the desired mode.

In terms of its suppression, wrong mode may be suppressed by keeping the parasitic lines from being resonant. This can be accomplished by terminating these lines in their characteristic impendence in such a way as to leave the operation in the main mode unaffected. A method commonly used consists in using parasitic suppressing resistors ($R_p$in Figure. 3.24). The value of these resistors is usually determined experimentally and it ordinarily ranges from 10 to 100 ohms in transmitters and several thousand ohms in local oscillators of super heterodyne receivers. The other types of oscillator of UHF are discussed below.

3.14. Multicavity Magnetron

The magnetron is a high-powered vacuum tube that works as self-excited microwave oscillator. Crossed electron and magnetic fields are used in the magnetron to produce the high-power output required in radar equipment.It was developed during Second World War and it is the most suitable microwave generator capable of delivering large power output at high efficiency under pulsed condition. It was developed principally for use in high pore microwave pulsed radar (Radio detection and Ranging). The principal elements of a typical multicalvity magnetron are shown in Figure3.25. It consist of a cylindrical heated oxide coated cathode surrounded by a laminated copper anode structure that possesses several identical cavities (usually 8 in number) opening into cathode-anode space by means of slots as shown. The output power is drawn by means of a coup long-loop as shown in Figure 3.25 (b) or by means of a trapered waveguide. The magnetron requires an external magnetic field lines parallel to the axis of the cathode. This field is usually proved by permanent magnet.



(a) Cathode structure (b) Cross-section

Figure3.25. Construction of multicavity magnetron.

Operationally, a high D.C potential is applied at the anode with respect to the cathode, thereby setting up a strong radial electric field. An axial magnetic field is created usually by permanent magnet or sometimes by an electromagnet. Electrons on emission from the cathode, experience force due to both the radial electric field and axial magnetic field. The electric field urges the electron radically outward while the magnetic fields exercise a force perpendicular to the instantaneous directly of motion of the electron. As a result of these combine force, the electron follows a spiral path of curvature of which decreases with the increase of the magnetic field strength. Accordingly, for a given value of anode D.C poetical there is a critical value of magnetic flux density for which the electrons just graze the anode.

For magnetic flux densities exceeding this critical value B, the electrons miss the anode and spiral back to cathode as shown in Figure3.26. Magnetic flux densities of the order of one or two times B are normally needed for oscillations.

With zero magnetic flux density ($B=O$), an electron emitted the cathode directly towards the anode. With weak magnetic field $(B < B)$, the electron path gets deflected as shown by path P_2 in Figure3.26.

However, with magnetic flux density exceeding the critical value B_c the electron is turned back towards the cathode without ever reaching the anode, as by path $P_{4.}$ with $B = B_{c,}$ the electron just grazes the anode surface as shown by path $P_{3.}$ With $B > B_{c,}$ in the absence of oscillation, all emitted electron return to cathode and current is zero.

Next let us assume that oscillationexist in the resonant structure. Then it will be shown that with $B < B_{c,}$ there results an interaction between the electrons and theelectric field which under favorablecircumstances causes the oscillationto receive energy from the electronsin the interaction space. Let usconsider the electrons in the interactionspace. Let us consider the π −modeof oscillations, i.e. oscillations withphase difference of π radians between adjacent poles. These oscillations, in turn produce radio frequency fringing field extending into the interaction space as illustrated in Figure3.27.

In the absence of this R.F field, electrons emitted at points 1 and 2 travels along the dotted paths $P_2\ and\ P_1$ as shown in Figure3.27. In actual practice, the R.F fields act on these electrons and modify the orbits as shown by solid lines$P_1\ and\ P_2$. Electron emitted from point 1 has its tangential velocity opposed by the R.F field. Hence, this electron gets slowed down by the oscillation. Further since electron slows down, the deflecting force exerted on it by the magnetic field get reduced. Accordingly, this electron shifts closer to the anode and follows the solid path $P_{1.}$

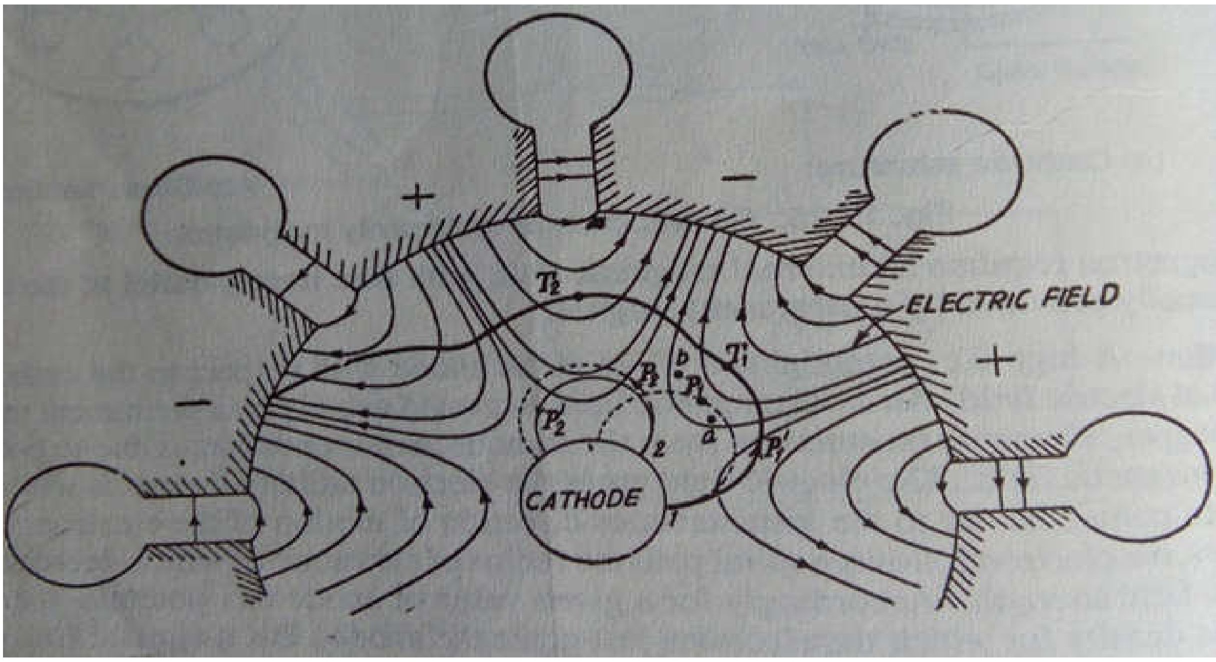


Figure 3.27. Electron pathway in a magnetron under oscillating condition.

Let the d.c anode voltage and the magnetic flux density B be so chosen that the tangential velocity of the electron makes the time of travel of the electron from position T_1 to T_2in Figure3.27 approximately equal half the periodic time of the R.F oscillation. Then this electron on reaching the point T_2 finds that the R.F field has reversed its polarity from that shown in Figure 3.27. As a result this electron continues to slow down and drift towards the anode. The velocity of this electron, however, does not alter appreciably since the energy acquired by the electron from the d.c anode-cathode voltage is a large extent delivered to the R.F. oscillation. Finally electron strikes anode surface.

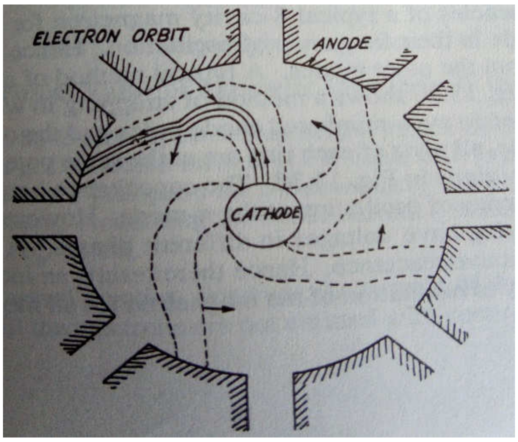


Figure3.28. Rotating space charge under oscillating condition

3.15. Magnetron Oscillator Conventional Tube Design

In a conventional model, we can identify the following types of tube design.

3.15.1. Hull or Single-Anode Magnetron

At very high magnetic field settings, the electrons are forced back onto the cathode, preventing current flow. At the opposite extreme, with no field, the electrons are free to flow straight from the cathode to the anode. There is a point between the two extremes, the critical value or Hull cut-off magnetic field (and cut-off voltage), where the electrons just reach the anode. At fields around this point, the device operates similar to a triode. However, magnetic control, due to hysteresis and other effects, results in a slower and less faithful response to control current than electrostatic control using a control grid in a conventional triode (not to mention greater weight and complexity), so magnetrons saw limited use in conventional electronic designs.

Evidently, when the magnetron is operating at the critical value, it would emit energy in the radio frequency spectrum.This occurs because a few of the electrons, instead of reaching the anode, continue to circle in the space between the cathode and the anode. Due to an effect now known as cyclotron radiation, these electrons radiate radio frequency energy. The effect is not very efficient. Eventually the electrons hit one of the electrodes, so the number in the circulating state at any given time is a small percentage of the overall current. It was also noticed that the frequency of the radiation depends on the size of the tube, and even early examples were built that produced signals in the microwave region.

Early conventional tube systems were limited to the high frequency bands, and although very high frequency systems became widely available in the late 1930s, the ultra high frequency and microwave regions were well beyond the ability of conventional circuits. The magnetron was one of the few devices able to generate signals in the microwave band and it was the only one that was able to produce high power at centimeter wavelengths.

3.15.2.Split-Anode Magnetron

Why Split-anode magnetron? Note that the original magnetron was very difficult to keep operating at the critical value, and even then the number of electrons in the circling state at any time was fairly low. This meant that it produced very low-power signals. Nevertheless, as one of the few devices known to create microwaves, interest in the device and potential improvements was widespread.The first major improvement was the split-anode magnetron, also known as a negative-resistance magnetron. As the name implies, this design used an anode that was split in two — one at each end of the tube — creating two half-cylinders. When both were charged to the same voltage the system worked like the original model. But by slightly altering the voltage of the two plates, the electron's trajectory could be modified so that they would naturally travel towards the lower voltage side. The plates were connected to an oscillator that reversed the relative voltage of the two plates at a given frequency.At any given instant, the electron will naturally be pushed towards the lower-voltage side of the tube. The electron will then oscillate back and forth as the voltage changes. At the same time, a strong magnetic field is applied, stronger than the critical value in the original design. This would normally cause the electron to circle back to the cathode, but due to the oscillating electrical field, the electron instead follows a looping path that continues toward the anodes.

Since all of the electrons in the flow experienced this looping motion, the amount of RF energy being radiated was greatly improved. And as the motion occurred at any field level beyond the critical value, it was no longer necessary to carefully tune the fields and voltages, and the overall stability of the device was greatly improved. Unfortunately, the higher field also meant that electrons often circled back to the cathode, depositing their energy on it and causing it to heat up. As this normally causes more electrons to be released, it could sometimes lead to a runaway effect, damaging the device.

3.15.3. Cavity Magnetron

The major advancement in magnetron design was the resonant cavity magnetron or electron-resonance magnetron, which works on entirely different principles. In this design the oscillation is created by the physical shaping of the anode, rather than external circuits or fields.

Before we proceed, a basic understanding of an oscillator is needed.

Now, an electronic oscillator can be made from a klystron tube, by providing a feedback path from output to input by connecting the "catcher" and "buncher" cavities with a coaxial cable or waveguide. When the device is turned on, electronic noise in the cavity is amplified by the tube and fed back from the output catcher to the buncher cavity to be amplified again. Because of the high Q of the cavities, the signal quickly becomes a sine wave at the resonant frequency of the cavities.

Mechanically, the cavity magnetron consists of a large, solid cylinder of metal with a hole drilled through the center of the circular face as shown in Figure 3.29. A wire acting as the cathode is run down the center of this hole, and the metal block itself forms the anode. Around this hole, known as the "interaction space, are a number of similar holes (resonators) drilled parallel to the interaction space, separated only a very short distance away. A small slot is cut between the interaction space and each of these resonators. The resulting block looks something like the cylinder on a revolver, with a somewhat larger central hole.

The parallel sides of the slots act as a capacitor while the anode block itself provides an inductor analog. Thus, each cavity forms its own resonant circuit, the frequency of which is defined by the energy of the electrons and the physical dimensions of the cavity.

The magnetic field is set to a value well below the critical, so the electrons follow arcing paths towards the anode. When they strike the anode, they cause it to become negatively charged in that region. As this process is random, some areas will become more or less charged than the areas around them. The anode is constructed of a highly conductive material, almost always copper, so these differences in voltage cause currents to appear to even them out. Since the current has to flow around the outside of the cavity, this process takes time.

During that time, additional electrons will avoid the hot spots and be deposited further along the anode, as the additional current flowing around it arrives too. This causes an oscillating current to form as the current tries to equalize one spot, then another.

The oscillating currents flowing around the cavities and their effect on the electron flow within the tube, causes large amounts of microwave radiofrequency energy to be generated in the cavities. The cavities are open on one end, so the entire mechanism forms a single, larger, microwave oscillator. A tap, normally a wire formed into a loop, extracts microwave energy from one of the cavities. In some systems the tap wire is replaced by an open hole, which allows the microwaves to flow into a waveguide.

As the oscillation takes some time to set up, and is inherently random at the start, subsequent startups will have different output parameters. Phase is almost never preserved, which makes the magnetron difficult to use in phased array systems. Frequency also drifts from pulse to pulse, a more difficult problem for a wider array of radar systems. Neither of these presents a problem for continuous-wave radars, nor for microwave ovens.

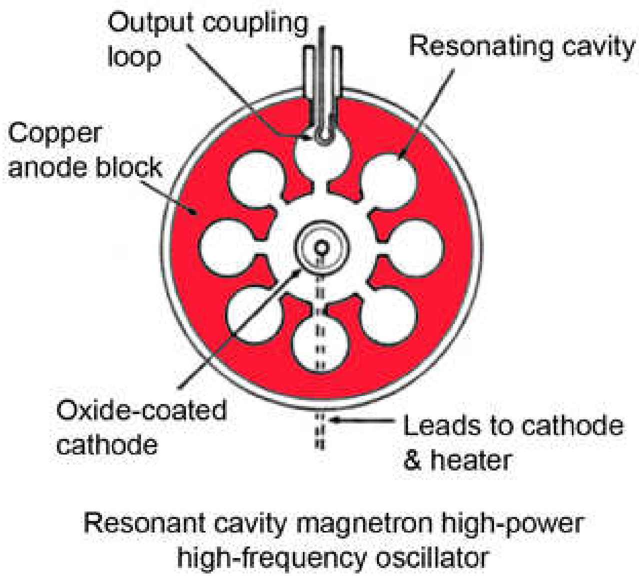


Figure 3.29. Diagram of a resonant cavity magnetron.

3.15.4. Common Features of Magnetron

All cavity magnetrons consist of a heated cathode placed at a high (continuous or pulsed) negative potential created by a high-voltage, direct-current power supply. The cathode is placed in the center of an evacuated, lobed, circular chamber. A magnetic field parallel to the filament is imposed by a permanent magnet. The magnetic field causes the electrons, attracted to the (relatively) positive outer part of the chamber, to spiral outward in a circular path, a consequence of the Lorentz force. Spaced around the rim of the chamber are cylindrical cavities.

Slots are cut along the length of the cavities that open into the central, common cavity space. As electrons sweep past these slots, they induce a high-frequency radio field in each resonant cavity, which in turn causes the electrons to bunch into groups. (This principle of cavity resonator is very similar to blowing a stream of air across the open top of a glass pop bottle.) A portion of the radio frequency energy is extracted by a short antenna that is connected to a waveguide (a metal tube, usually of rectangular cross section).

The waveguide directs the extracted RF energy to the load, which may be a cooking chamber in a microwave oven or a high-gain antenna in the case of radar.
The sizes of the cavities determine the resonant frequency, and thereby the frequency of the emitted microwaves. However, the frequency is not precisely controllable. The operating frequency varies with changes in load impedance, with changes in the supply current, and with the temperature of the tube. This is not a problem in uses such as heating, or in some forms of radar where the receiver can be synchronized with an imprecise magnetron frequency. Where precise frequencies are needed, other devices, such as the klystron are used.

The magnetron is a self-oscillating device requiring no external elements other than a power supply. A well-defined threshold anode voltage must be applied before oscillation will build up; this voltage is a function of the dimensions of the resonant cavity, and the applied magnetic field. In pulsed applications there is a delay of several cycles before the oscillator achieves full peak power, and the build-up of anode voltage must be coordinated with the build-up of oscillator output.Where there are even numbers of cavities, two concentric rings can connect alternate cavity walls to prevent inefficient modes of oscillation. This is called pi-strapping because the two straps lock the phase difference between adjacent cavities at pi radians (180°).
The modern magnetron is a relativelyan efficient device. In a microwave oven, for instance, a 1.1-kilowatt input will generally create about 700 watts of microwave power, an efficiency of around 65%. (The high-voltage and the properties of the cathode determine the power of a magnetron.) Large *S* band magnetrons can produce up to 2.5 megawatts peak power with an average power of 3.75 kW.Some large magnetrons are water cooled. The magnetron remains in widespread use in roles which require high power, but where precise control over frequency and phase is unimportant

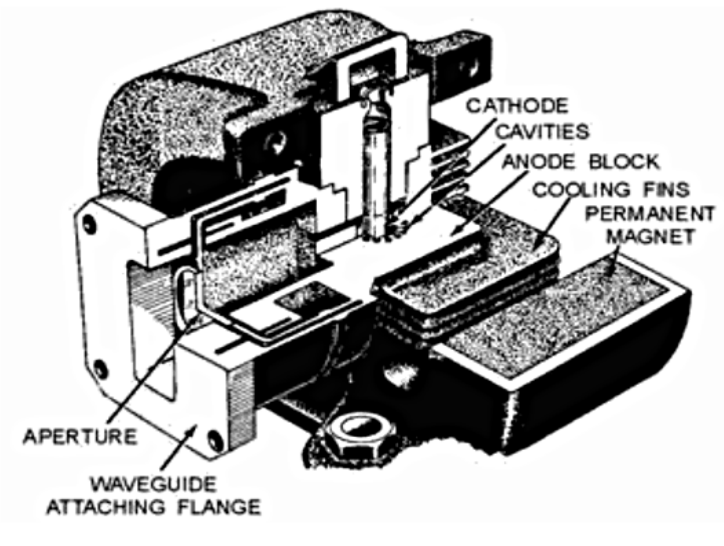


Figure3.30. Magnetron cutaway.

