Companies

Electrical & Electronics Engineering Series

ELECTROMAGNETIC WAVES

R K SHEVGAONKAR

ELECTROMAGNETIC WAVES

About the Author

Dr. R K Shevgaonkar is Professor and Head of the Electrical Engineering Department, Indian Institute of Technology (IIT), Bombay. He was a scientist at University of Maryland, USA, Indian Institute of Astrophysics/Raman Research Institute, and Visiting Professor at University of Nebraska, Lincoln, USA, and ETH, Zurich, Switzerland. His research interests include Antennas, Linear and Non-Linear Fiber Optics, Numerical Electromagnetics, Radio Astronomy and Digital Image Processing. His teaching and research experience spans over 30 years. He has published more than 125 papers in international journals and conferences, and co-authored a book on Transmission Lines.

He is Fellow of the Indian Academy of Engineering, the National Science Academy of India, the Institution of Electronics and Telecommunication Engineers, senior member of IEEE, and member of many professional bodies like International Astronomical Society, Optical Society of India and Indian Astronomical Society. He has received numerous awards for his scientific achievements including the 'Excellence in Teaching' award at IIT, Bombay.

ELECTROMAGNETIC WAVES

R K Shevgaonkar

Professor and Head
Department of Electrical Engineering
Indian Institute of Technology
Bombay

Tata McGraw Hill Education Private Limited
NEW DELHI

McGraw-Hill Offices

New Delhi New York St Louis San Francisco Auckland Bogotá Caracas Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal San Juan Santiago Singapore Sydney Tokyo Toronto

Tata McGraw-Hill

Fourth reprint 2010
RLLBRDDFRQRAD

ISBN-13: 978-0-07-059116-5
ISBN-10: 0-07-059116-4

Published by Tata McGraw Hill Education Private Limited,
7 West Patel Nagar, New Delhi 110 008, typeset in LATEX at Script Makers,
19, A1-B, DDA Market, Paschim Vihar, New Delhi 110 063 and printed at
Sai Printo Pack, New Delhi 110 020

The McGraw·Hill Companies

To
Mama and Mai
who saw a dream
and to
Sarita, Kamini *and* ***Chaitanya***
who made the dream a reality.

Contents

Preface

THE BACKGROUND

Electromagnetics is one of the foundation subjects of electrical engineering and physics. My liking towards the subject of electromagnetic waves dates back to my undergraduate days. Of course that time the subject was taught in a fashion to make it appear like a vector calculus course. Later when I studied the subject of antenna and propagation during my graduate years from an excellent teacher, I was fascinated by it. My association with radio astronomy further increased my understanding of electromagnetics. The more I understood the subject the more I realized that electromagnetics is rather conceptual and not mathematical. Unfortunately, the conceptual nature of the subject is not emphasized at the undergraduate level. A course on electromagnetics then appears as a vector algebra course. The students, thus, quickly loose interest in the subject. In absence of any conceptual understanding of the subject they find no relevance of electromagnetics in the rest of the engineering curriculum. They perceive the course as a formula substitution exercise, which is certainly not the true nature of the course. The true nature of the course is conceptual. The course appears difficult at the outset because it requires a fair amount of imagination in visualizing the vector fields in three-dimensional space. However, once these fields are properly visualized, the subject becomes more of a fun than a burden. When I adopted the conceptual approach to the course in my classes, the students got more interested in the subject. They started enjoying the concepts of electromagnetics and finding their applications in real life problems.

THE APPROACH

At the undergraduate level the challenge is to make the subject interesting without compromising on mathematical rigor. This book, *Electromagnetic Waves*, is an attempt to make the subject exciting, conceptually sound and useful to the students of electrical engineering.