3.16. Klystron

3.16.1. KlystronDefinition

A klystron is a specialized linear-beam vacuum tube, invented in 1937 by American electrical engineers Russell and Sigurd Varian,which is used as an amplifier for high radio frequencies, from UHF up into the microwave range. The problem of transit time can be overcome by use of a klystron. Originally this type of microwave tube was developed as a low-power device, but the power-handling capacity of modern tubes is high. Klystron tubes are superior to magnetrons in that they can function as an oscillator, amplifier or detector. The klystron uses a process known as velocity modulation in which the velocity of the electron stream is varied as it moves through the tube. To be of greater use the electrons are converted into a density-modulated beam from which energy can be extracted. Low-power klystrons are used as oscillators in terrestrial microwave relay communications links, while high-power klystrons are used as output tubes in UHF television transmitters, satellite communication, and radar transmitters, and to generate the drive power for modern particle accelerators.In a klystron, an electron beam interacts with radio waves as it passes through resonant cavities, metal boxes along the length of a tube.The electron beam first passes through a cavity to which the input signal is applied. The energy of the electron beam amplifies the signal, and the amplified signal is taken from a cavity at the other end of the tube. The output signal can be coupled back into the input cavity to make an electronic oscillator to generate radio waves. The gain of klystrons can be high, 60 dB (one million) or more, with output power up to tens of megawatts, but the bandwidth is narrow, usually a few percent although it can be up to 10% in some devices.

A reflex klystron is an obsolete type in which the electron beam was reflected back along its path by a high potential electrode, used as an oscillator.Klystrons amplify RF signals by converting the kinetic energy in a DC electron beam into radio frequency power. A beam of electrons is produced by a thermionic cathode (a heated pellet of low work function material), and accelerated by high-voltage electrodes (typically in the tens of kilovolts). This beam is then passed through an input cavity resonator. RF energy is fed into the input cavity at, or near, its resonant frequency, creating standing waves, which produce an oscillating voltage which acts on the electron beam. The electric field causes the electrons to "bunch": electrons that pass through when the electric field opposes their motion are slowed, while electrons which pass through when the electric field is in the same direction are accelerated, causing the previously continuous electron beam to form bunches at the input frequency.

To reinforce the bunching, a klystron may contain additional "buncher" cavities as discussed later in this unit. The beam then passes through a "drift" tube in which the faster electrons catch up to the slower ones, creating the "bunches", then through a "catcher" cavity. In the output "catcher" cavity, each bunch enters the cavity at the time in the cycle when the electric field opposes the electrons' motion, decelerating them. Thus, the kinetic energy of the electrons is converted to potential energy of the field, increasing the amplitude of the oscillations. The oscillations excited in the catcher cavity are coupled out through a coaxial cable or waveguide. The spent electron beam, with reduced energy, is captured by a collector electrode.

3.16.2. Reflex Klystron Oscillator

The reflex klystron is a low power klystron tube with a single cavity, which functioned as an oscillator. It was used as a local oscillator in some radar receivers and a modulator in microwave transmitters the 1950s and 60s, but is now obsolete, replaced by semiconductor microwave devices discussed later.In the reflex klystron the electron beam passes through a single resonant cavity. Figure 3.31a depicts a cutaway Reflex Klystron with its internal structure.

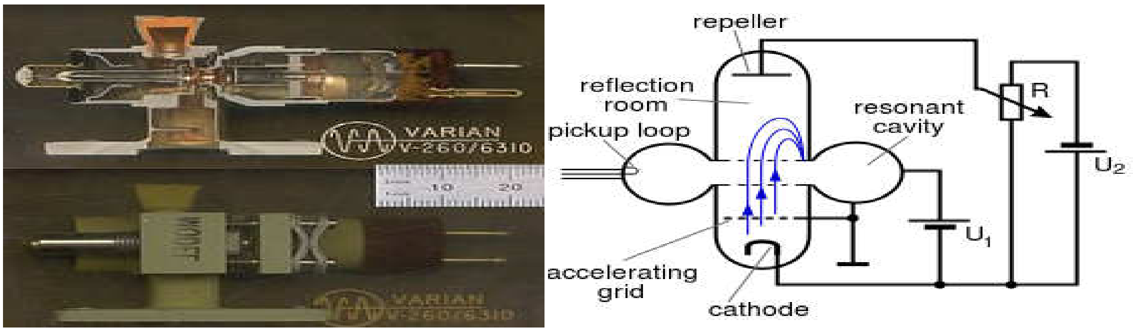


Figure 3.31a. Cutaway Reflex Klystron with its internal structure

The electrons are fired into one end of the tube by an electron gun. After passing through the resonant cavity, they are reflected by a negatively charged reflector electrode for another pass through the cavity, where they are then collected. The electron beam is velocity modulated when it first passes through the cavity. The formation of electron bunches takes place in the drift space between the reflector and the cavity. The voltage on the reflector must be adjusted so that the bunching is at a maximum as the electron beam re-enters the resonant cavity, thus ensuring a maximum of energy is transferred from the electron beam to the RF oscillations in the cavity. The reflector voltage may be varied slightly from the optimum value, which results in some loss of output power, but also in a variation in frequency. This effect is used to good advantage for automatic frequency control in receivers, and in frequency modulation for transmitters. The level of modulation applied for transmission is small enough that the power output essentially remains constant. At regions far from the optimum voltage, no oscillations are obtained at all.

There are often several regions of reflector voltage where the reflex klystron will oscillate; these are referred to as modes. The electronic tuning range of the reflex klystron is usually referred to as the variation in frequency between half power points—the points in the oscillating mode where the power output is half the maximum output in the mode. Figure 3.31b shows the physical representation of reflex klystron.

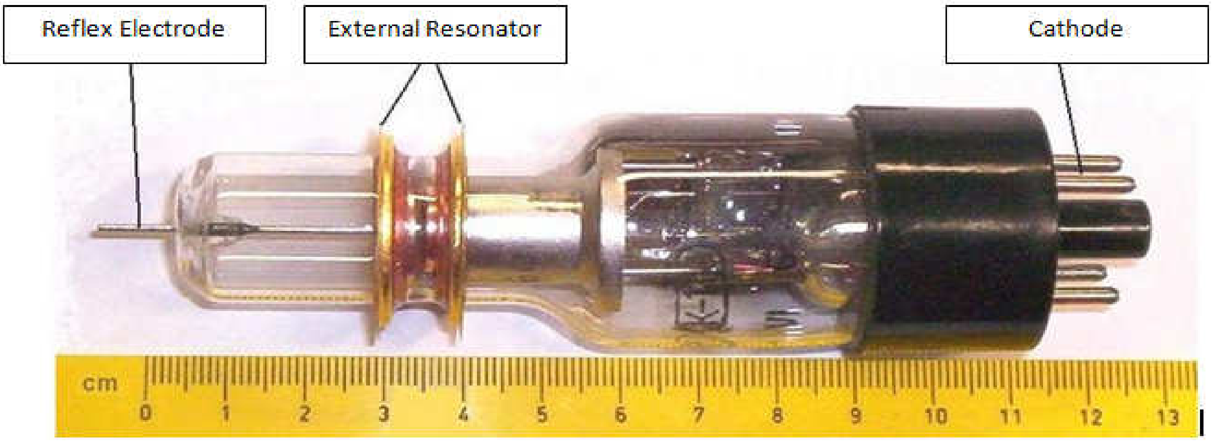


Figure 3.31b. Physical representation of reflex klystron.

So far, we have established that reflex klystron is a single cavity klystron. We can use Figure 3.32 to show its basic structure for proper understanding of Figure 3.31a, and 3.31b. It consists of a cathode, a focusing electrode placed at the cathode potential, a re-entrant cavity resonator and a repeller electrode. The resonator also severs as an anode. The repeller electrode is kept at a moderate negative with respect to the cathode. An electrode beam formed by the cathode and the focusing electrode passes through a gap in the resonator as shown in Figure 3.32 and travel towards the repeller. But the repeller being at negative voltage with respect to the cathode, turns the electron beam back toward the anode, after the electron have travelled to some point (such as point P in Figure 3.32) in the vicinity of the repeller. These returning electrons pass through the gap in the resonator a second time.

3.16.3. Reflex klystron Oscillatory Mechanism

In order to study the mechanism of a reflex klystron oscillator in a more practical context, we first assume that oscillations are sustained. The R.F oscillation assumes to be present in the resonator, produce an R.F voltage across the gap. The R.F voltage acts on the electron beam travelling toward repeller, causing the velocity of the electrons emerging from the gap into the repeller space to vary with time in accordance with this R.F voltage. Thus the electron stream entering the repeller space gets velocity modulated. Hence the electrons which pass the gap at different times in the R.F cycle, take different length of time to return the gap. The net result is the formation of bunches of electrons on reaching the gap on the return journey.

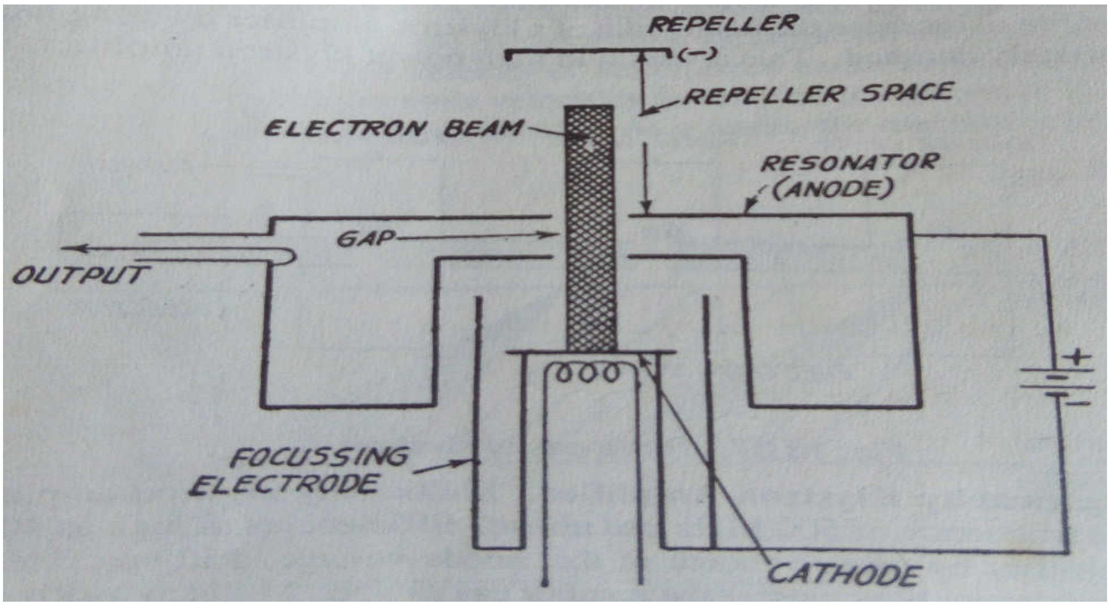


Figure3.32. Reflex klystron Frame layout.

Figure 3.32 gives the variation position in repeller space with time for a few typical electrons. Consider the electron which passes through the gap in its journey towards repeller at the instant when the gap voltage is zero and is just going negative. This may be called the reference electron and has been indicated by E_R in Figure 3.33. The distance versus time curve for this electron and has been indicated E_a which passes the gap a little easier the reference electronE_R. This electron enters the repeller space with greater velocity than the reference electron $E_R$and, therefore, penetrate further in the repeller space against the retarding field. It takes longer time to return to the anode and tend to arrive at the anode on its return path at the same time as the reference electron. Similarly an electronE_L, which passes the gap a little than the reference electron E_R enters the repeller space with lesser velocity than the reference electronE_R. This electron penetrates less in the repeller space, takes lesser time to return to the anode and tends to arrives at the anode on its return path at the same time as the reference electron.

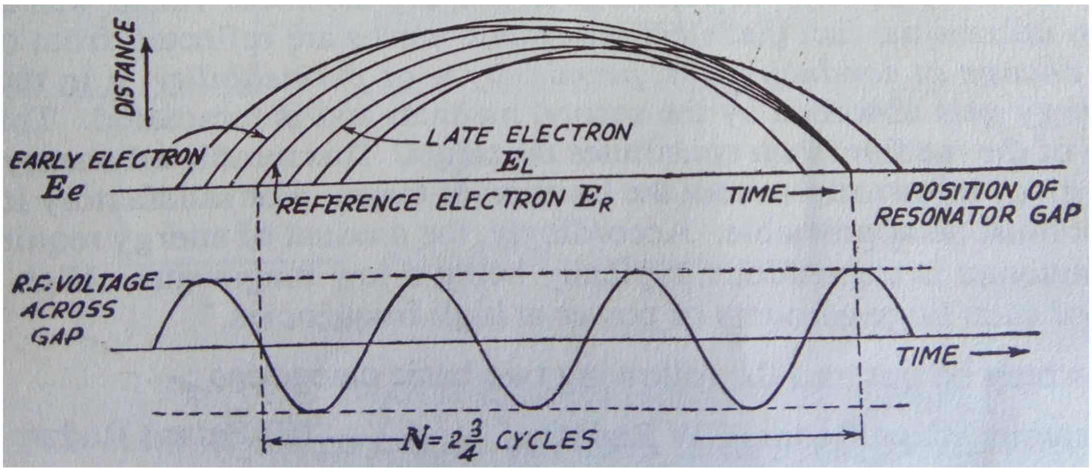


Figure3.33. Position versus time curves of electrons in repeller space of reflex klystron

3.16.4. Multi-cavity Klystron Amplifier

Among the tubes, which actually utilize the transit in achieving their normal operation, the more important types are: multicavity klystron, reflex klystron, travelling wave tube, backward wave oscillator and the magnetron and the magnetron.We have already discussed some of them but will emphasis the single, double and multicavity klystron.For example, Figure 3.35a and Figure 3.35b gives the basic construction of a two-cavity klystron. It consistent of the following constituent part (i) Cathode (ii) Field free draft space (iv) Catcher cavity and (v) Collector.A high-velocity electron beam produced by the cathode is passed through the input cavity resonator (called the buncher), through the field –free drift space and the output cavity resonator (called catcher), before the beam encounters the collector electrode. The cathode typically uses a pierce gun to produce the electron beam.


Figure 3.35a. Basic construction of a two-cavity klystron.

The two-cavity klystron of Figure3.35b is basically designed to work as an amplifier. However, it may be used as an oscillator if a small part of the output power is fed back to the input by suitably coupling the buncher and catcher cavities.

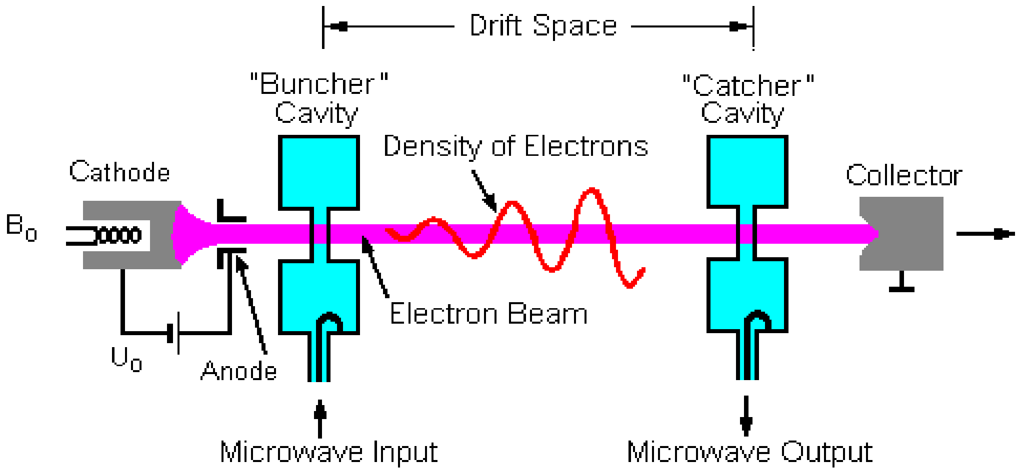


Figure 3.35b. Two-Cavity Klystron.

i. Two-Cavity Klystron

The simplest klystron tube is the two-cavity klystron shown in Figure 3.35b. In this tube, there are two microwave cavity resonators, the "catcher" and the "buncher". When used as an amplifier, the weak microwave signal to be amplified is applied to the buncher cavity through a coaxial cable or waveguide, and the amplified signal is extracted from the catcher cavity.

At one end of the tube is the hot cathode heated by a filament which produces electrons. The electrons are attracted to and pass through an anode cylinder at a high positive potential; the cathode and anode act as an electron gun to produce a high velocity stream of electrons. An external electromagnet winding creates a longitudinal magnetic field along the beam axis which prevents the beam from spreading

The beam first passes through the "buncher" cavity resonator, through grids attached to each side. The buncher grids have an oscillating AC potential across them, produced by standing wave oscillations within the cavity, excited by the input signal at the cavity's resonant frequency applied by a coaxial cable or waveguide. The direction of the field between the grids changes twice per cycle of the input signal. Electrons entering when the entrance grid is negative and the exit grid is positive encounter an electric field in the same direction as their motion, and is accelerated by the field. Electrons entering a half-cycle later, when the polarity is opposite, encounter an electric field which opposes their motion, and are decelerated.

Beyond the buncher grids is a space called the drift space. This space is long enough so that the accelerated electrons catch up to the retarded electrons, forming "bunches" longitudinally along the beam axis. Its length is chosen to allow maximum bunching at the resonant frequency, and may be several feet long.

The electrons then pass through a second cavity, called the "catcher", through a similar pair of grids on each side of the cavity. The function of the catcher grids is to absorb energy from the electron beam. The bunches of electrons passing through excite standing waves in the cavity, which has the same resonant frequency as the buncher cavity. Each bunch of electrons passes between the grids at a point in the cycle when the exit grid is negative with respect to the entrance grid, so the electric field in the cavity between the grids opposes the electrons motion. The electrons thus do work on the electric field, and are decelerated; their kinetic energy is converted to electric potential energy, increasing the amplitude of the oscillating electric field in the cavity. Thus the oscillating field in the catcher cavity is an amplified copy of the signal applied to the buncher cavity. The amplified signal is extracted from the catcher cavity through a coaxial cable or waveguide.After passing through the catcher and giving up its energy, the lower energy electron beam is absorbed by a "collector" electrode, a second anode which is kept at a small positive voltage.

ii. Three-cavity Klystron

This is the case where additional cavities are placed between the buncher and the catcher activities. Figure 3.36 shows a three-cavity klystron. As usual, the input signal is fed to the buncher cavity. But the oscillation is excited in the middle cavity also by the partially bunched electron stream passing the gap G_m of the middle cavity. An A.C voltage develops across this middle gap G_m and this voltage in turn acts on the electron stream. The cavity is slightly detuned by keeping its resonant frequency slightly the above the signal frequency. Hence causes the phase of the voltage across gap G_m to be so related to the electron stream at G_m as to cause further velocity modulation. This result in increased voltage gain and increased efficiency of the tube. It also becomes possible to increase the bandwidth of a klystron amplifier by using intermediate cavities and keeping them appropriately detuned. This is useful in high power klystron amplifiers used television.

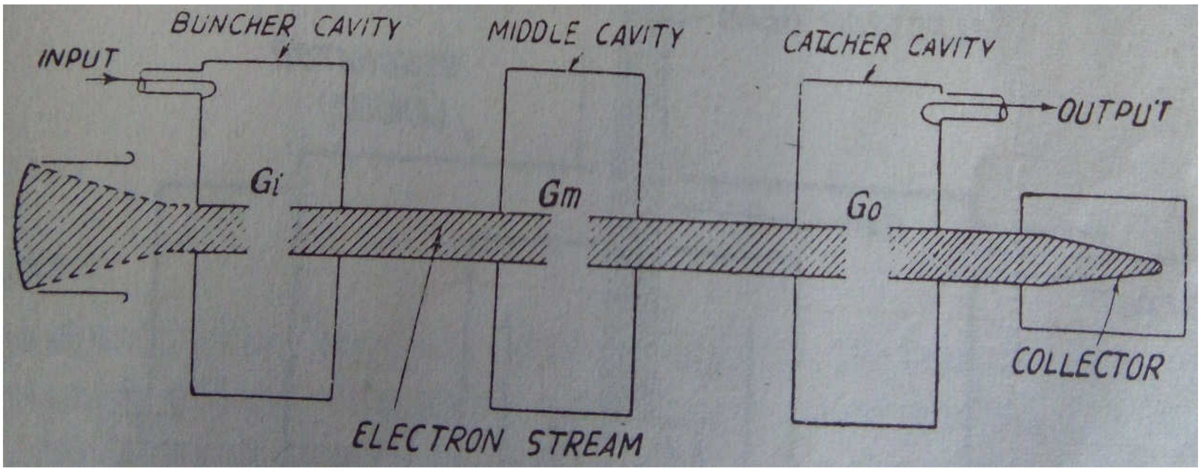


Figure3.36. Three-cavity klystron

3.16.5. Performance of Multicavity Klystron Amplifier

We should understand that the multicavity klystron amplifiers are principally used as power amplifiers at frequency of 500 MHz and above. Efficiencies as high as 40% may be achieved in mulitcavity klystron amplifier by proper choice of D.C anode voltage, drift-space length, and the signal amplitude so as to secure maximum bunching at the catcher cavity gap. Multicavity klystron amplifiers may deliver continuous wave power exceeding 30,000 kW at 3000 MHz. The power gain exceeding 30dB may be obtained in three-cavity klystron amplifier. The band with obtainable in multicavity klystron amplifier exceeds 6MHz and is, therefore adequate television. Further, klystron power amplifier maintains excellent linearity of amplifier up to about 80% of the maximum output power.However, the noise figure of Klystron tube is large being greater than 25dB. Accordingly, klystron amplifier is not used for amplification of weak microwave signals. For amplification of weak microwave signal, the use of travelling wave tubes is we preferred.

3.17. Semiconductor Substitute

Modern semiconductor technology has effectively replaced the reflex klystron and other related oscillators in most applications. Why? Now, let's talk about Semiconductors as an alternative briefly.

What is a semiconductor material? This is a material that has an electrical conductivity value falling between that of a conductor, such as copper, and an insulator, such as glass. Their resistance decreases as their temperature increases, which is behavior opposite to that of a metal. Their conducting properties may be altered in useful ways by the deliberate, controlled introduction of impurities ("doping") into the crystal structure. Where two differently-doped regions exist in the same crystal, a semiconductor junction is created. The behavior of charge carriers which include electrons, ions and electron holes at these junctions is the basis of diodes, transistors and all modern electronics. Semiconductor devices can display a range of useful properties such as passing current more easily in one direction than the other, showing variable resistance, and sensitivity to light or heat. Because the electrical properties of a semiconductor material can be modified by doping, or by the application of electrical fields or light, devices made from semiconductors can be used for amplification, switching, and energy conversion. Some useful properties of semiconductor in this context include:

i. Variable conductivity - Semiconductors in their natural state are poor conductors because a current requires the flow of electrons, and semiconductors have their valence bands filled, preventing the entry flow of new electrons. There are several developed techniques that allow semiconducting materials to behave like conducting materials, such as doping or gating. These modifications have two outcomes: n-type and p-type. These refer to the excess or shortage of electrons, respectively. An unbalanced number of electrons would cause a current to flow through the material.

ii. Heterojunctions – These occur when two differently doped semiconducting materials are joined together. For example, a configuration could consist of p-doped and n-doped germanium. This results in an exchange of electrons and holes between the differently doped semiconducting materials. The n-doped germanium would have an excess of electrons, and the p-doped germanium would have an excess of holes. The transfer occurs until equilibrium is reached by a process called recombination, which causes the migrating electrons from the n-type to come in contact with the migrating holes from the p-type. A product of this process is charged ions, which result in an electric field.

iii. Excited electrons - A difference in electric potential on a semiconducting material would cause it to leave thermal equilibrium and create a non-equilibrium situation. This introduces electrons and holes to the system, which interact via a process called ambipolar diffusion. Whenever thermal equilibrium is disturbed in a semiconducting material, the amount of holes and electrons changes. Such disruptions can occur as a result of a temperature difference or photons, which can enter the system and create electrons and holes. The process that creates and annihilates electrons and holes are called generation and recombination.

iv. Light emission - In certain semiconductors, excited electrons can relax by emitting light instead of producing heat. These semiconductors are used in the construction of light-emitting diodes and fluorescent quantum dots.

v. Thermal energy conversion - Semiconductors have large thermoelectric power factors making them useful in thermoelectric generators, as well as high thermoelectric figures of merit making them useful in thermoelectric coolers.

4.0 Conclusion

Microwave devices can be used to generate UHF needed for signal broadcasting. When greater power is required, more complex devices must be used. Also the process of miniaturization imposes problems at higher frequencies due to higher power dissipation. Because of these restrictions alternative designs have been developed, among them multicavity magnetrons, klystrons, reflex klystrons and travelling wave tubes. With the advances in technological applications, you must always look out for the best semiconductor substitute in place of the microwave devices owing to the benefits of semiconductor materials.

5.0 Summary

UHF is the radio frequencies in the range between 300 megahertz (MHz) and 3gigahertz (GHz). There are factors that affect Ultra-High Frequency Responses such as Lead Inductance and Internal Capacitances, Grid-To-Plate Capacitance Vs Input Admittance, Cathode Lead Inductance and Grid-Capacitance Vs Input Admittance, Transit Time Effect in Diodes, Tube Output Conductance and Tube Condition of Current. The remedy to power issues in UHF is presented. The various categories of UHF tube construction were discussed such as Single ended construction, Double ended construction and Disk seal construction. The various types of oscillators for UHF include multicavity magnetrons, klystrons, reflex klystrons , travelling wave tubes, and semiconductors.

6.0 Tutor-Marked Assignment (TMA)

1. Explain the term UHF and discuss why it is used for television broadcasting?
2. In representing and electron tube, outline three quantizes that need to be studied.
3. Explain the effect of Input Admittance on Grid-To-Plate Capacitance. Also, what is the effect of the Input Admittance and Lead Inductance with Internal Capacitances in an UHF system
4. Using the Millman's theorem, derive the total input admittance in a Triode amplifier.
5. State the Three factors which limit the range of frequency of electron tube oscillators.
6. What is Tank Circuit Q?
7. What do you understand by Parasitic Oscillation and how can this be Suppressed?

7.0 References/Further Readings

1. The Magnetron (2009)". Bournemouth University. Retrieved 23 August 2009.
2. Schroter, B. (Spring 2008). "How important was Tizard’s Box of Tricks?" (PDF). Imperial Engineer. **8**: 10. Retrieved 2009-08-23.
3. Ma, L. "3D Computer Modeling of Magnetrons." University of London Ph.D. Thesis. December 2004. Accessed 2009-08-23.

4. L.W. Turner,(ed), Electronics Engineer's Reference Book, 4th ed. Newnes-Butterworth, London 1976 ISBN 9780408001687, pages 7-71 to 7-77
5. Albert W. Hull (1921) "The effect of a uniform magnetic field on the motion of electrons between coaxial cylinders," Physical Review, 18 (1) : 31-57. See also: Albert W. Hull, "The magnetron," Journal of the American Institute of Electrical Engineers, vol. 40, no. 9, pages 715-723
6. W. Kaiser, "The Development of Electron Tubes and of Radar technology: The Relationship of Science and Technology," pp. 217 – 236.
7. Blumtritt, H. Petzold and W. Aspray, eds.(1994). Tracking the History of Radar, IEEE, Piscataway, NJ, USA.
8. James E. Brittain(1985).The magnetron and the beginnings of the microwave age," Physics Today, vol. 38, pages 60-67.
9. Japanese engineers: Hidetsugu Yagi,(1928); Beam transmission of ultra-short waves," Proceedings of the Institute of Radio Engineers, vol. 16, no. 6, pages 715-741
10. Willshaw, W. E.; L. Rushforth; A. G. Stainsby; R. Latham; A. W. Balls; A. H. King (1946). "The high-power pulsed magnetron: development and design for radar applications". The Journal of the Institution of Electrical Engineers - Part IIIA: Radiolocation. 93 (5). doi:10.1049/ji-3a-1.1946.0188. Retrieved 22 June 2012.
11. Jr. Raymond C. Watson (2009). Radar Origins Worldwide: History of Its Evolution in 13 Nations Through World War II. Trafford Publishing. pp. 315–. ISBN 978-1-4269-2110-0. Retrieved 24 June 2011.
12. https://www.nonstopsystems.com/radio/pdf-hell/article-UHF-TX-tubes.pdf
13. C.E. Haller(1942), The Design and Development of Three New UltraHigh-Frequency Transmitting Tubes, **:** Proceedings of the IRE , Vol. 30, Issue1Pp**.** 20 – 26, **DOI:** 10.1109/JRPROC.1942.232001.
14. Andrew Leven (2000). Telecommunication Circuits and Technology, ISBN 0 7506 5045 1, MPG Books, Bodmin Cornwall, India

UNIT 4. HIGH IMPEDANCE CIRCUITS

1.0 Introduction
2.0 Objectives
3.0 Main Content
3.1. Definition of Impedance
3.2. Why high input and low output impedance
3.3. Four Terminal Transmission Line Model Elements
3.3.1. High-frequency transmission
3.3.2. Characteristics Impedance
3.4. Special cases of lossless transmission lines
3.4.1. Half wave length
3.4.2. Quarter wave length
3.4.3. Matched Load
3.4.4. Short Circuit
3.4.5. Open Circuit
3.4.6. Stepped Transmission Line
3.5. Types of Transmission Line
3.6 Insertion Loss
3.7. Two Port NetworkRepresentations
3.7.1. Z-Parameters
3.7.2. Y-Parameters
3.7.3. H-Parameters
3.7.4. Transmission Parameters
3.8. Interconnection of Two Port Network
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0 References/Further Readings

1.0 INTRODUCTION

So far we have looked at the communication system model were we discussed the elemental models of a communication system. The issue of noise analysis and signal computation was presented. In the communication system, we looked at modulation and types of modulation. Frequency Modulation and ultra high frequency tubes and oscillators were also studied. These are all vital in electronic communications systems. We shall now consider the concept of impedance in such systems in this module.

Now, in electronic communication systems, high impedance implies that a point in a circuit (a node) allows a relatively small amount of current through, per unit of applied voltage at that point. High impedance circuits are low current and potentially high voltage, whereas low impedance circuits are the opposite (low voltage and potentially high current). More accurately, a high impedance connection has a well-defined current (which, in many applications, can be conveniently chosen to be near zero) and a loosely-defined voltage.

A low-impedance connection has a well-defined voltage (also chosen to be near zero if possible) and a loosely-defined current. Also, in digital circuits, a high impedance (also known as hi-Z, tri-stated, or floating) output is not being driven to any defined logic level by the output circuit. The signal is neither driven to a logical high nor low level; this third condition leads to the description "tri-stated”. Such a signal can be seen as an open circuit (or "floating" wire) because connecting it to a low impedance circuit will not affect that circuit; it will instead itself be pulled to the same voltage as the actively driven output. This is the basis for bus-systems in computers, among many other uses.

In analog circuits, a high impedance node is one that does not have any low impedance paths to any other nodes in the frequency range being considered. Since the terms low and high depend on context to some extent, it is possible in principle for some high impedance nodes to be described as low impedance in one context, and high impedance in another; so the node (perhaps a signal source or amplifier input) has relatively low currents for the voltages involved.High impedance nodes have higher thermal noise voltages and are more prone to capacitive and inductive noise pick up.

When testing, they are often difficult to probe as the impedance of an oscilloscope or multimeter can heavily affect the signal or voltage on the node. High impedance signal outputs are characteristic of some transducers (such as crystal pickups); they require a very high impedance load from the amplifier to which they are connected. Vacuum tube amplifiers and field effect transistors more easily supply high-impedance inputs than bipolar junction transistor-based amplifiers, although current buffer circuits or step-down transformers can match a high-impedance input source to a low impedance amplifier.

This unit will look at high impedance network with particular emphasis on two–port network which is widely used in the communications systems. We shall discuss the characteristic impedanceZ_0 and the propagation coefficienctsγ. Furthermore, Z-parameters, and Y-parameters, are discussed at elementary level. Practical examples regarding 2-port network interconnectivity is discussed.

2.0 Objectives

After going through this unit, you should be able to:

- ✓ understand the concept of impedance
- ✓ differentiate between different types of impedance elements
- ✓ understand the reason for high input and low output impedance in networks
- ✓ compare the different transmission line topologies
- ✓ distinguish between characteristics impedance and propagation constant
- ✓ derive the various models of two port network
- ✓ know the importance of using internetworked two port network system
- ✓ solve mathematical problems on two port networks

3.0 Main Content

3.1. Definition of Impedance

You may ask, what is the meaning of Impedance?

Impedance, denoted Z, is an expression of the opposition that an electronic component, circuit, or system offers to alternating and/or direct electric current. It is a vector (two-dimensional) quantity consisting of two independent scalar (one-dimensional) phenomena: resistance and reactance.

Now, Resistance, denoted R, is a measure of the extent to which a substance opposes the movement of electrons among its atoms. The more easily the atoms give up and/or accept electrons, the lower the resistance, which is expressed in positive real number ohms. Resistance is observed with alternating current (AC) and also with direct current (DC). Examples of materials with low resistance, known as electrical conductors, include copper, silver, and gold. High-resistance substances are called insulators or dielectrics, and include materials such as polyethylene, mica, and glass. A material with intermediate levels of resistance is classified as a semiconductor (See previous unit). Examples are silicon, germanium, and gallium arsenide.

Reactance, denoted X, is an expression of the extent to which an electronic component, circuit, or system stores and releases energy as the current and voltage fluctuate with each AC cycle. Reactance is expressed in imaginary number ohms. It is observed for AC, but not for DC. When AC passes through a component that contains reactance, energy might be stored and released in the form of a magnetic field, in which case the reactance is inductive (denoted $+ jX_L$); or energy might be stored and released in the form of an electric field, in which case the reactance is capacitive (denoted $-jX_C$). Reactance is conventionally multiplied by the positive square root of -1, which is the unit imaginary number called the *j operator*, to express Z as a complex number of the form $R + jX_L$ (when the net reactance is inductive) or $R - jX_C$(when the net reactance is capacitive).
The illustration in Figure 4.1(2)shows a coordinate plane modified to denote complex-number impedances. Resistance appears on the horizontal axis, moving toward the right.(The left-hand half of this coordinate plane is not normally used because negative resistances are not encountered in common practice.).
Inductive reactance appears on the positive imaginary axis, moving upward. Capacitive reactance is depicted on the negative imaginary axis, moving downward. As an example, a complex impedance consisting of 4 ohms of resistance and $+j5$ ohms of inductive reactance is denoted as a vector from the origin to the point on the plane corresponding to $4 + j5$.

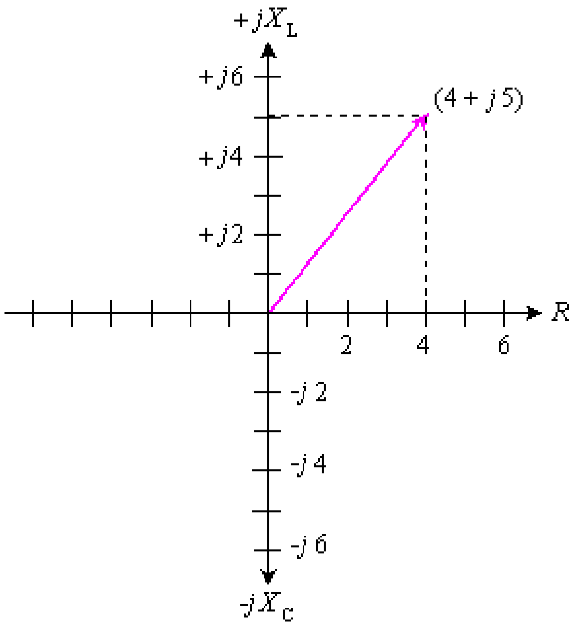


Figure 4.1. Polar diagram of complex impedance.

In series circuits, resistances and reactances add together independently. Suppose a resistance of 100.00 ohms is connected in a series circuit with an inductance of 10.000H?.At 4.0000 MHz, the complex impedance is:

$Z_{RL} = R + jX_L = 100.00 + j251.33$ (4.1)

If a capacitor of 0.0010000 ? F is put in place of the inductor, the resulting complex impedance at 4.0000 MHz is:

$Z_{RC} = R - jX_C = 100.00 - j39.789$ (4.2)

If all three components are connected in series, then the reactances add, yielding a complex impedance of:

$Z_{RLC} = 100 + j251.33 - j39.789 = 100 + j211.5.$ (4.3)

This is the equivalent of a 100-ohm resistor in series with an inductor having $+j211.5$ ohms of reactance. At 4.0000 MHz, this reactance is presented by an inductance of 8.415 ?H, as determined by plugging the numbers into the formula for inductive reactance and working backwards.(See the definition of for this formula, and for the corresponding formula for capacitive reactance.).

Parallel RLC circuits are more complicated to analyze than are series circuits. To calculate the effects of capacitive and inductive reactance in parallel, the quantities are converted to inductive susceptance and capacitive susceptance. Susceptance is the reciprocal of reactance. Susceptance combines with conductance, which is the reciprocal of resistance, to form complex admittance, which is the reciprocal of complex impedance. An intermediate electronics text or reference book is recommended for further study.

3.2. Why high input and low output impedance

This is a very good question regarding the need high input and low output impedance in an amplifier circuit? Now, high input impedance allows the amplifier circuitto be connected to many kinds of source devices without concern that it will load down the source. Low output impedance allows the amplifier to be connected to many different kinds of load devices without concern that its output may be reduced by the load. Ahigh impedance means that the circuit draws or provides little power on the signal. Low impedance means that the circuit draws or provides more power on the signal.