Having adapted the conceptual approach to electromagnetics, the availability of proper textbooks became a major concern. All major textbooks on electromagnetics follow the vector-algebra-dominated approach. Most of the books start with vector algebra and calculus followed by Maxwell's equations and their solutions. Before junior level, where the subject of electromagnetic waves is usually taught, the students acquire a good knowledge of circuit theory. A subject that starts with three-dimensional vector calculus does not immediately gel with the previous circuit-oriented courses. As a result, the students find an abrupt change in electrical engineering concepts. However, if a gradual transition from the circuit theory to the field theory is made, students find it more digestible. The subject of transmission line is precisely this transition. While analyzing transmission lines we still retain circuit terminology, like voltage, current, resistance, capacitance, etc. but introduce the concept of space. An electromagnetic wave related terminology is slowly introduced as the discussion on transmission line progresses. By the time the discussion on transmission lines is completed, the students already have some idea about the electromagnetic waves, albeit in one dimension. Extending the concepts from there to three dimensions is not very difficult.

The book has a large number of solved and unsolved problems of practical nature that would be useful in enhancing the analytical skills of the students. The problems are designed to test the conceptual understanding of the students rather than their algebraic manipulation capabilities. The review questions can help them in self-evaluating their understanding of the subject. For proper understanding of electromagnetic waves, the students are expected to solve all the solved and unsolved problems to the last detail and to answer all the review questions.

ORGANISATION OF THE BOOK

The book starts with a brief introduction to Electromagnetic Waves and their applications in Chapter 1. Chapter 2 is on transmission lines. In this chapter the limitations of circuit approach in analyzing high frequency circuits are highlighted and the concept of distributed element is introduced. It is shown that the natural solution for voltage and current is a wave type solution. Concepts of reflection, impedance transformation, impedance mismatch are discussed in detail and finally the applications of transmission line are explained. Chapter 3 is on Maxwell's equations. It is assumed that the students are familiar with basic vector operations like dot product, cross product, curl, divergence, etc. If not, the basics of vector algebra and calculus are given in an appendix at the end of the book. Maxwell's equations are derived in integral and differential form from the basic laws of electromagnetics. Further the boundary conditions are derived from the integral form of Maxwell's equations. At this stage there is no undue usage of different coordinate systems, etc. However, basics of the three coordinate systems have been given at the beginning of the chapter. Maxwell's equations are directly derived for the time varying fields which could be reduced to those for the static fields by putting time derivatives to zero.

In Chapter 4, the solution of Maxwell's equation for the time varying fields in an unbound medium is derived. Hence the concept of uniform plane wave is introduced. Ample references are made to transmission lines to show the basic

similarity between the two cases. Polarization of a wave has been discussed in detail highlighting its importance in real life situations. Further propagation of an EM wave in a conducting medium is investigated and the concept of skin depth, complex dielectric constant, surface current, etc are introduced in this chapter. The chapter concludes with the derivation of the Poynting theorem and introduction of the Poynting vector.

Chapter 5 discusses propagation of EM waves across media interfaces. Using physical reasoning, the Snell's law is established. Also the important concept of phase and group velocity is introduced in this chapter. An in-depth understanding to total internal reflection is developed at this stage. The reflection of EM waves from conducting boundaries is discussed in such a way that the reader can almost foresee the parallel plane waveguide.

Chapter 6 and 7 are on waveguides. The concept of modal propagation is developed in these chapters. Chapter 6 mainly discusses metallic waveguides in parallel plane and rectangular form. Chapter 7 is on dielectric waveguides that are important in optical technology, like thin film optical devices and optical fibers.

Chapter 8 and 9 are on antennas and antenna arrays. First, the basic philosophy of radiation is established and the concept of magnetic vector potential is introduced. The relation between the current and the magnetic vector potential is logically developed. Radiation characteristics of Hertz dipole and other dipoles are discussed further. Limitations of simple antennas are pointed out and a case is made for antenna arrays. After basic analysis of linear uniform arrays, the final Fourier transform relationship between the current distribution and the antenna radiation pattern has been established. Students who are already familiar with the Fourier transform find this relationship extremely exciting.

Chapter 10 is on propagation of EM waves for radio wave communications. Different transmission modes like ground waves, space waves and sky waves are discussed in this chapter. Propagation of an EM wave in the Ionosphere is discussed in great detail. Dielectric constant for ionized medium, with and without magnetization is derived and new interesting features of EM waves are discussed. A flavor of scattering of EM waves from the ionospheric irregularities is provided at the end.