In general it is a good idea to follow the rule that the circuit should have a high input impedance (so that it can interface correctly with any other circuit that may or may not be able to provide enough power) and a low output impedance (so that the circuit may interface correctly with another, that draws a lot of input power).

As long as the output impedance of the signal is lower than the input impedance of the circuit that the signal is going into, everything works well.Hence, it can be opined that we avoid Input signal voltage attenuation by high impedance at the input side and at the same time we also avoid Output current attenuation by lowering the resistance at the output side for optimaloperation of an amplifier circuitry.

3.3. Four Terminal Transmission Line Model Elements

3.3.1. High-frequency transmission

High-frequency transmission lines can be defined as those designed to carry electromagnetic waves whose wavelengths are shorter than or comparable to the length of the line. Under these conditions, the approximations useful for calculations at lower frequencies are no longer accurate. This often occurs with radio, microwave and optical signals, metal mesh optical filters, and with the signals found in high-speed digital circuits. Figure 4.2 shows the schematic variations for a typical transmission line.

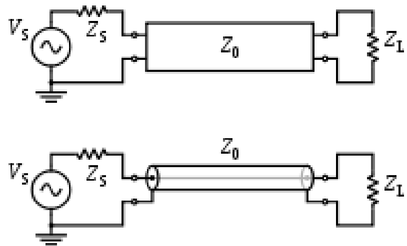


Figure 4.2. Transmission line Schematics.

In electronic communications, a transmission line is a specialized cable or other structure designed to conduct alternating current of radio frequency, that is, currents with a frequency high enough that their wave nature must be taken into account. Transmission lines are used for purposes such as connecting radio transmitters and receivers with their antennas, distributing cable television signals, trunk lines routing calls between telephone switching centers, computer network connections and high speed computer data buses. This section covers two-conductor transmission line such as parallel line (ladder line), coaxial cable, stripline, and microstrip. For the purposes of analysis, an electrical transmission line can be modeled as a two-port network shown in Figure 4.3.

Figure 4.3 Electrical transmission line

Let us say that the network is assumed to be linear (i.e. the complex voltage across either port is proportional to the complex current flowing into it when there are no reflections), and the two ports are assumed to be interchangeable. If the transmission line is uniform along its length, then its behavior is largely described by a single parameter called the characteristic impedance, symbol Z_0. This is the ratio of the complex voltage of a given wave to the complex current of the same wave at any point on the line as shown in Figure 4.4. Typical values of Z_0 are 50 or 75 ohms for a coaxial cable, about 100 ohms for a twisted pair of wires, and about 300 ohms for a common type of untwisted pair used in radio transmission.

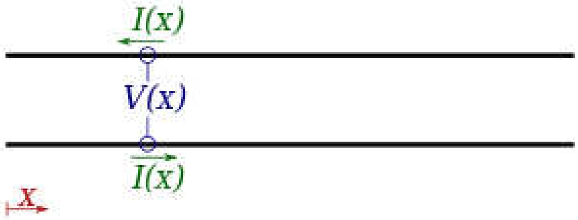


Figure 4.4. Electrical transmission line depicting characteristic impedance

When sending power down a transmission line, it is usually desirable that as much power as possible will be absorbed by the load and as little as possible will be reflected back to the source. This can be ensured by making the load impedance equal to Z_0, in which case the transmission line is said to be matched.

Some of the power that is fed into a transmission line is lost because of its resistance. This effect is called ohmic or resistiveloss (e.g. ohmic heating). At high frequencies, another effect called dielectric loss becomes significant, adding to the losses caused by resistance. Dielectric loss is caused when the insulating material inside the transmission line absorbs energy from the alternating electric field and converts it to heat (e.g. dielectric heating). The transmission line is modeled with a resistance (R) and inductance (L) in series with a capacitance (C) and conductance (G) in parallel. The resistance and conductance contribute to the loss in a transmission line.The total loss of power in a transmission line is often specified in decibels per metre (dB/m), and usually depends on the frequency of the signal. The manufacturer often supplies a chart showing the loss in dB/m at a range of frequencies. A loss of 3 dB corresponds approximately to a halving of the power.

3.3.2. Characteristics Impedance

As you can see from Figure 4.4, energy travels along a transmission line in the form of an electromagnetic wave, i.e., the wave set up by the signal source known as incident (or forward) wave. It is only when the load impedance at the receiving end is a reflectionless match for the line will all the energy be transferred to the load. If refelctionless matching is not achieved, energy will be reflected back along the line in the form of a reflected wave (hence, the name reflectionless matching). Because of the distributed nature of a transmission line, the question may be asked: Precisely to what impedance must the load be matched? This can be answered by considering a hypothetical line, infinite in length ad for which no reflection can occur, since the incident wave never reaches the end. The ratio of maximum voltage to maximum current at any point on such a line is found to be constant, that is, independent of position. We have earlier called this the characteristic impedance Z_0. Now, if a finite length of line is terminated in a load $Z_L = Z_0$, this will appear as an infinite line to the incident wave since at all points, including the load termination, the ratio of voltage to current will equal Z_0. Thus, the characteristic impedance of a transmission line is the ratio of voltage to current at any point along the line on which no reflected wave exists.

With a sinusoidal signal of angular frequency ω rad/sec, the characteristic impedance in terms of the primary constants is given as

$$Z_0 = \sqrt{\frac{R+j\omega L}{G+j\omega C}}\Omega \quad (4.4)$$

At low frequency such that R>> $\omega C,$ the expression for Z_0 reduces to

$$Z_0 = \sqrt{\frac{R}{G}}\ \Omega \quad (4.5)$$

and at high frequencies such that R<<ωL and G<< ωC, it becomes

$$Z_0 = \sqrt{\frac{L}{C}}\ \Omega \quad (4.6)$$

From Equ.4.4 to 4.6, it is obvious that each limiting value is purely resistive (there is no *j* coefficient) and independent of frequency. Between these limits, Z_0 is complex and frequency dependent. It is known that for most practical lines, it is capacitive. However, above a few tens of kilohertz for two-wire lines and a few hundred kilohertz for coaxial lines, the high-frequency approximation for Z_0 is sufficiently accurate for most practical purposes. See reference for detailed mathematical models for transmission lines.

3.4. Special cases of lossless transmission lines

3.4.1. Half wave length

For the special case where $\beta\ell = n\pi$ where n is an integer (meaning that the length of the line is a multiple of half a wavelength), the expression reduces to the load impedance so that

$$Z_{in} = Z_L \quad (4.7)$$

For alln. his includes the case when $n = 0$, meaning that the length of the transmission line is negligibly small compared to the wavelength. The physical significance of this is that the transmission line can be ignored (i.e. treated as a wire) in either case.

3.4.2. Quarter wave length

For the case where the length of the line is one quarter wavelength long, or an odd multiple of a quarter wavelength long, the input impedance becomes

$$Z_{in} = \frac{Z_0^2}{Z_L} \quad (4.8)$$

3.4.3. Matched Load

As you can see from Figure 4.5, another special case is when the load impedance is equal to the characteristic impedance of the line (i.e. the line is matched), in which case the impedance reduces to the characteristic impedance of the line so that

$Z_L = V_s = Z_{in} = Z_L = Z_0$, For all ℓ and all λ.

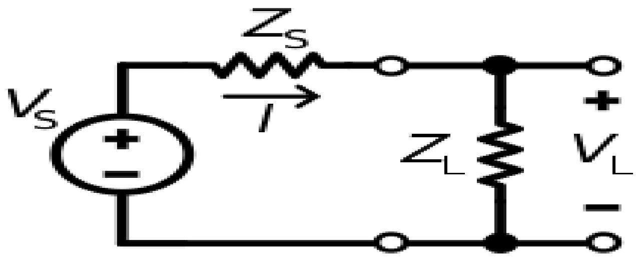


Figure 4.5. Electrical transmission line depicting characteristic impedance

3.4.4. Short Circuit

For the case of a shorted load $Z_L = 0$, the e input impedance is purely imaginary and a periodic function of position and wavelength (frequency)

$$Z_{in}(\ell) = -\mathrm{j}Z_0\tan(\beta\ell), \tag{4.9}$$

3.4.5. Open Circuit

For the case of an open load (i.e. $Z_L = \infty$), the input impedance is once again imaginary and periodic.

$$Z_{in}(\ell) = -\mathrm{j}Z_0\mathrm{Cot}(\beta\ell). \tag{4.10}$$

3.4.6. Stepped Transmission Line

A stepped transmission lineis used for wide range impedance matching. It can be considered as multiple transmission line segments connected in series, with the characteristic impedance of each individual element to be Z_0, ias shown in Figure4.6.

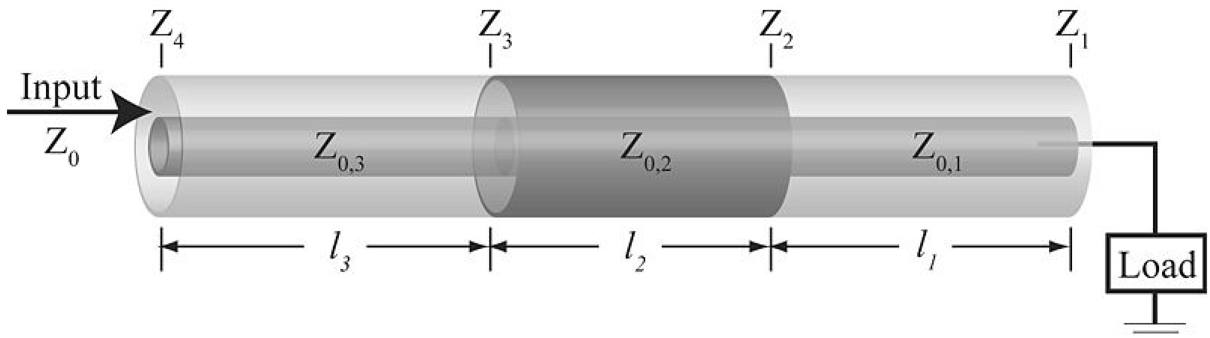


Figure 4.6. Impedance of Stepped Transmission Line.

The input impedance can be obtained from the successive application of the chain relation

$$Z_{i+1} = Z_{0,i}\,\frac{Z_{i+}\,jZ_{0,i}\tan\,(\beta_i\ell_i)}{Z_{i+}\,jZ_{1,i}\tan\,(\beta_i\ell_i)} \tag{4.11}$$

Where β_i = is the wave number of the *i-th* transmission line segment and β_ℓ is the length of the segment that loads the *i-th* segment.

3.5. Types of Transmission Lines

We shall now introduce practical examples of typical transmission lines. Some classical examples includes: coaxial cable, microstrip, stripline, balanced lines, twisted pair, star quad twin-lead, lecher lines and single-wire lines. These could be synthesized into a two port networks. These are used widely in electronic communication systems. The most critical attributes of these type of network are the characteristic impedance Z_0 and the propagation coefficiencts γ. the characteristic impedance Z_0 is defined as the input impedance of an infinite number of two port networks in cascade while the propagation coefficient explains the natural logarithm of the ratio of the input and output currents of a network terminated on an iterative fashion at both ends.The iterative impedance of a two port network is best defines as the value of the impedance measured at one pair of terminals when the other pair of terminals is terminated with an impedance of the same value as depicted in Figure4.12.

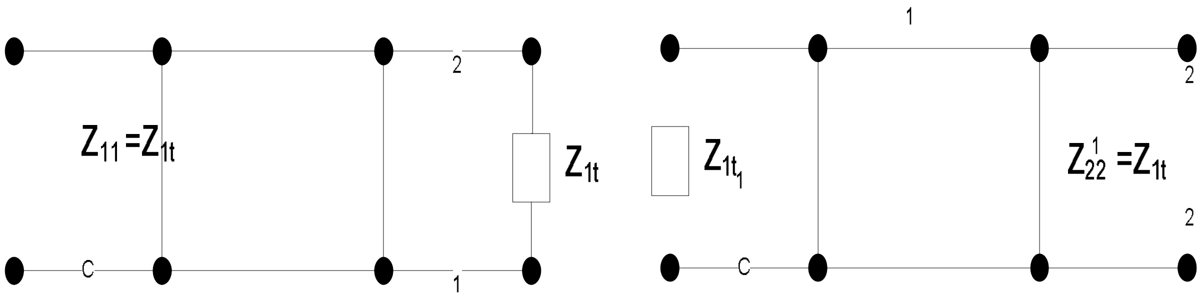


Figure 4.12. Iterative impedance Z_{it} and Iterative impedance Z_{it2} of a 2-port network.

The image impedance Z_{01} and Z_{02} of a two port network are those impedances such that if one of them, say $Z0_2$ is connected across terminals 22^I, then the impedance seen when looking in at terminals 11^I is equal to the other i.e. Z_{01}. Likewise if Z_{01} terminates terminal 11^I, then the impedance seen looking in at terminals 22^I is $Z_{02.}$ Figure 4.13explains this scenario.

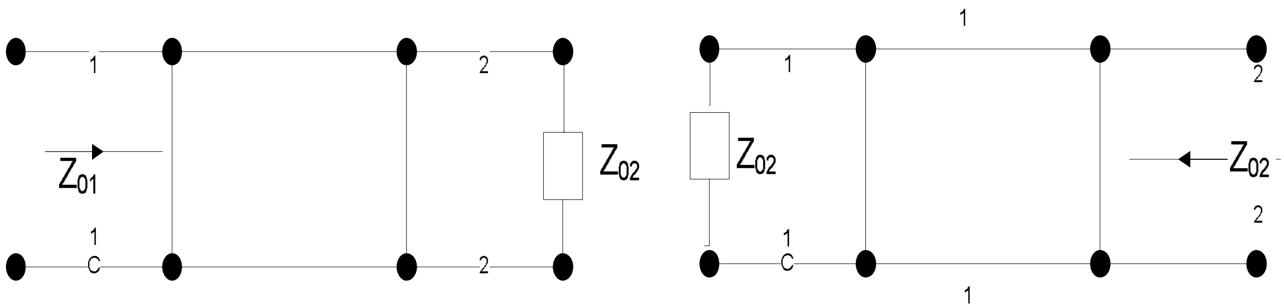


Figure 4.13. Image impedance of a 2-port network.

By a new definition image impedances can be derived as the two terminating impedances which are such that when they are simultaneously connected to the appropriate terminals of the network, each terminating impedance is equal to the impedance presented to it.

3.6. Insertion Loss

Another interesting concept here is the insertion loss. Basically, when a network is inserted between a generator and load, there is a loss of power in the load due to the mismatch between the generator and the network. There is also a loss of power between the network and load leading to an attendant attenuation through the network. The loss is called the insertion loss. If $P_1$is the power in the load without the network inserted and P_2 is the power in the load while the network inserted, then, the insertion loss is defined as

Insertion loss = $10\log\frac{P_1}{P_2}$ dB = $20\log\frac{V_1}{V_2}$ or $20\log\frac{I_1}{I_2}$ (4.12)

3.7. Two-Port Network Representations

A general two-port network is shown in Figure 4.14. However, we can look at the two-port network in a more simplified perspective. A two-port network is an circuit with two pairs of terminals to connect to external circuits. Two terminals constitute a port if the currents applied to them satisfy the essential requirement known as the port condition: the electric current entering one terminal must equal the current emerging from the other terminal on the same port. The ports constitute interfaces where the network connects to other networks, the points where signals are applied or outputs are taken. In a two-port network, often port 1 is considered the input port and port 2 is considered the output port.The two-port network model is used in mathematical circuit analysis techniques to isolate portions of larger circuits. A two-port network is regarded as a "black box" with its properties specified by a matrix of numbers. This allows the response of the network to signals applied to the ports to be calculated easily, without solving for all the internal voltages and currents in the network. It also allows similar circuits or devices to be compared easily. For example, transistors are often regarded as two-ports, characterized by their h- which are listed by the manufacturer. Any linear circuit with four terminals can be regarded as a two-port network provided that it does not contain an independent source and satisfies the port conditions.

Examples of circuits analyzed as two-ports are filters, matching networks, transmission lines, transformers, and small-signal models for transistors (such as the hybrid-pi model). The analysis of passive two-port networks is an outgrowth of reciprocity theorems first derived by Lorentz. In two-port mathematical models, the network is described by a 2 by 2 square matrix of complex numbers. The common models that are used are referred to as z-parameters, y-parameters, h-parameters, g-parameters, and ABCD-parameters, each described individually below. These are all limited to linear networks since an underlying assumption of their derivation is that any given circuit condition is a linear superposition of various short-circuit and open circuit conditions. They are usually expressed in matrix notation, and they establish relations between the variables which are shown in Figure 1.

i. Voltage across port 1 V_1

ii. Current into port 1, I_1

iii. Voltage across port 2, V_2

iv. Current into port 2, I_2

An example of a two-port network with symbol definitions is demonstrated in Figure 4.14. . Notice the port condition is satisfied: the same current flows into each port as leaves that port.

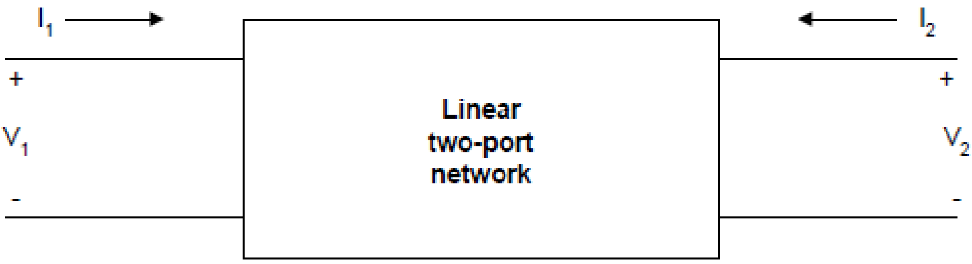


Figure 4.14. Two-port network with symbol definitions.

In this case, I_1 and V_1 are the input current and voltage, respectively. Also, I_2 and V_2 are output current and voltage, respectively. It is assumed that the linear two-port circuit contains no independent sources of energy and that the circuit is initially at rest (no stored energy). Furthermore, any controlled sources within the linear two-port circuit cannot depend on variables that are outside the circuit

The difference between the various models lies in which of these variables are regarded as the independent variables. These current and voltage variables are most useful at low-to-moderate frequencies. At high frequencies (e.g., microwave frequencies), the use of power and energy variables is more appropriate, and the two-port current–voltage approach is replaced by an approach based upon scattering parameters.

3.7.1. Z-Parameters

A two-port network can be described by z-parameters as

$$V_1 = Z_{11}I_1 + Z_{22}I_1$$
$$V_2 = Z_{21}I_1 + Z_{22}I_1 \qquad (4.13)$$

In matrix form, the above equation can be rewritten as

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \qquad (4.14)$$

The zparameter can be found as follows

$$Z_{11} = \frac{V_1}{I_1} |I_2=0$$
$$Z_{12} = \frac{V_1}{I_2} |I_1= 0$$
$$Z_{21} = \frac{V_2}{I_1} |I_2= 0 \qquad (4.15)$$
$$Z_{22} = \frac{V_2}{I_2} |I_1= 0$$

The z-parameters are also called open-circuit impedance parameters since they are obtained as a ratio of voltage and current and the parameters are obtained by open-circuiting port 2 (I_2 = 0) or port1 (I_1= 0). The following example shows a technique for finding the z-parameters of a simple circuit.

Example 1
For the T-network shown in Figure 4.15, find the z-parameters.

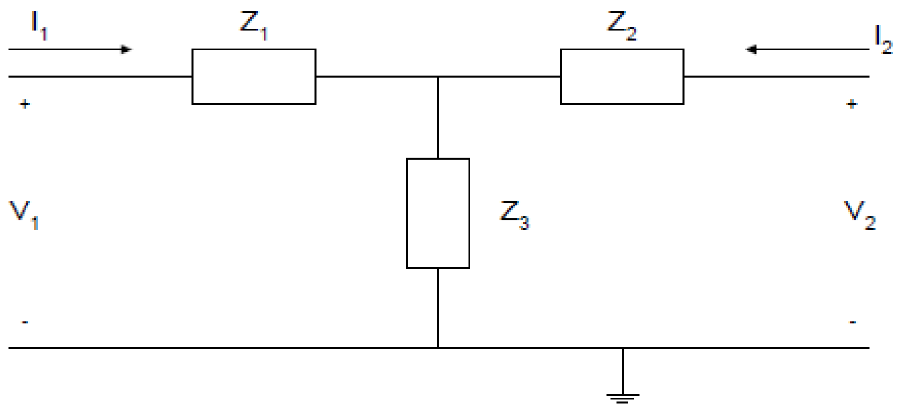


Figure 4.15.Block diagram of T-Network

Using KVL
$V_1 = Z_1 I_1 + Z_3\,(I_1 + I_2) = (Z_2 + Z_3) I_1 + Z_3 I_2$ (4.16)
$V_2 = Z_2 I_2 + Z_3\,(I_1 + I_2) = (Z_2) I_1 + (Z_2 + Z_3) I_2$ (4.17)
Thus

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_1 + Z_3 & Z_3 \\ Z_3 & Z_2 + Z_3 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \quad (4.18)$$

and the z-parameters are

$$[Z] = \begin{bmatrix} Z_1 + Z_3 & Z_3 \\ Z_3 & Z_2 + Z_3 \end{bmatrix} \quad (4.19)$$

3.7.2. Y-Parameters

A two-port network can also be represented using y-parameters. The describingequations are
$I_1 = y_{11} V_1 + y_{12} V_2$ (4.20)
$I_2 = y_{21} V_1 + y_{22} V_2$ (4.21)
Where
V_1 and V_2 are independent variables and
I_1 and I_2 are dependent variables.

In matrix form, the above equations can be rewritten as

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

The y-parameters can be found as follows:
$y_{11} = \frac{I_1}{V_1} | V_2 = 0$

$$y_{12} = \frac{I_1}{V_2} | V_1 = 0 \qquad (4.22)$$
$$y_{21} = \frac{I_2}{V_1} | V_2 = 0$$
$$y_{22} = \frac{I_2}{V_2} | V_1 = 0$$

The y-parameters are also called short-circuit admittance parameters. They are obtained as a ratio of current and voltage and the parameters are found by short-circuiting port 2 ($V_2 = 0$) or port 1 ($V_1 = 0$). The following two examples show how to obtain the y-parameters of simple circuits.

Example 2
Find the y-parameters of the pi (π) network shown in Figure 4.16.

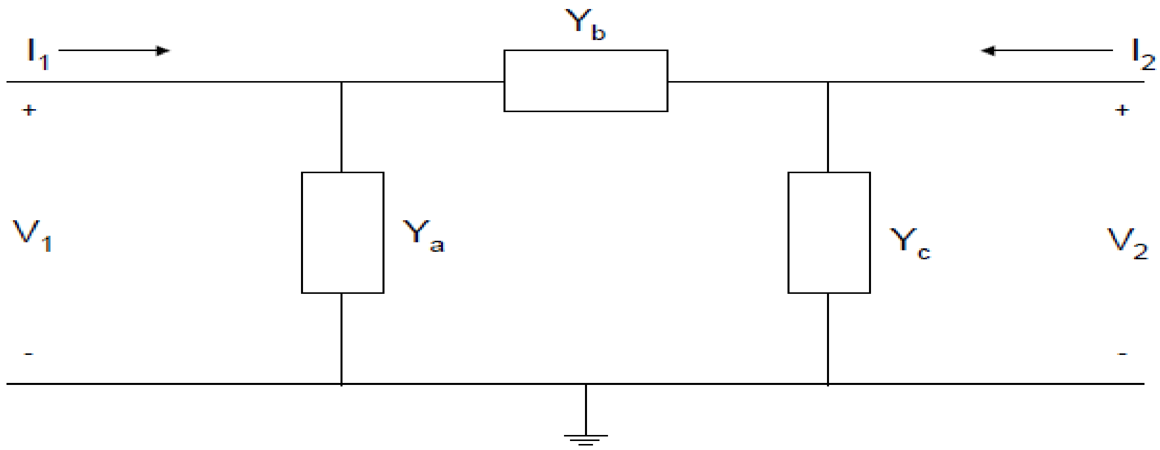


Figure 4.16. Pi-Network.

Solution
Using KCL, we have

$$I_1 = V_1 y_a + (V_1 - V_2) y_b = V_1 (y_a + y_b) - V_2 y_b \qquad (4.23)$$
$$I_2 = V_2 y_c + (V_2 - V_1) y_b = -V_1 y_b + V_2 (y_b + y_c) \qquad (4.24)$$

Comparing Equ. (4.23) and (4.24) to Equ. (4.20) and (4.21), the y parameters are

$$[Y] = \begin{bmatrix} Y_a + Y_b & -Y_b \\ -Y_b & Y_b + Y_c \end{bmatrix} \qquad (4.25)$$

Example 3
Figure 13 shows the simplified model of a field effect transistor. Find its yparameters.

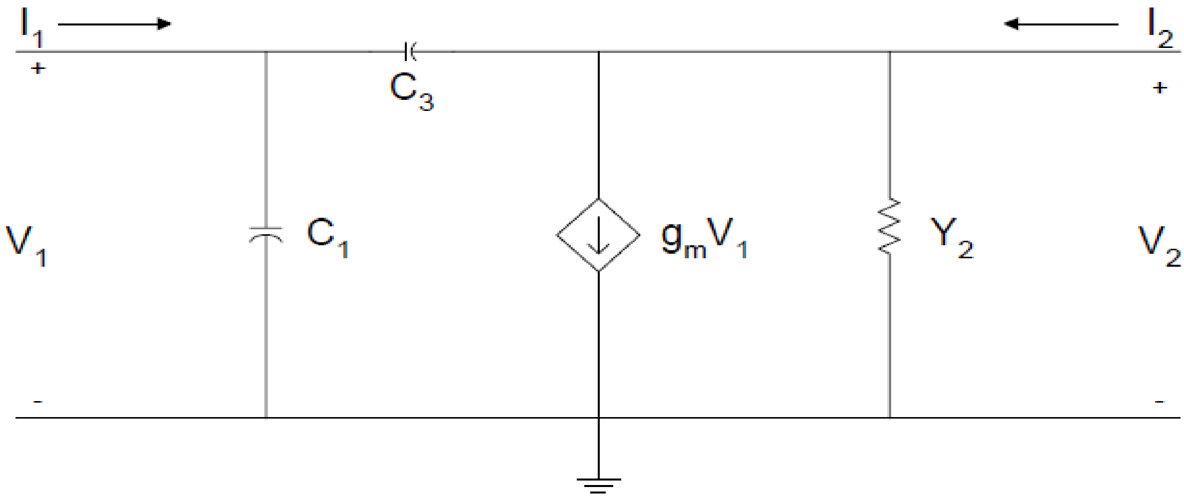


Figure 4.17.Simplified Model of a Field Effect Transistor

Using KCL,
Comparing the above two equations (4.20) and (4.21), the y-parameters are given by

$$I_1 = V_1 sC_1 + (V_1 - V_2)sC_3 = V_1(sC_1 + sC_3) - V_2(-sC_3) \tag{4.26}$$

$$I_2 = V_2 Y_2 + g_m V_1 + (V_2 - V_1)sC_3 = V_1(g_m - sC_3) + V_2(Y_2 + sC_3) \tag{4.27}$$

$$[Y] = \begin{bmatrix} sC_1 + sC_3 & -sC_3 \\ g_m - sC_3 & Y_2 + sC_3 \end{bmatrix} \tag{4.28}$$

3.7.3. H-parameters

A two-port network can be represented using the h-parameters. The describingequations for the h-parameters are given I Equ. 4.29 and 4.30.

$$V_1 = h_{11} I_1 + h_{12} V_2 \tag{4.29}$$

$$I_2 = h_{21} I_1 + h_{22} V_2 \tag{4.30}$$

Where
I_1 and V_2 are independent variables and
V_1 and I_2 are dependent variables.
In matrix form, the above two equations become Equ 4.31.

$$\begin{bmatrix} V_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{12} \end{bmatrix} \begin{bmatrix} I_1 \\ V_2 \end{bmatrix} \tag{4.31}$$

The h-parameters can be found as follows:

$$h_{11} = \frac{V_1}{I_1} | V_2 = 0 \tag{4.32}$$

$$h_{12} = \frac{V_1}{V_2} | I_1 = 0$$

$$h_{21} = \frac{I_2}{I_1} | V_2 = 0$$

$$h_{22} = \frac{I_2}{V_2} | I_1 = 0$$

The h-parameters are also called hybrid parameters since they contain bothopen-circuit parameters ($I_1 = 0$) and short-circuit parameters ($V_2 = 0$). Theh-parameters of a bipolar junction transistor are determined in the example below

Example 4
A simplified equivalent circuit of a bipolar junction transistor is shown in Figure14, find its h-parameters.

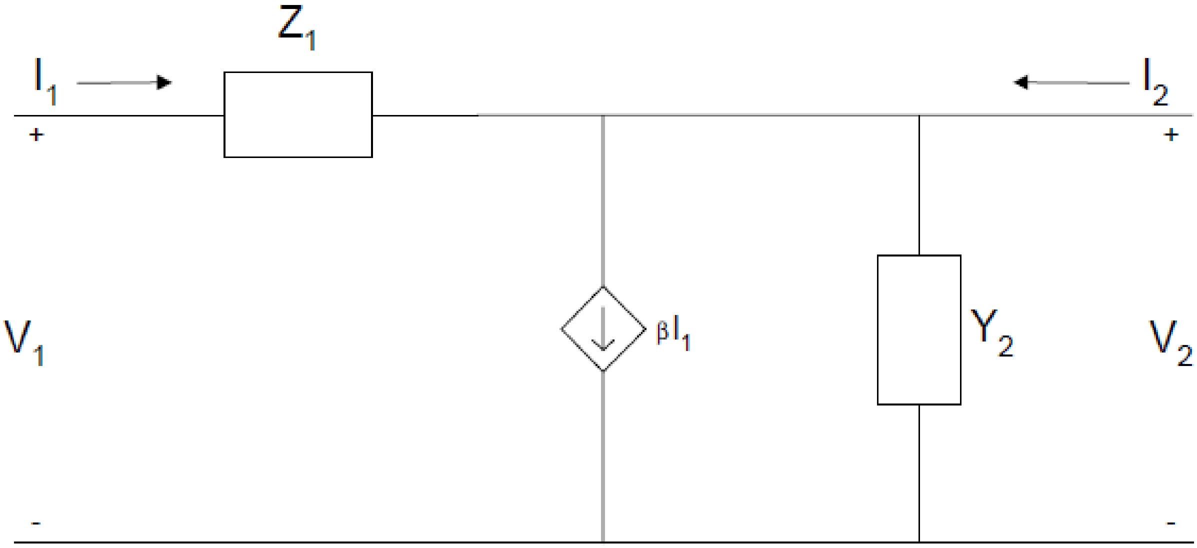


Figure 4.18.Simplified Equivalent Circuit of a Bipolar Junction Transistor

Solution
Using KCL for port 1,
$V_1 = I_1 Z_1$ (4.33)
Using KCL at port 2, we get,
$I_2 = \beta I_1 + Y_2 V_2$ (4.34)
Comparing the above two equations to Equations (4.29) and (4.30) we get the h-parameters.
$[h] = \begin{bmatrix} Z_1 & 0 \\ \beta & Y_2 \end{bmatrix}$ (4.35)

3.7.4. Transmission parameters

A two-port network can be described by transmission parameters. The describingequations are
$V_1 = a_{11}V_2 - a_{12}I_2$ (4.36)
$I_1 = a_{21}V_2 - a_{22}I_2$ (4.37)
Where
V_2 and I_2 are independent variables and
V_1 and I_1 are dependent variables.

In matrix form, the above two equations can be rewritten as

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{12} \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix}$$

The transmission parameters can be found as

$a_{11} = \frac{V_1}{V_2} | I_2 = 0$

$a_{12} = -\frac{V_1}{I_2} | V_2 = 0$

$a_{21} = \frac{I_1}{V_2} | I_2 = 0$

$a_{22} = -\frac{I_1}{I_2} | V_2 = 0$

The transmission parameters express the primary (sending end) variables $V_1$and I_1 in terms of the secondary (receiving end) variables V_2 and - I_2 . Thenegative of I_2 is used to allow the current to enter the load at the receivingend. Examples5 show some techniques for obtaining the transmissionparameters of impedance and admittance networks.

Example 5

Find the transmission parameters of Figure 4.19.

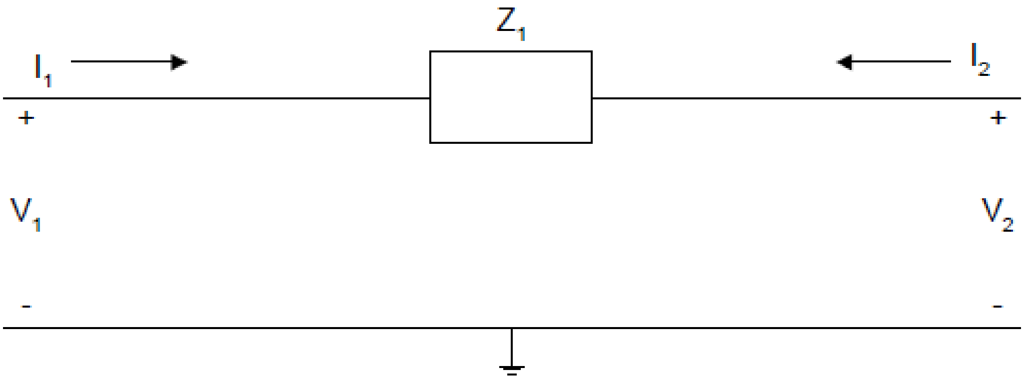


Figure 4.19. Simple Impedance Network

Solution

By inspection,

$I_1 = -I_2$ (4.38)

Using KVL,

$V_1 = V_2 + Z_1 I_1$ (7.43)

Since $I_1 = -I_2$, Equ (4.39) becomes

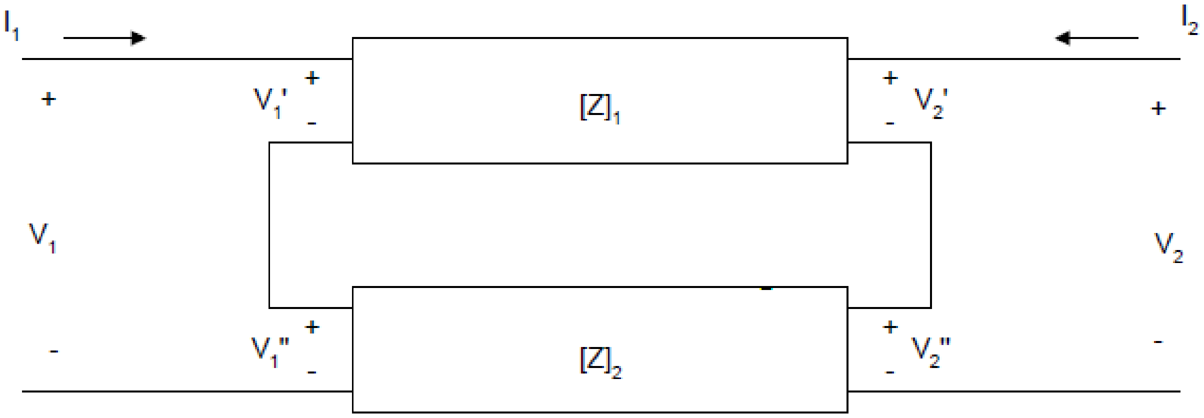
$V_1 = V_2 - Z_1 I_2$ (4.40)

Comparing Equations (4.39) and (4.40) to Equations (4.29) and (4.30), wehave

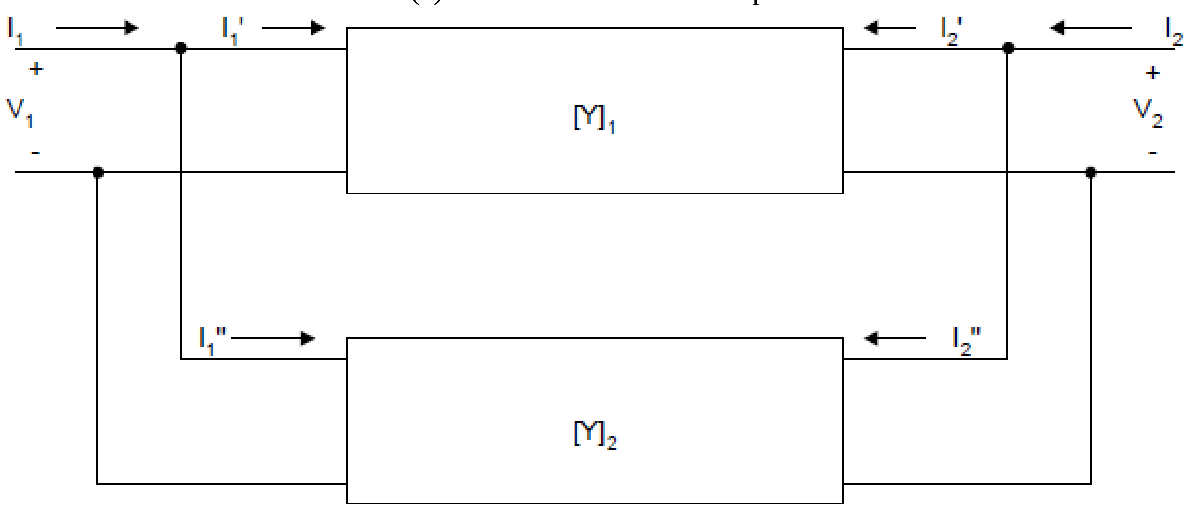
a_{11}=1 ; a_{12}=1 ; a_{21}=0 ; a_{22}= 1 (4.41)

3.8. Interconnection Of Two-Port Networks

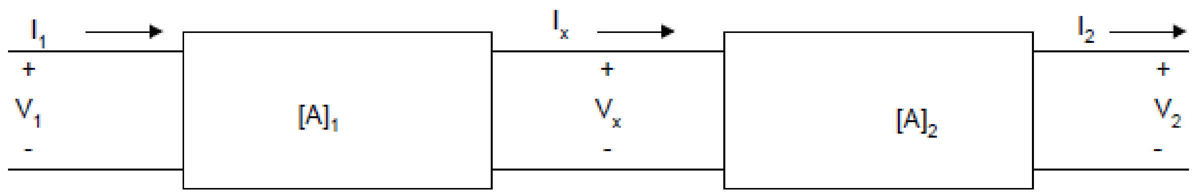
So far, we have seen various configurations of the two port network. However, two-port networks can be connected in series, parallel or cascade. Figure4.20 shows the various two-port interconnections.