All the chapters have numerous solved and unsolved problems besides conceptual review questions at the end of each chapter. Unnecessary complex formulae and their derivations are avoided in the book. Also there are no problems that would merely test the memory of a student. Formula substitution type problems are also avoided in the book. The problems are essentially designed to test the conceptual understanding of the subject and not the computational and algebraic skills of the students.

It is sincerely hoped that the presentation in the book and the problems will help the students in developing conceptual understanding of the subject without getting lost in the mathematical manipulations, and the students will start seeing EM wave related phenomena around them. If this book helps in making the subject of electromagnetic waves enjoyable, the whole exercise is worth the efforts.

WHO WILL BENEFIT

The book is targeted at junior students of electrical engineering in general, and electronics and telecommunication engineering in particular. A semester long course can be offered using Chapters 3–8 for general electrical/electronics engineering students. For telecommunication students, two-semester long courses can be offered using the whole content of the book. Depending upon the background of the students, the vector calculus can be covered before the Maxwell's.

The book will also be useful as reference material to practicing engineers working in the area of RF communication, radio broadcasting, microwave engineering and radar, optical and satellite communication, mobile communication, telemetry, remote sensing, radio astronomy, etc.

ACKNOWLEDGEMENTS

A large number of people have contributed directly or indirectly to this book. The first and the foremost are the wonderful teachers that I was fortunate enough to have during my student days. Prof. N.C. Mathur, Prof. D.K. Paul, Prof. A. Paul, and Prof. S. Mahaptra have played a very important role in igniting my curiosity about electromagnetics. Later during my professional career I interacted with great personalities like Prof. V. Radhakrishnan, Prof. Ch. V. Sastry, Prof. G. Swarup, who not only provided an in-depth understanding of electromagnetics and antenna, but also inspired me in many ways. Professional colleagues like Dr. Avinash Deshpande, Dr. K.S. Dwarakanath, Prof. B.N. Dwivedi, Prof. S.K. Jain, Prof. S.V. Kulkarni and many others who through their unending discussions polished my understanding of electromagnetics. I am indebted to all my teachers and colleagues for making me what I am today.

My understanding of the subject has been under constant scrutiny of the students of the electrical engineering department at IIT, Bombay. The students through their very intriguing questions force me to think deeper, which further refines my understanding of the subject. Every time I teach this subject I learn something new and exciting. I am indeed grateful to the students for making the teaching of this subject enjoyable.

My thanks are also due to the following referees of the manuscript whose comments positively contributed in improving the presentation.

1. Dr. Yaduvir Singh, Dept. of Electrical & Instrumentation Engineering, Thapar Institute of Engineering & Technology (Deemed University), Patiala.
2. Mr. R. Selvaraj, School of EEE (Communication wing), SASTRA Deemed University, Thanjavur.
3. Prof. Partha Pratim Bhattacharya, BS Institute of Engineering & Management, Kolkata.
4. Mr. N.V. Venkateswara Rao, Associate Professor, EEE Department, N.B.K.R. Institute of Science & Technology, Vidyanagar, Andhra Pradesh.

My family has contributed both, directly and indirectly, to this book. My wife, Sarita, who herself has doctorate in electrical engineering with specialization in microwaves and microwave photonics, has been a regular critic of this work. Her comments and critical suggestions have greatly improved the contents and presentation of the book. I am indebted to her for all the professional sacrifice she has made for my progress. My daughter, Kamini (who also helped in preparing part of the manuscript) and son, Chaitanya have been a constant source of energy needed to take up such a mammoth task. The love and affection bestowed on me by my parents and my family is beyond words.

Finally, my thanks are due to Ms. Seema Mane, Mr. R.S. Patwardhan, Mr. S.L. Parab who provided administrative and secretarial support in developing the manuscript. They carried out the job in a most pleasing fashion.

This book is a tribute to all those who taught me something in life.

R K SHEVGAONKAR

CHAPTER 1

Introduction

1.1 WHY STUDY ELECTROMAGNETIC WAVES?

Electromagnetic waves, a subject so classical and at the same time so modern. In ancient times people used to investigate things like

- why do small paper clips get attracted to an amber rod rubbed with silk,
- what causes lightening,
- why does a magnetic needle deflect when it is kept close to current carrying conductor,
- why do stars twinkle but the planets do not,
- how does the light reach the earth from the sun when there is no medium in between,
- why do light rays get focused when passed through a curved mirror or a lens
- why do colors get separated when a light beam passes through a prism, and so on.