(a) Series-connected Two-port Network



(b) Parallel-connected Two-port Network



(c) Cascade Connection of Two-port Network

Figure 4.20. Interconnection of Two-port Networks (a) Series (b) Parallel (c) Cascade

It can be shown that if two-port networks with z-parameters$[Z]_1$,$[Z]_2$,$[Z]_3$, ………………., $[Z]_n$are connected in series, then the equivalent twoportz-parameters are given as

$[Z]_{eq}$ = $[Z]_1 + [Z]_2$,$+[Z]_3$, $+$ ……+. $[Z]_n$ (4.42)

If two-port networks with y-parameters$[Y]_1$,$[Y]_2$,$[Y]_3$, ………………., $[Y]_n$..., , are connected in parallel, then the equivalent two-port y-parameters are given as

$[Y]_{eq}$ = $[Y]_1 + [Y]_2$,$+[Y]_3$, $+$ ……+. $[Y]_n$ (4.43)

When several two-port networks are connected in cascade, and the individualnetworks have transmission parameters $[A]_1$ = $[A]_2 + [A]_3$……+. $[A]_n$,then the equivalent two-port parameter will have a transmission parameter given as

$$[A]_{eq} = [A]_1 * [A]_2 * [A]_3 * \ldots\ldots.. * [A]_n \tag{4.44}$$

We shall now use an example to illustrate the determination of the equivalent parameters of interconnected two-port networks.

Example 6

Find the equivalent y-parameters for the bridge T-network shown in Figure4.21

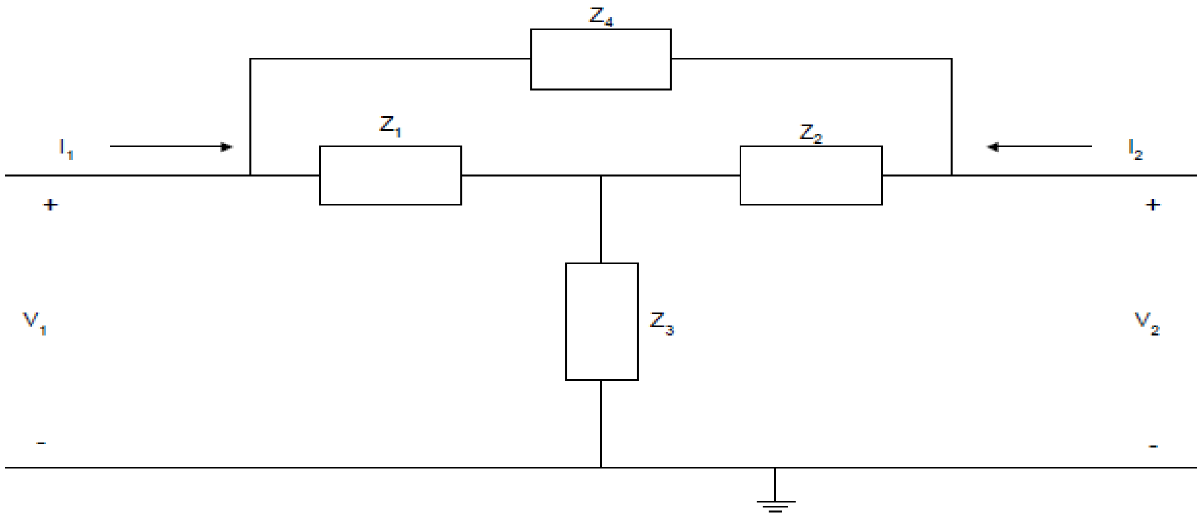


Figure 4.21.Bridge-T Network

Solution

The bridge-T network can be redrawn as shown in Figure 4.22.

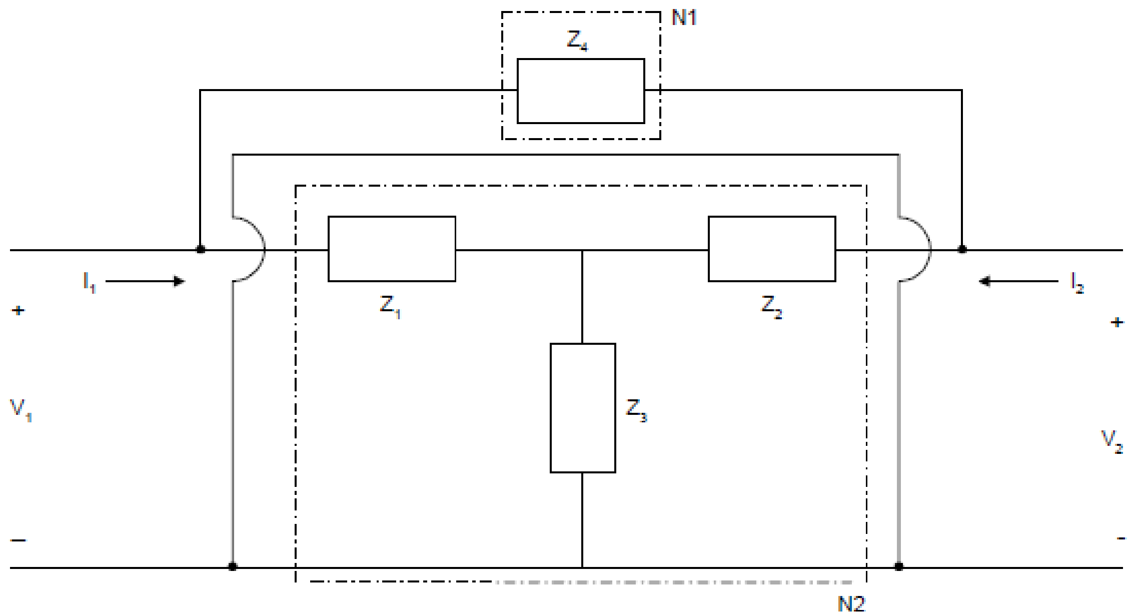


Figure 4.22. An Alternative Representation of Bridge-T Network

From the previous cases in Example 1, the z-parameters of network N_2 are

$$[z] = \begin{bmatrix} Z_1 + Z_2 & Z_3 \\ Z_3 & Z_2 + Z_3 \end{bmatrix}$$

We can convert the z-parameters to y-parameters and we get

$$y_{11} = \frac{Z_2+Z_3}{Z_1Z_2+Z_1Z_3+Z_2Z_3} \quad (4.45)$$

$$y_{12} = \frac{-Z_3}{Z_1Z_2+Z_1Z_3+Z_2Z_3} \quad (4.46)$$

$$y_{21} = \frac{-Z_3}{Z_1Z_2+Z_1Z_3+Z_2Z_3} \quad (4.46)$$

$$y_{22} = -\frac{Z_1+Z_3}{Z_1Z_2+Z_1Z_3+Z_2Z_3} \quad (4.47)$$

From Example 5, the transmission parameters of network $N_1$are

$$a_{11} = 1\ ;\ a_{12} = Z_4; a_{12} = Z_4\ ;\ a_{21} = 0; a_{21} = 1.$$

We convert the transmission parameters to y-parametersand we get

$$y_{11} = \frac{1}{Z_4} \quad (4.48)$$

$$y_{12} = -\frac{1}{Z_4} \quad (4.49)$$

$$y_{21} = -\frac{1}{Z_4} \quad (4.50)$$

$$y_{22} = \frac{1}{Z_4} \quad (4.51)$$

Using Equ. (4.48) to (4.51) the equivalent y-parameters of the bridge-T network, we now have

$$y_{11_{eq}} = \frac{1}{Z_4} + \frac{Z_2+Z_3}{Z_1Z_2+Z_1Z_3+Z_2Z_3}$$

$$y_{12_{eq}} = -\frac{1}{Z_4} - \frac{Z_3}{Z_1Z_2+Z_1Z_3+Z_2Z_3}$$

$$y_{21_{eq}} = -\frac{1}{Z_4} - \frac{Z_3}{Z_1Z_2+Z_1Z_3+Z_2Z_3}$$

$$y_{22_{eq}} = \frac{1}{Z_4} + \frac{Z_1+Z_3}{Z_1Z_2+Z_1Z_3+Z_2Z_3}$$

5.0 Conclusion

High impedance network is very important in communication systems. A good understand of circuit impedance as well as the different types of impedance elements have been explained. In communication systems, we have analyzed the reason for high input and low output impedance. Transmission line topologies, characteristics impedance and propagation constant in communication network have been simplified.

The various models of two port network and the internetworked two port network system has been well established while deriving mathematical models for two port networksproblems

6.0 Summary

- A two-port network is an electrical network (circuit) or device with two *pairs* of terminals to connect to external circuits.
- Two terminals constitute a port if the currents applied to them satisfy the essential requirement known as the port condition: the electric current entering one terminal must equal the current emerging from the other terminal on the same port.
- The ports constitute interfaces where the network connects to other networks, the points where signals are applied or outputs are taken.
- In a two-port network, often port 1 is considered the input port and port 2 is considered the output port.
- Examples of circuits analyzed as two-ports are filters, matching networks, transmission lines, transformers, and small-signal models for transistors (such as the hybrid-pi model). The analysis of passive two-port networks is an outgrowth of reciprocity theorems first derived by Lorentz.

6.0 Tutor-Marked Assignment (TMA)

- Define the concept of Impedance in a communication system.If ina series circuit, resistances and reactances add together independently, suppose a resistance of 120.00 ohms is connected in a series circuit with an inductance of 100.000H? At 54MHz, show the complex impedance. Give the impedance when a capacitor of 0.0010000F is introduced.
- Explain the objective of high input and low output impedance in a circuit.
- Briefly describe Characteristics Impedance and insertion loss.
- List 5 examples of typical transmission lines where impedance effects are visible.
- With a well labeled diagram, show a:

- Series-connected Two-port Network
- Parallel-connected Two-port Network
- Cascade Connection of Two-port Network

7.0 References/Further Readings

[1]. Biran, A. and Breiner, M. (1995), MATLAB for Engineers, Addison-Wesley,
[2]. Etter, D.M., Engineering Problem Solving with MATLAB, 2nd Edition,Prentice Hall, 1997.
[3]. MathWorks, Inc., (1995)MATLAB, High-Performance Numeric Computation Software,.
[4]. Nilsson, J.W., (1990),Electric Circuits, 3rd Edition, Addison-Wesley Publishing Company,.
[5]. Meader, D.A. (1991), Laplace Circuit Analysis and Active Filters,Prentice Hall,
[6]. Dennis Roddy, John Coolen, (2006). Electronic Communications, 4th Edition, Prentice-Hall of India.
[7]. Carlin, HJ, Civalleri, PP, (1998) Wideband circuit design, CRC Press, ISBN 0-8493-7897-4.

[8]. William F. Egan, (2003), Practical RF system design, Wiley-IEEE, ISBN 0-471-20023-9.
[9]. Farago, PS,(1961.)An Introduction to Linear Network Analysis, The English Universities Press Ltd,
[10]. Gray, P.R.; Hurst, P.J.; Lewis, S.H.; Meyer, R.G. (2001). Analysis and Design of Analog Integrated Circuits (4th ed.). New York: Wiley. ISBN 0-471-32168-0.
[11]. Ghosh, Smarajit, Network Theory: Analysis and Synthesis, Prentice Hall of India ISBN 81-203-2638-5.
[12]. Jaeger, R.C.; Blalock, T.N. (2006). Microelectronic Circuit Design (3rd ed.). Boston: McGraw–Hill. ISBN 978-0-07-319163-8.
[13]. Matthaei, Young, Jones, (1964)Microwave Filters, Impedance-Matching Networks, and Coupling Structures, McGraw-Hill..
[14]. Mahmood Nahvi, Joseph Edminister (2002), Schaum's outline of theory and problems of electric circuits, McGraw-Hill Professional, ISBN 0-07-139307-2.
[15]. Dragica Vasileska, Stephen Marshall Goodnick, Computational electronics, Morgan & Claypool Publishers, 2006 ISBN 1-59829-056-8.
[16]. Clayton R. Paul, (2008),Analysis of Multiconductor Transmission Lines, John Wiley & Sons, ISBN 0470131543, 9780470131541.
[17]. Greg.N.Onoh (2001). Principles of Telecommunications. 2nd Edition, ISBN 978-036-029-8.
[18]. https://en.wikipedia.org/wiki/Two-port_network#General_properties
[19]. https://en.wikipedia.org/wiki/Transmission_line

Module 4 Demodulation systems in Electronic Communication
UNIT 1 AM DETECTION AND DEMODULATION STRATEGIES

1.0 Introduction
2.0 Objectives
3.0 Main Content
3.1 What is Modulation?
3.2 Classification of Detection devices
3.2.1. Square Law Diode Detector
3.2.2. Linear Diode Detector
3.3 AM Demodulation Methods
3.3.1 Envelope detector
3.3.1.1. Envelope detector advantages
3.3.1.2. Envelope detector disadvantages
3.3.2 Product Detector
3.3.2.1.Advantages of Product Detector
3.3.2.2. Drawbacks of Product Detector
3.3.2.3. Simple Product Detector Mathematical Model
3.4 Time RC Constant in linear Diode Detector
3.5 RC Time constant analytical derivation
3.6 Distortion in Linear Diode Detector
3.7. Linear Diode Detector with filter
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment (TMA)
7.0 References/Further Readings

1.0 INTRODUCTION

We want to establish here that in the Unit 1 of Module 1, we did not properly highlight the limitations of AM. We shall look at this constraint in this module. Now, one disadvantage of all amplitude modulation techniques (not only standard AM) is that the receiver amplifies and detects noise and electromagnetic interference in equal proportion to the signal. For this reason, AM broadcast is not favored for music and high fidelity broadcasting, but rather for voice communications and broadcasts (sports, news, talk radio etc.). Another disadvantage of AM is that it is inefficient in power usage; at least two-thirds of the power is concentrated in the carrier signal. The carrier signal contains none of the original information being transmitted (voice, video, data, etc.).

However, its presence provides a simple means of demodulation using envelope detection, providing a frequency and phase reference to extract the modulation from the sidebands. In some modulation systems based on AM, a lower transmitter power is required through partial or total elimination of the carrier component, however receivers for these signals are more complex and costly. The receiver may regenerate a copy of the carrier frequency (usually as shifted to the intermediate frequency) from a greatly reduced pilot carrier (in reduced-carrier transmission or DSB-RC) to use in the demodulation process. Even with the carrier totally eliminated in double-sideband suppressed-carrier transmission, carrier regeneration is possible. However, this doesn't work for single-sideband suppressed-carrier transmission (SSB-SC).But Single sideband is nevertheless used widely in amateur radio and other voice communications both due to its power efficiency and bandwidth efficiency (cutting the RF bandwidth in half compared to standard AM).

On the other hand, in medium wave and short wave broadcasting, standard AM with the full carrier allows for reception using inexpensive receivers. The broadcaster absorbs the extra power cost to greatly increase potential audience.

In Section 3.1, we shall elaborate the exact meaning of demodulation as applied to communication systems in general. You will learn the various classifications of detection methods such as square law diodes, and linear diode detectors. You will be exposed to sources of distortion in linear diode detector as well as the choice of time constant in linear diode detector among other things.

2.0 Objectives

After going through this unit, you should be able to

- ✓ explain the concept of demodulation with respect to general communication systems.
- ✓ distinguish Square Law Diode Detector from Linear Diode Detector
- ✓ derive RC Time Constant analytically.
- ✓ explain Linear Diode Detector with filter
- ✓ describe distortion in Linear Diode Detector

3.0 Main Content

3.1. What is Demodulation?

In other to answer this question, you must understand that the process of detection or demodulation consists in recovering the original modulating voltage from the modulated carrier voltage. Demodulation is extracting the original information-bearing signal from a modulated carrier wave. A demodulator is an electronic circuit that is used to recover the information content from the modulated carrier wave. There are many types of modulation so there are many types of demodulators. The signal output from a demodulator may represent sound (an analog audio signal), images (an analog video signal) or binary data (a digital signal).These terms are traditionally used in connection with radio receivers, but many other systems use many kinds of demodulators. For example, in a modem, which is a contraction of the terms modulator/demodulator, a demodulator is used to extract a serial digital data stream from a carrier signal which is used to carry it through a telephone line, coaxial cable, or optical fiber.

Therefore, detection is a process reverse of the modulation. The detection process is accomplished by mixing the carrier with the sideband components carrying the intelligence, in a non-linear device. This mixing process results in sum and different frequency terms. Each sideband beats with the carrier to produce the different term which is nothing but the original modulation frequency term.

As far as the mixing operation is concerned, the detection process is similar to the modulation process. Accordingly, similar circuits may be used for both modulation and detection purpose except for certain differences in the output circuits for selection etc. thus in modulation process, the carrier voltage and the audio frequency modulation voltage are mixed in the modulated amplifier to produce the sum and the different frequency terms and the output circuit is tuned to the carrier frequency so that it responds to the carrier sideband terms and rejects all other frequency terms.

There are several ways of demodulation depending on how parameters of the base-band signal such as amplitude, frequency or phase are transmitted in the carrier signal. For example, for a signal modulated with a linear modulation like AM (amplitude modulation), we can use a synchronous detector. On the other hand, for a signal modulated with an angular modulation, we must use an FM (frequency modulation) demodulator or a PM (phase modulation) demodulator. Different kinds of circuits perform these functions.