While in the modern days people ask questions like

- how do we receive TV and radio signals,
- how do transmitted signals get affected while propagating through the earth's atmosphere,
- why do we not receive radio stations inside a train compartment,
- why is the radio reception good in some corner of the room but not in the other,
- why do medium wave radio stations not show any temporal variation in their signal strengths but the short wave radio stations do,
- why do the cell phones have signal fluctuation,
- why are TV antennas mounted on the top of the building but the radio antennas are not,
- why do some things get heated in a microwave oven and others not,

- why is there a disturbance on a radio but not on a TV when a motor cycle is started in the vicinity,
- why does a printed circuit board which was working satisfactorily at low frequencies start malfunctioning at high frequencies, and so on.

The list is unending. All these phenomena, which appear to come from different areas, have a common thread running through them and that is the *Electromagnetism*.

Initially the subject of electricity and magnetism was a matter of intellectual curiosity. Also in early days electricity and magnetism were considered unrelated phenomena. As time progressed the relationship between electricity and magnetism became evident and word 'Electromagnetism' emerged. Today, we hardly find any electrical or electronic device around us, which does not work on the principles of electromagnetism, or is not influenced by it. The things which were considered science fictions a hundred years ago have become a reality due to tremendous progress of electromagnetics and its engineering employment.

Today, the applications of electromagnetics can be broadly divided into two categories (i) low frequency but high power (ii) high frequency but low power. There are a few applications which have high frequency and high power. For low frequency applications, the analysis of static fields like electrostatics and magnetostatics is adequate to investigate the electromagnetic phenomenon. However, as the frequency increases, electric and magnetic fields get more strongly coupled and we always observe a composite phenomenon of electric and magnetic fields. One can then conveniently divide the subject of electromagnetics into two parts, the static electromagnetics and the time varying electromagnetics. As will be clear subsequently, the time varying electric and magnetic fields always constitute a wave phenomenon called the electromagnetic wave which is the prime subject of discussion of this book. (In this book, the discussion is focused on the time varying electric and magnetic fields, their inter relationship, their spatial and temporal characteristics, etc.)

The phenomenon of electromagnetism in totality is governed by the four Maxwell's equations, which can be derived from the physical laws like the Gauss Law, the Ampere's law and the Faraday's law of electromagnetic induction. Generally, we find two schools of thought regarding Maxwell's equations. Some people believe that the laws of electricity and magnetism were first established through experimentation and the mathematical representation of these experimental laws led to Maxwell's equations. People following this line of thought consider the origin of the electromagnetism in experimentation and therefore seek stronger physical reasoning to explain an electromagnetic phenomenon. The people belonging to the other school of thought, take the four Maxwell's equations as mathematical postulates and the physical laws like the Gauss law, etc. as the experimental verification of the postulates. Both approaches have their merits. The first approach, which has a foundation in experimentation, prevents people from getting lost in mathematical manipulations. This approach

keeps reminding an investigator that electromagnetism is a physical phenomenon and therefore one should develop a habit of finding a physical picture for every mathematical manipulation one carries out. Consequently, one should not lose touch with the physical phenomenon. The second approach is more accurate and elegant as it is free from experimental errors. With only experimental measurements one could never establish such exact relationships like inverse square law and so on.

The electromagnetic theory is the generalization of the circuit theory, or the circuit theory is rather a special case of the electromagnetic theory. The circuit theory deals with quantities like voltage, current, resistance, inductance and capacitance which are scalar in nature. The electromagnetic theory on the other hand deals with quantities like the electric and magnetic fields which are vector quantities. The complexity of the electromagnetic theory is several times higher compared to the circuit theory. Although every phenomena of electricity and magnetism can be analyzed in the frame work of electromagnetic theory, at low frequencies the circuit approach is adequate. As the frequency increases the inadequacy of the circuit approach is evident and one is forced to follow the electromagnetic field approach. For example, sending electrical signals through free space can never be visualized using circuit approach. Similarly, it is difficult to visualize a current flow in a wire which is connected to a source at one end but is left open at the other end. The correct approach therefore would be to apply the simpler circuit approach as far as it can be applied and make a transition to the field approach when the circuit model tends to break down.