Many techniques such as carrier recovery, clock recovery, bit slip, frame synchronization, rake receiver, pulse compression, Received Signal Strength Indication, error detection and correction, etc., are only performed by demodulators, although any specific demodulator may perform only some or none of these techniques.Many things can act as a demodulator, if they pass the radio waves on nonlinearly.
Let us pause for a while and ask ourselves: What the different types of demodulation or detection techniques? We shall now explain the various types in Section 3.2. Owing to the complexities of the mathematical models associated with this topic, we shall limit this module to Unit contents only.

3.2. Classification of Detection devices
In all forms of detection models, there are two functional classifications or types of detectors. These are:

i. Square law diode detectors
ii. Linear diode detectors.

In the first instance, the square law diode detector utilizes non-linear (approximately square law) portion of dynamic current voltage characteristic of an electron device. In square law detector, the current through the detecting device flow continuously and hence the behavior of the same may be conveniently analyzed using a power series representation.
On the other hand, the linear diode detectors use the rectification property of a diode. In this case, the current in this rectifying diode flows in pulse. The behavior of such a circuit, therefore, cannot be analyzed using the conventional power series representation. A linear relation, however, exists between the detected output voltage and the carrier amplitude.
In the previous discussion, we studied the two types of detection systems found in any communication model. In context, demodulation is essentially a reverse phenomenon of modulation existing at the receiving end of the communication system. We have outlined the two types of detection systems. Let us now study these two types of detection schemes.

3.2.1. Square Law Diode Detector
This type normally utilizes the non-linear portion of the dynamic-voltage characteristic of a diode. Remember that every diode has non-linear dynamic characteristics. How the square law type differs from the linear diode detector is that in this case the applied input carrier voltage is of small magnitude and hence is restricted to the excessively non-linear portion of the dynamic characteristic, whereas in linear diode detector, a large amplitude modulated carrier voltage is applied to the diode and most of the operation takes place over the linear region of the characteristic.

Figure 1.1 demonstrates the basic circuit of square law diode detector. The diode is biased positively to shift the zero-signal operating point to the small current nonlinear region of the dynamic current-voltage characteristic as shown in Figure. 1.2 (a). The capacitor- resistor combination constitutes the load.
Interestingly, to study the operation of this detector, we may consider first, the only resistor R to constitute the load impedance.
Then, Figure 1.2 (a) gives the dynamic current- voltage characteristic of the diode. Superposition of modulated carrier voltage on the dynamic characteristic is also illustrated in Figure. 1.2 (a). This results in the output current wave form as shown in Figure 1.2 (b).
Since the operation takes places over the nonlinear region of the characteristic, the current waveform has its lower half compressed. This results in the average value of current as shown in Figure. 1.2 (b). This average current consists of a steady or d.c component I_{dc} and a time varying component at the modulation frequency. The shunt capacitor C by passes all the radio frequency components leaving only the average component to flow through the load resistor R producing the desired detected-output.

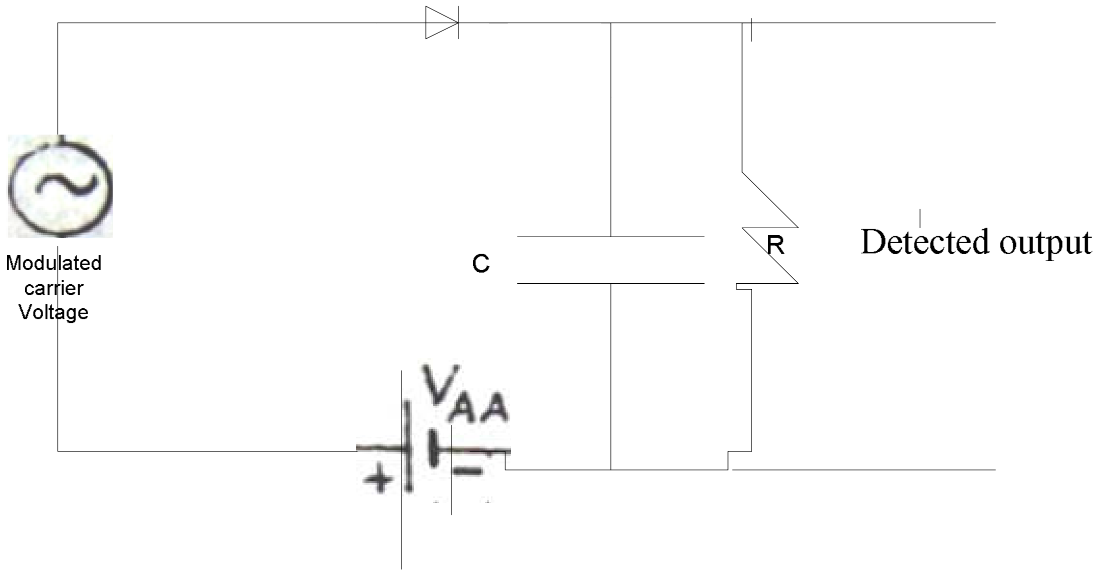


Figure 1.1. Basic circuit of square law diode detector

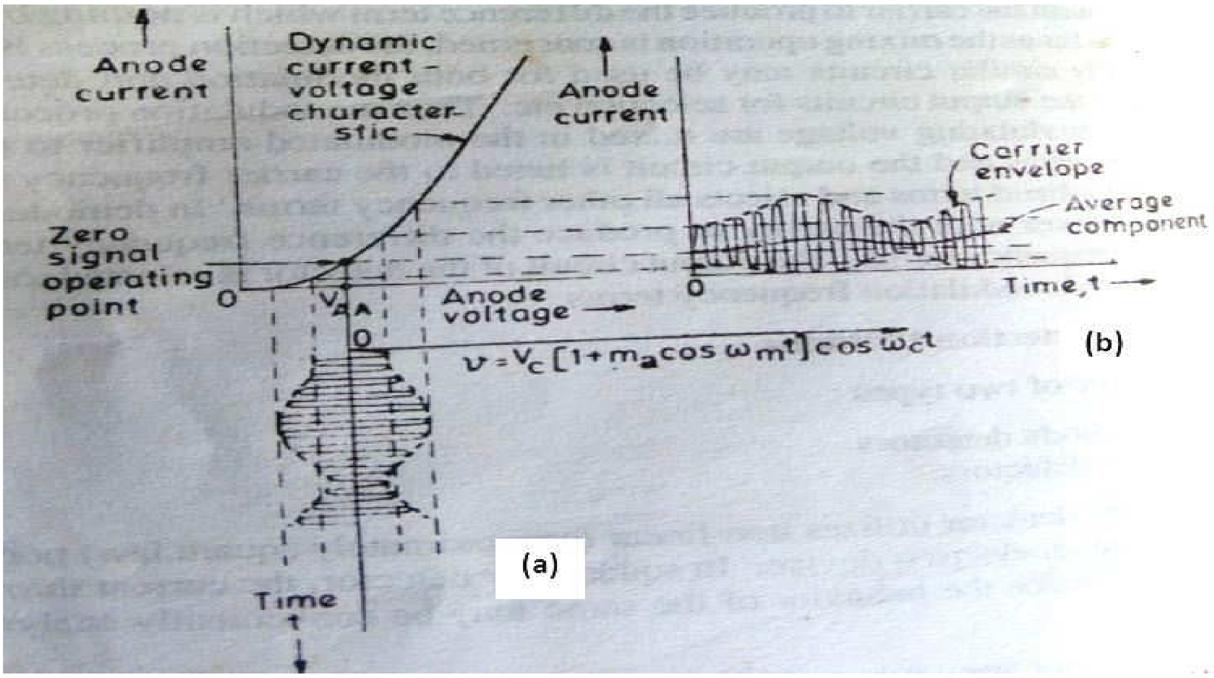


Figure 1.2. Graphical illustration of detection process in square law diode detector.

Figure 1.2(a) shows the superposition of modulated carrier voltage while Figure 1.2(b) shows the output current of diode detector with over the dynamic characteristic of diode resistance load. It worthy to mention that the dynamic current-voltage characteristic obeys approximately the square-law relation as shown below.

$$i_a = a_1.v_a + a_2.v_a^2 \tag{1.1}$$

Where i_a is a.c. anode current? and anode voltage given by

$$v_a = V_c\ [1 + m_a\ cos\omega_c t] \tag{1.2}$$

and contains terms in frequencies $\omega_,$ $(\omega_c + \omega_m)$ and $(\omega_c - \omega_m)$.
Evidently the second term in Equ. (1.1) gives term in frequencies $2\omega c_, 2(\omega_c + \omega_m)$ $2(\omega_c - \omega_m)\omega_mand\omega_{m.}$the R.F. terms bypassed through shunt capacitor C whereas terms in frequencies $\omega_m$and2_{ω_m} are developed across the load resistor R. the term in frequency $\omega_m$constitutes the desired output whereas the term in frequency 2_{ω_m} forms the distortion term.
Having established the anode voltage in Equ. 1.2, lets now look at the linear diode detector.

3.2.2. Linear Diode Detector

It might interest you to know that the linear diode detectors are popularly used in commercial radio receivers. However, a linear diode detector for satisfactory operation requires modulated carrier voltages of 5volts or more. With such high carrier voltage, the cut-in voltage V_v of the diode may be neglected and the operation may be considered to be taking place over essentially linear region of the dynamic current- voltage characteristic of the diode. With small carrier voltage, operation takes place over the excessively nonlinear region of the characteristic giving square law detection with associated distortion.

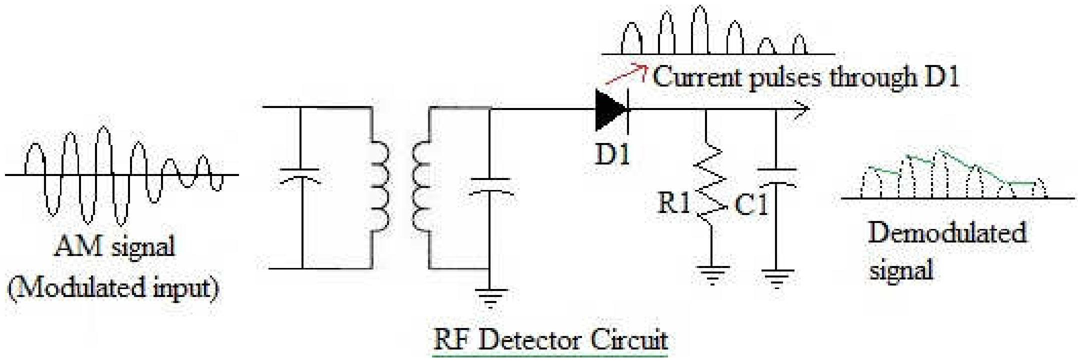


Figure 1.3. Operation of linear diode detector.

There is a superposition of modulated carrier voltage on the linear $i_a - v_a$ characteristic of a diode while generating an output voltage of linear diode detector with capacitor C.

Linear detector utilizes the rectification characteristic of a diode.Figure 1.4 gives the idealized linear dynamic current-voltage characteristic of the diode detector. The modulated carrier voltage is applied to the series combination of diode and the load impedance consisting of resistor R in shunt with capacitor C. Since applied voltage is of large magnitude, the operation takes place essentially over the linear region of the dynamic current-voltage characteristic of the diode. At this point, it is important we look at the envelop detector.

3.3. AM Demodulation Methods

You will recall that an AM signal encodes the information onto the carrier wave by varying its amplitude in direct sympathy with the analogue signal to be sent. In this regard, there are two methods used to demodulate AM signals:

3.3.1. Envelope detector

AM signal demodulation using envelope detector is one method. Now, an envelope detector is an electronic circuit that takes a high-frequency signal as input and provides an output which is the envelope of the original signal. The capacitor in the circuit stores up charge on the rising edge, and releases it slowly through the resistor when the signal falls. The diode in series rectifies the incoming signal, allowing current flow only when the positive input terminal is at a higher potential than the negative input terminal.

Most practical envelope detectors use either half-wave or full-wave rectification of the signal to convert the AC audio input into a pulsed DC signals. Filtering is then used to smooth the final result. This filtering is rarely perfect and some "ripple" is likely to remain on the envelope follower output, particularly for low frequency inputs such as notes from a bass guitar. More filtering gives a smoother result, but decreases the responsiveness; thus, real-world designs must be optimized for the application.For the diode detector diagram shown in Figure 5.4, this represents the simplest form of envelope detector.

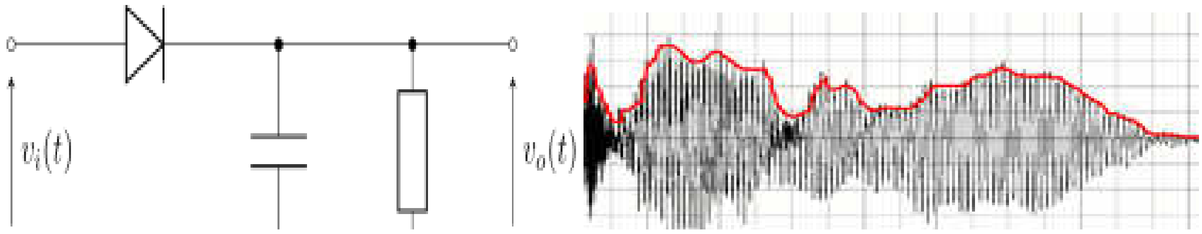


Figure 1.4. Simple envelop detector with demodulated output voltage

Now, a diode detector is simply a diode between the input and output of a circuit, connected to a resistor and capacitor in parallel from the output of the circuit to the ground. If the resistor and capacitor are correctly chosen, the output of this circuit should approximate a voltage-shifted version of the original (baseband) signal. A simple filter can then be applied to filter out the DC component.

You can now remember that the envelope detector is a very simple method of demodulation that does not require a coherent demodulator. It consists of an envelope detector that can be a rectifier (anything that will pass current in one direction only) or other non-linear component that enhances one half of the received signal over the other and a low-pass filter. The rectifier may be in the form of a single diode or may be more complex. Many natural substances exhibit this rectification behavior, which is why it was the earliest modulation and demodulation technique used in radio. The filter is usually an RC low-pass type but the filter function can sometimes be achieved by relying on the limited frequency response of the circuitry following the rectifier. The crystal set exploits the simplicity of AM modulation to produce a receiver with very few parts, using the crystal as the rectifier and the limited frequency response of the headphones as the filter.

An envelope detector can be used to demodulate a previously modulated signal by removing all high frequency components of the signal. The capacitor and resistor form a low-pass filter to filter out the carrier frequency. Such a device is often used to demodulate AM radio signals because the envelope of the modulated signal is equivalent to the baseband signal.

Mathematically, any AM or FM signal $X(t)$ can be written in the following form

$$X(t) = R(t)Cos(\omega t + \varphi(t)) \qquad (1.3)$$

In the case of AM, $\varphi(t)$ (the phase component of the signal) is constant and can be ignored. In AM, the carrier frequency ω is also constant. Thus, all the information in the AM signal is in $R(t)$. $R(t)$ is called the envelope of the signal. Hence, an AM signal is given by the function

$$X(t) = (C + m(t))Cos(\omega t)) \qquad (1.4)$$

With $m(t)$ representing the original audio frequency message, C the carrier amplitude and $R(t)$ equal to $C + m(t)$. So, if the envelope of the AM signal can be extracted, the original message can be recovered.

In the case of FM, the transmitted signal has a constant envelope $R(t) = R$ and can be ignored. However, many FM receivers measure the envelope anyway for received signal strength indication.

We shall look at the advantages and disadvantages of a diode envelope detector below. We have established previously that the AM diode envelope detector has been available for many years and has been widely used. Although amplitude modulation is used less these days, and other forms of AM detector can be easily incorporated into integrated circuits, the simple diode detector still has some advantages.

3.3.1.1. Envelope detector advantages:

- Low cost: The diode detector only requires the use of a few low cost components. This made it ideal for use in transistor (and valve / vacuum tube) radios using discrete components.
- Simplicity: Using very few components, the Diode AM detector was easy to implement. It was reliable and did not require any setup.

3.3.1.2. Envelope detector disadvantages:

- Distortion: As the diode detector is non-linear it introduces distortion onto the detected audio signal.
- Selective fading: One of the issues often experienced on the short and medium wavebands where the AM transmissions are located is that of selective fading. The diode envelope detector is not able to combat the effects of this in the way that some other detectors are able, and as a result, distortion occurs when selective fading occurs.
- Sensitivity: The diode detector is not as sensitive as some other types. If silicon diodes are used, these have a turn on voltage of around 0.6 volts as a result, germanium or Schottky diodes are used which have a lower turn on voltage of around 0.2 to 0.3 volts. Even with the use of the Schottky diode, the diode envelope detector still suffers from a poor level of sensitivity.

Other drawbacks of the envelope detector include:

- The input to the detector must be band-pass filtered around the desired signal, or else the detector will simultaneously demodulate several signals. The filtering can be done with a tunable filter or, more practically, a superheterodyne receiver
- It is more susceptible to noise than a product detector
- If the signal is overmodulated, distortion will occur

Most of these drawbacks are relatively minor and are usually acceptable tradeoffs for the simplicity and low cost of using an envelope detector.

3.3.2. Product Detector

It its simplest term, a product detector is a type of demodulator used for AM and SSB signals. Rather than converting the envelope of the signal into the decoded waveform like an envelope detector, the product detector takes the product of the modulated signal and a local oscillator, hence the name. A product detector is a frequency mixer.

Kindly note that the product detectors can be designed to accept either IF or RF frequency inputs. A product detector which accepts an IF signal would be used as a demodulator block in a superheterodyne receiver, and a detector designed for RF can be combined with an RF amplifier and a low-pass filter into a direct-conversion receiver.This type multiplies the incoming signal by the signal of a local oscillator with the same frequency and phase as the carrier of the incoming signal. After filtering, the original audio signal will result. You must note that the SSB is a form of AM in which the carrier is reduced or suppressed entirely, which require coherent demodulation. The simplest form of product detector mixes (or heterodynes) the RF or IF signal with a locally derived carrier (the Beat Frequency Oscillator- BFO) to produce an audio frequency copy of the original audio signal and a mixer product at twice the original RF or IF frequency. This high-frequency component can then be filtered out, leaving the original audio frequency signal.

3.3.2.1.Advantages of Product Detector

We shall now outline the merits and demerits of the product detector/demodulator scheme. Interestingly, the product demodulator has some advantages over an envelope detector for AM signal reception. Now,

- The product demodulator can decode overmodulated AM and AM with suppressed carrier.
- A signal demodulated with a product detector will have a higher signal to noise ratio than the same signal demodulated with an envelope detector.
- The envelope detector is a simple and relatively inexpensive circuit, and it can provide higher fidelity, since there is no possibility of mistuning the local oscillator. A product detector (or equivalent) is needed to demodulate SSB signals.

3.3.2.2. Drawbacks of Product Detector

Although this simple detector works, it has two major drawbacks:

- The frequency of the local oscillator must be the same as the frequency of the carrier, or else the output message will fade in and out in the case of AM, or be frequency shifted in the case of SSB
- Once the frequency is matched, the phase of carrier must be obtained, or else the demodulated message will be attenuated, but the noise will not be.
- The frequency of an AM carrier can be accurately determined with a phase-locked loop, but for SSB, the only solution is to construct a highly stable oscillator which has its own tradeoffs.

3.3.2.3.Simple Product Detector Mathematical Model

We can now express the product detector mathematically below.
Now, If $x(t) = \big(C + m(t)\big)Cos(\omega t)$.
Multiplying the AM signal $x(t)$ by an oscillator at the same frequency as and in phase with the carrier component yields

$$y(t) = \big(C + m(t)\big)Cos(\omega t)Cos(\omega t), \tag{1.5}$$

This can be re-written as

$$y(t) = \big(C + m(t)\big)(\tfrac{1}{2} + \tfrac{1}{2}Cos(2\omega t), \tag{1.6}$$

After filtering out the high-frequency component based around$Cos(2\omega t)$ and the DC component C, the original message will be recovered

It must be noted that a coherent demodulator is the AM/FM type used for recovering the phase modulation from an input intermediate frequency signal. This includes a phase locked loop having switchable frequency and phase detectors. During initial acquisition the phase locked loop frequency is adjusted to be equal to the input intermediate frequency by comparing the loop frequency with the input frequency in the frequency detector and using the resulting error signal to adjust the loop frequency.

After acquisition the phase detector maintains the loop frequency phase locked to the input frequency by now using the resulting error signal to adjust loop frequency phase, the error signal from the phase detector being the recovered phase modulation. A second control loop includes a narrow pass filter tuned to the desired frequency of the loop frequency. The error signal from the second loop is summed with the first mentioned error signal to precisely maintain loop frequency.

Coherent systems need carrier phase information at the receiver and they use matched filters to detect and decide what data was sent, while non-coherent systems do not need carrier phase information and use methods like square law to recover the data.

3.4. Time RC Constant in linear Diode Detector

From Section 3.2.2., one might wish to know the implication of keeping the time constant RC either too high or too low? Of course the time constant cannot be kept either too high or too low. If the time constant RC is very low, the discharges curve during the no conducting period is almost vertical resulting in large fluctuations in an output voltage. On the other hand, if the time constant RC is very large the discharge curve is almost horizontal and it then misses several peaks of the rectified output voltage during negative peaks as shown in Figure 1.5. Hence, with high time constant RC, the circuit cannot handle large depths of modulation and the signal gets clipped at the negative peaks.

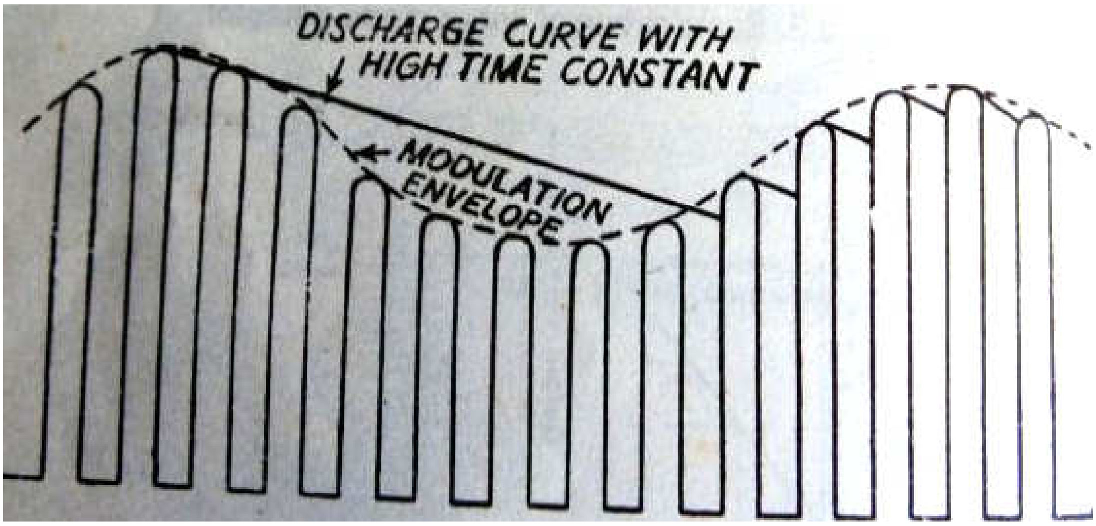


Figure. 1.5. Effect of high time constant on the performance of the linear diode detector.

We can now concluded that the time constant used should be as large as possible subject to the provision that the clipping during negative peaks of modulated voltage enveloped does not take place. The maximum allowable value of time constant is such that the rate of discharge of capacitor C is the same as the rate of decrease of the modulation envelope. This maximum possible value of the time constant may be determined analytically below.

3.5. RC Time Constant Analytical Derivation

Having understood the implication of the RC time constant in Section 3.4, we can now show the useful analytical model. But recall that the equation for the envelope of the modulated voltage is given by,

$$v = V_c\,[1 + m_a cos\omega_m t] \qquad (1.7)$$

Where$\omega_m$is the maximum modulating frequency permissible

The slope of the envelope is then given by,

$$\frac{dv}{dt} = -V_c\omega_m . m_a\ sin\omega_{m^t} \qquad (1.8)$$

At any time $t = \ t_o$the value of envelope voltage is given by,

$$V_o = V_c\,[1 + \omega_a \cos\omega_m t_o] \qquad (1.9)$$

and the corresponding slope is given by

$$\left(\frac{dv}{dt}\right)_{t=t_o} = \ -\omega_m m_a V_c \sin\omega_m t_o \qquad (1.10)$$

Let time $t_o$be the instant when the capacitor C just starts discharging through the resistor R.

Eventually at $t = \ t_o$, capacitor is charged to a voltage v_{k_o}equal to the envelope voltage.

Then, $v_{k_o = } v_o = V_c\,[1+m_a \cos\omega_m t_o]$ (1.11)

At any subsequent time t, the exponential decay of voltage $v_k$across the capacitor takes place according to the constant RC and the voltage v_k across capacitor C is given by,

$$v_k = v_{k_o}\epsilon^{-(t-t_o)IRC} \qquad (1.12)$$

The rate of change of $v_k$is given by,

$$\frac{d_{v_k}}{dt} = \ - \ v_{k_o}\frac{\epsilon-(t-t_o)IRC}{RC} \qquad (1.13)$$

At time t = $t_{o,}$the rate of change of voltage across the capacitor is given by,

$$\left(\frac{d_{v_k}}{dt}\right)_{t=t_o} = -\frac{1}{RC}v_{k_o} = -\frac{1}{RC}V_c\,[1 + m_a \cos\omega_m t_o] \qquad (1.14)$$

If clipping is to be avoided, at time $t = t_{o,}$the slope of capacitor voltage $v_k$should be algebraically equal to or less than the slope of envelope voltage.

Then,

$$-\ \frac{1}{RC}V_c\,[1 + m_a \cos w_m t_o] \leq -\ V_c\omega_m m_a sin\omega_m t_o \qquad (1.15)$$

$$\frac{1}{RC}V_c\,[1 + m_a \cos\omega_m t_o] \geq V_c\omega_m m_a \sin\omega_m t_o \qquad (1.16)$$

i.e., at time $t = \ t_{o,}$ the rate of decay of capacitor voltage should be equal to or greater than the rate of decay of envelope voltage.

$$\frac{1}{RC} \ \geq \ \omega_m \frac{m_a \sin\omega_m t_o}{1+m_a \cos\omega_m t_o} \qquad (1.17)$$

The most severe condition imposed on the value of time constant RC is when the quantity $m_a sin\omega_m t_0/(1 + m_a \cos\omega_m t_o)$ is a maximum. The condition under which this fraction is maximum is obtained by equating to zero the time rate change of this fraction i.e.

$$\frac{d}{dt}\left[\frac{m_a\ sin\omega_m t_o}{1+m_a \cos\omega_m t_o}\right] = 0 \qquad (1.18)$$

Equ. (1.18) yields,

$\cos \omega_{t_0} = -m_a$ (1.19)

$\sin \omega_{t_0} = \sqrt{1-m_a^2}$ (1.20)

Hence, Equ. (1.17) reduce to,

$$\frac{1}{RC} \geq \omega_m m_a \frac{\sqrt{1-m_a^2}}{1+m_a(-m_a)}$$

$$\frac{1}{RC} \geq \omega_m \frac{m_a}{\sqrt{1-m_a^2}} \quad (1.21)$$

As the modulation index m_a increase, the ratio $\frac{m_a}{\sqrt{1-m_a^2}}$ increase and hence requirement of RC becomes more sever. Thus, when the modulation approaches 100%, the ratio $\frac{m}{\sqrt{1-m_a^2}}$ approaches infinity. With zero time contain, carrier frequency components are not filtered out and then the output constants the modulation frequency as well as carrier frequency components.

The above analysis results in Equ. (11.21) which gives the maximum permissible value of time constant RC for no clipping. But the above analysis neglects many factors which appreciably modify the results. Consequently maximum permissible value of RC is based on empirical formula. Again, from experimental observation, if the amount of harmonic generation or distortion is not excessive, the maximum value of time constant RC should be selected in accordance with the relation.

$$\frac{1}{RC} \geq \omega_m m_a \, . \quad (1.22)$$

3.6. Distortion in Linear Diode Detector

We shall now look at the two sources of distortion in linear diode detector. So far, you can now see the implication of RC time constant on modulation index. While studying demodulation in the previous section, we have noticed distortion as an inherent feature of the linear diode detector. Now, one source of distortion in linear diode detector has already been discussed, namely distortion due to improper selection of time constant RC. If RC is too low, removal of radio frequency components is inadequate and the output voltage waveform is spiky in nature. On the other hand, if RC is too large, distortion may result due to clipping during the negative peaks of the modulation wave.

The second source of distortion in linear diode detector is the curvature of the diode characteristics. As a result of this curvature, the efficiency of rectification varies according to the amplitude of the envelope. Such a distortion may, however, be reduced by two means:

i. By selecting the value of load resistance large compared with the anode resistance of the detector diode and

ii. By applying the carrier voltage of large envelope amplitude. Under such conditions, the detection efficiency is large usually greater than 80% and the distortion due to curvature of diode characteristics is less than 2%. The carrier envelope amplitude reduces the distortion increase.

3.7. Linear Diode Detector with filter

Consider the circuit in Figures 1.1 and Figure 1.3,C passes all RF components of current and the output voltage developed across R contains very small RF components. To effect better elimination of RF voltage from the output, a π- type $R - C$ filter may be used as shown in Figure 1.7. Then the voltage across the load resistor R_L consists of the desired modulation frequency components plus a DC voltage proportional to the carrier amplitude. These two components are separated out. The DC components may be used for the automatic gain control (AGC) whereas the modulation frequency component is further amplifier in one or more stages of audio frequency amplifier. The modulation frequency voltage is coupled to the audio amplifier through a coupling capacitor C_0 which simultaneously eliminates the DC components from reaching the audio amplifier input. In the circuit of Figure1.7. The DC resistance offered to the detector diode is simply $(R_I + R_L)$. the AC resistance differs from this DC value because of the shunting effect of (i) filter capacitors C_I and C_2 and (ii) base circuit resistance R_b of the next stage.

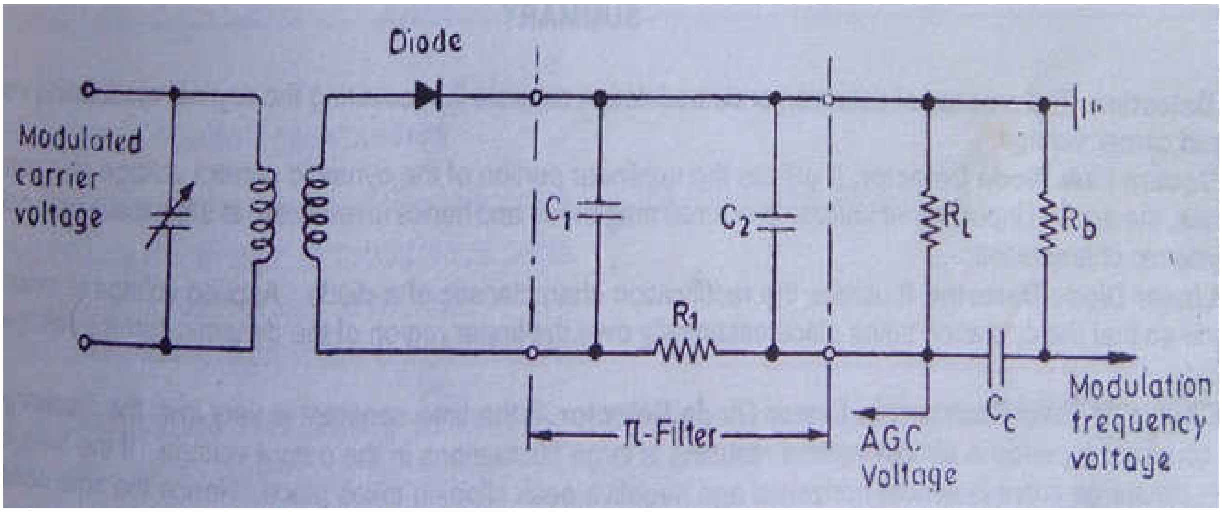


Figure 1.7. Linear diode detector with π-filter.

The values of resistances $R_{I,}R_L$ and R_b have to be carefully chosen in order that distortion does not take place at the maximum value of modulation index. Thus considering the specific case of sinusoidal modulation voltage, the diode output current consists of a DC component and a sinusoidal modulation frequency component.
In order to avoid negative peak clipping and the associated distortion the amplitude of modulation frequency diode current should not exceed the DC diode current i.e. $I \leq I_{d_c}$.
But the modulation index m_a is given by,

$$m_a = \frac{peak\ value\ of\ modulation\ frequency\ output\ voltage}{DC\ output\ voltage}$$

Or $m_a = \frac{V_m}{V_{d_c}}$ (1.23)

Or $m_a = \frac{I_m |Z_m|}{I_{d_c}.R}$ (1.24)

Where I_m is the peak value of modulation frequency diode current.

$Z_m$is the impedance of the diode output circuit at the frequency.
I_{d_c} is the diode current.
and R is the diode output circuit DC resistance.
To avoid negative peak clipping the maximum value $(m_a)_{m_{ax}}$of modulation index is such that I_m just equals I_{dc}.
Hence substituting, $i_m = I_{dc}$ in Equ. (5.10), we get

$$(m_a)_{m_{ax}} = \frac{|Z_m|}{R} \quad (1.25)$$

The component values are so chosen that at the modulation frequency, reactance of C_1 and C_2 are large and the reactance of coupling capacitor C_k is small. Hence impedance of the output circuit of detector at modulation frequency is given by,

$$|Z_m| = R_1 + \frac{R_L R_b}{R_L + R_b}$$

Hence, $$(m_a)_{max} = \frac{R_L + R_L R_b / (R_L + R_b)}{R_L + R_b} \quad (1.26)$$

The resistance *P* is usually a variable resistance and is used to adjust the value of detected output voltage. This provides the manual volume control in radio receiver using a linear diode detector.

4.0. Conclusion

In this course unit, we have explained and understood the practical theory behind demodulation with respect to general communication systems. Various types with respect to communication systems have been highlighted. We have clearly established the differences between Square Law Diode Detector and Linear Diode Detector. Analytical derivation of the RC Time Constant has been discussed. Also, we looked at distortion in Linear Diode Detector.

5.0 Summary

Demodulator- A demodulator is an electronic circuit that is used to recover the information content from the modulated carrier wave.

- Detection. The process of detection or demodulation consists in recovering the original modulating voltage from the modulated carrier voltage.
- Square Law Diode Detector. It utilizes the nonlinear portion of the dynamic current voltage characteristic a diode. In this case, the applied input carrier voltage is of small magnitude and hence is restricted to the excessively nonlinear portion of the dynamic characteristic.
- Linear Diode Detector. It utilizes the rectification characteristic of a diode. Applied voltage is maintained at large magnitude so that the operation takes place essentially over the linear region of the dynamic current-voltage characteristic of the diode.
- Choice of Time Constant in Linear Diode Detector. If the time constant is very low, the discharge curve during the non-conducting period is almost vertical resulting is large fluctuations in the output voltage. If the time constant is very large the discharge curve is almost horizontal and negative peak clipping takes place. Hence the time constant should be as large as possible but must avoid negative peak clipping. Mathematically,

$$\frac{1}{RC} \geq \omega_m \frac{m_a}{\sqrt{1-m_a^2}}$$

- Distortion in Linear Diode Detector. There are two sources of distortion in linear diode detector. (i) Improper selection of time constant RC (ii) curvature of the diode characteristics.
- Linear Diode Detector with π- filter results in better elimination of RF voltage from the output.

6.0 Tutor-Marked Assignment (TMA)

1. Define the process of detection. Describe the basic principle of detection.
2. Give the basic principle of square law diode detection. Draw the circuit of square law diode detector and describe how detection takes place.
3. What is meant by linear diode detection? Draw the circuit of a linear diode detector using a simple capacitor filter and describe graphically the detection process performed by this circuit.
4. Explain why the time constant RC of load circuit in linear diode detector cannot be kept too low or too high.
5. Derive expression for the maximum value of time constant RC in a linear diode detector so as to avoid negative peak clipping.
6. What are the principle sources of distortion in linear diode detector and how can distortion be kept low?
7. Draw the circuit and explain the working of a linear diode detector using π- filter. Discuss the considerations involved in the selection of resistive elements in the output circuit of this detector. Obtain an expression for the maximum value of modulation index which the detector can handle without distortion.

7.0 References/Further Readings

1. kennedy, G.(1977.), Electronic Communication Systems' 2nd edition McGraw Hill Kogukusha Ltd.
2. ryder, J.D.(1978), Electronic Fundamentals and Applications', 5th edition, Prentice Hall of India Pvt. Ltd, New Delhi,
3. De Vrijer, F.W.: Modulation', Phillips Technical Review, 36, pp 305-362, 1967, No. 11/12.
4. https://en.wikipedia.org/wiki/Envelope_detector Retrieved 3rd September,2017
5.https://www.electronics-notes.com/articles/radio/modulation/am-diode-detector-demodulator.php[accessed Sep 2, 2017].
6.What is the difference between coherent signals and non- coherent signals. Available from: https://www.researchgate.net/post/What_is_the_difference_between_coherent_signals_and_non_coherent_signals_in_a_simple_declaration [accessed Sep 2, 2017].