1.2 APPLICATIONS OF ELECTROMAGNETIC WAVES

Time varying electric and magnetic fields, in general, constitute a wave phenomenon. Some of the applications where electromagnetic waves can be encountered are given in the following sections:

1.2.1 Transmission Lines

At low frequencies, an electrical circuit is completely characterized by the electrical parameters like resistance, inductance, etc. and the physical size of the electrical components plays no role in the circuit analysis. Simple Kirchhoff's laws are adequate for analyzing a circuit. However, as the frequency increases the size of the components becomes important and the space starts playing a role in the performance of the circuit. The voltage and currents exist in the form of waves. Even a change in the length of a simple connecting wire may alter the behavior of the circuit. The circuit approach then has to be re-investigated with inclusion of the space into the analysis. This approach is then called the transmission line approach. Two conductor transmission media like the coaxial cable, flat ribbon cable, etc. are the examples of transmission lines. Although, the primary objective of a transmission line is to carry electromagnetic energy efficiently from one location to other, they find wide applications in high frequency circuit designs.

1.2.2 High Frequency and Microwave Circuits

As the frequency increases, any discontinuity in the circuit path leads to electromagnetic radiation. Also at high frequencies the transit time of the signals cannot be ignored. In the era of high speed computers, where data rates are approaching to a few Gb/sec, the phenomena related to the electromagnetic waves, like the bit distortion, signal reflection, impedance matching play a vital role in high speed communication networks.

1.2.3 Antennas

An antenna is a device which can launch and receive electromagnetic waves efficiently. An antenna which appears as a passive looking device in a communication system, is one of the most important devices. But for the large antennas, the communication between an earth station and a satellite is practically impossible. The communication which can be established with a few watts of power, would need few MW of power in the absence of proper antennas. Many types of antennas have been in use over several decades. However, the antenna research is still very active. With recent advances in mobile communication, design of compact, efficient, multi-frequency antennas have received a new impetus in the last decade.

1.2.4 Fiber Optic Communication

Fiber optic communication is the most modern form of guided wave communication. Electromagnetic theory is used to investigate propagation of light inside the optical fibers. The phenomenon of dispersion which has a direct bearing on the speed of communication, is due to the difference in speed of light for different modes of the optical fiber. The modal propagation inside an optical fiber is a direct consequence of the wave nature of light. Electromagnetic wave theory is also important in the analysis of lasers and photo detectors. By employing the complex phenomena of electromagnetic waves, a variety of fiber optic devices have been developed for efficient high speed long haul communication.

1.2.5 Mobile Communication

Knowledge of electromagnetic wave propagation plays a vital role in understanding the radio environment. Depending upon the variation of the signal strength as a function of distance, different frequency reuse schemes can be employed in a cellular system. Fading is one of the important aspects of mobile communication. Efficient signal processing algorithms need the knowledge of the radio environment to correctly predict the fading behavior.

1.2.6 Electromagnetic Inference (EMI) and Compatibility

An electrical circuit, which especially switches heavy current, tends to give electromagnetic radiation. This radiation interferes with the other circuit elements affecting the overall performance of the circuit. Switch mode power supply and high speed digital circuits create a substantial amount of EMI. To protect the circuits from EMI, shielding techniques are employed. Designing of proper EMI shields needs a thorough knowledge of electromagnetics.

1.2.7 Radio Astronomy

This is a branch of astronomy in which observations of the sky are carried out at radio frequencies. The radio signals received from the sky are very weak in nature. State of art communication receivers and antennas are required to detect these signals. Therefore radio astronomy is a combination of electronics engineering and physics. This is one of the fields where a thorough understanding of electromagnetic waves is necessary. In fact, almost all aspects of electromagnetic waves in some form or the other, are employed in radio astronomy.

These are some of the major areas where knowledge of electromagnetic waves is profoundly used. There are many more applications of electromagnetic waves in addition to these.

1.3 SUMMARY

In this chapter the importance of electromagnetic waves in electrical engineering has been highlighted. A wide application of electromagnetic waves have been metioned. The understanding of electromagnetic waves is useful in invertigating many classical and modern engineering problems. Many, apparently unrelated, fields like optical astronomy and wireless communication share the same principles of electromagnetic waves. In the following chapters we develop understanding of various aspects like, propagation, confinement, radiation, etc, of the electromagnetic waves.

CHAPTER 2

Transmission Lines

2.1 INTRODUCTION

The phenomenon of electromagnetic waves, in principle, is associated with the time varying electric or magnetic fields although its presence becomes more visible at higher frequencies. Like any other wave, the electromagnetic wave is a composite phenomenon of space and time. At low frequencies since the size of the circuit elements is negligible compared to the wavelength, the effect of space is generally ignored. A circuit is characterized by its temporal response and the circuit components are described by their electrical parameters. That is to say that at low frequencies we get correct circuit performance if the circuit elements have the correct electrical values irrespective of their physical size. This however is not true at frequencies beyond few hundreds of kHz. In fact, even the lengths of simple connecting wires alter the circuit performance. Naturally, there is a conceptual change as we increase the frequency of operation and an electrical circuit can no longer be analysed without taking the 'space' into consideration. This is due to the fact that the signal requires a finite time to travel along an electric circuit which is no more negligible compared to the time period of the signal. This effect is called the 'transit time effect'. Therefore at high frequencies the transit time effect has to be included in the circuit analysis.

In general, the electromagnetic wave phenomenon is described by behavior of the electric and the magnetic fields in the three dimensional space as a function of time. However, to develop the basic concepts let us first analyse a simpler but important case of electromagnetic waves, the *transmission line*. In the anlysis of a transmission line, we continue to use the electrical quantities like voltage, current, resistance, capacitance, etc. as used in low frequency circuits, but appropriately incorporate the effect of finite transit time. We then obtain the voltage and current which have wave type behavior. Analysis of a tranmission line therefore helps in understanding some of the fundamentals of the electromagnetic waves.

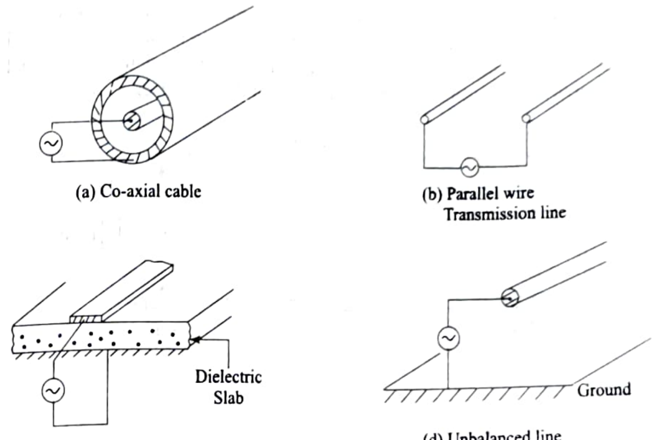


Fig. 2.1 *Transmission line configurations.*

2.2 CONCEPT OF DISTRIBUTED ELEMENTS

Figure 2.1 shows various types of transmission lines used in practice. However, generically a transmission line is a two-conductor system one end of which is connected to a source and the other end is connected to a load as shown in Fig. 2.2. Let the length of the transmission line be l and let it be excited by a voltage source at frequency f. Realizing that no signal can travel with infinite speed, let the speed of the signal on the tranmission line be v. Now, let us suppose that at some instant of time, the voltage corresponding to point P on the voltage waveform (say V_p) is connected to the transmission line at AA′. Due to finite speed of the signal the V_p voltage will not appear at BB′ instantaneously but will be delayed by the transit time, $t_r = l/v$. Obviously during this time the input signal does not remain at point P but changes to point Q. That means when the voltage is passing through a point Q at AA′, it is passing through point P at BB′. In other words there is a phase difference between the voltages at AA′ and BB′. It should be realized that there will be a phase difference between any two locations on the transmission line, and the phase difference will increase as the seperation between the locations increases. This will be true even when the conductors used in the transmission line are ideal. It is then interesting to find that even when the resistance of a conductor is zero there is a potential difference between its ends. With a little thinking, one would find that although the resistive voltage drop across an ideal conductor is zero, the reactive drop will be finite for any non-zero angular frequency, ω. As the frequency increases the inductive reactance (ωL) and the capacative susceptance (ωC) increase proportionately making reactive effect more observable. From transit time view-point, an increase in frequency reduces the signal time period, making the transit time a measureable fraction of the signal period. Therefore, the dominance of the circuit reactance at high frequencies and the effect of finite transit time are the two faces of the same phenomenon.

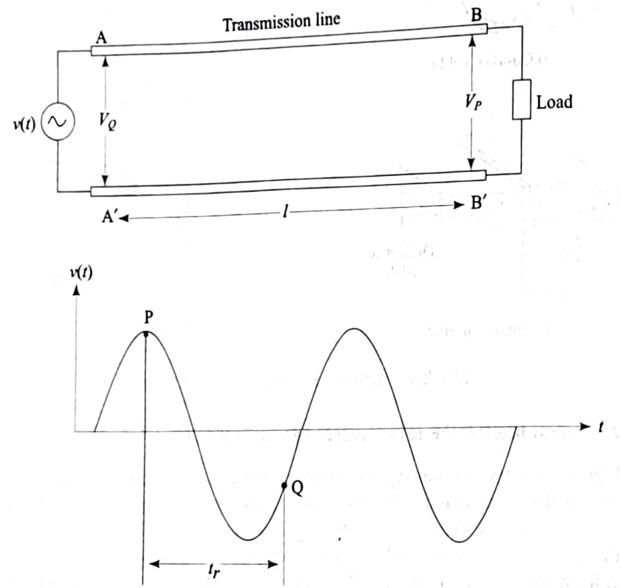


Fig. 2.2 *Voltage on a transmission line.*

The lumped element circuit analysis, at low frequencies, is valid provided the signal transit time is negligible compared to the signal period, i.e. if $l/v << T$, where $T = 1/f$. One can, therefore, model a transmission line by lumped elements by dividing it into small sections such that over each sub-section the transit time effect is negligible. It is clear that for any finite length of the subsection, the transit time effect will be appreciable beyond certain frequency, no matter how small the subsection is. If the lumped element model be applicable at all frequencies, the length of the subsection must tend to zero and the analysis must be carried out in some form of limit. The circuit elements now cannot be defined for the entire transmission line but has to be defined for a subsection. It is, therefore, appropriate to define circuit parameters for unit length of transmission line. The circuit elements are not located at a particular location of the line but are distributed all along its length. The high frequency analysis of a transmission line can be carried out by using the concept of distributed elements.

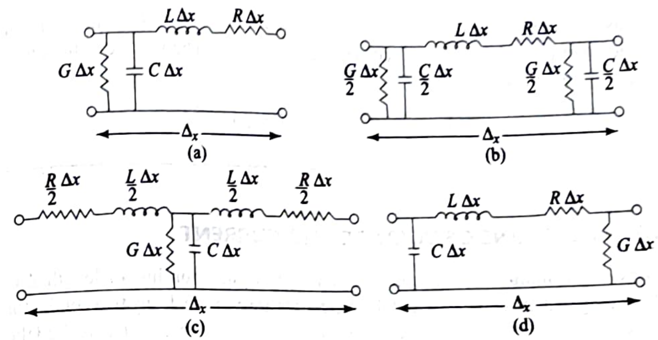


Fig. 2.3 *Lumped circuit equivalent for infinitesimal section of a transmission line.*

EXAMPLE 2.1 A two conductor transmission line is 10 cm long. A sinusoidal signal of 1V peak amplitude is applied to one end of the line. If the signal travels on the line with a speed of 2×10^8 m/sec, find (a) transit time on the line (b) frequency at which the transit time is 10% of the signal period (c) signal voltage on the other end of the line at an instant when the input signal is passing through +ve maximum for a frequency of 500 MHz.

Solution:

(a) Transit time $t_r = \dfrac{\text{Length of the line}}{\text{Velocity}} = \dfrac{0.1\text{ m}}{2 \times 10^8\text{ m/sec}} = 0.5$ nsec

(b) Transit time should be 10% of the period T, i.e.

$$\text{The transit time } t_r = 0.1\,\text{T}$$

$$\Rightarrow \quad T = \frac{t_r}{0.1} = 10\,t_r = 5\text{ nsec}$$

$$\text{Frequency} = \frac{1}{T} = \frac{1}{5 \times 10^{-9}} = 200\text{ MHz}$$

(c) Let the signal voltage be given as

$$v(t) = A\cos(2\pi f t)\text{V}$$

$$\text{Given that} \quad A = 1,\ f = 500\text{ MHz}.$$

We, therefore, get

$$v(t) = \cos(\pi \times 10^9 t)$$

Taking the instant as $t = 0$, when the input voltage $v(t) = 1$ V, the signal at the other end of the line will correspond to $t = -t_r$. The voltage at the other end of the line therefore is

$$\cos(-\pi \times 10^9 t_r) = \cos\left(-\frac{\pi}{2}\right) = 0 \text{ V}.$$

2.3 EQUATIONS OF VOLTAGE AND CURRENT

Consider an infinitesimally small section of a transmission line of length Δx. Assuming that the conductor and the dielectrics are non-ideal, any two conductor system can be represented by the four characteristic circuit parameters (called the 'primary constants' of the line) namely, the series resistance, the series inductance, the shunt capacitance and the shunt conductance. Let R, L, C and G represent these quantities respectively per unit length of the line. The choice of unit length is arbitary but in MKS system it is one meter. The equivalent lumped circuit for the infinitesimal section of the transmission line can be shown as in Fig. 2.3. All the circuits in Fig. 2.3 are equivalent for $\Delta x \to 0$. The resistance, inductance, conductance and capacitance of the infinitesimal section of the line are $R\Delta x$, $L\Delta x$, $G\Delta x$, and $C\Delta x$ respectively.

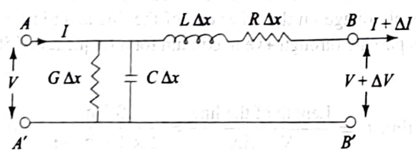


Fig. 2.4 *Voltage and current on an infinitesimal section of a transmission line.*

Let a sinusoidal voltage of angular frequency ω be applied between AA′, and let a current I flow into the terminal A (refer to Fig. 2.4). Now due to voltage drop in the series elements, R and L, the voltage at BB′ will not be same as that at AA′. Similarly, since some part of the input current will be bypassed through the shunt elements, C and G, the output current at point B will not be same as that at A. Let the current and the voltage at BB′ be $I + \Delta I$ and $V + \Delta V$ respectively. We can then write,

$$\Delta V = -(R\Delta x + j\omega L\Delta x)I \quad (2.1)$$

$$\Delta I = -(G\Delta x + j\omega C\Delta x)V \quad (2.2)$$

The negative sign indicates that the voltage and current at BB′ are less than their respective values at AA′.

Equations (2.1), (2.2) can be re-written as

$$\frac{\Delta V}{\Delta x} = -(R + j\omega L)I \quad (2.3)$$

$$\frac{\Delta I}{\Delta x} = -(G + j\omega C)V \quad (2.4)$$

As discussed earlier, if the lumped analysis has to be valid at all frequencies, the length of the section Δx must tend to zero. Taking the limit of Eqns (2.3) and (2.4) for $\Delta x \to 0$, we get

$$\lim_{\Delta x \to 0} \frac{\Delta V}{\Delta x} \equiv \frac{dV}{dx} = -(R + j\omega L)I \quad (2.5)$$

$$\lim_{\Delta x \to 0} \frac{\Delta I}{\Delta x} \equiv \frac{dI}{dx} = -(G + j\omega C)V \quad (2.6)$$

Therefore, we find that the voltage and current on a transmission line are governed by two coupled first order differential Eqns (2.5) and (2.6). Differentiating Eqn (2.5) with respect to x, we get

$$\frac{d^2V}{dx^2} = -(R + j\omega L)\frac{dI}{dx} \quad (2.7)$$

Now substituting for dI/dx from Eqn (2.6) we have

$$\frac{d^2V}{dx^2} = -(R + j\omega L)[-(G + j\omega C)V] \quad (2.8)$$

$$= (R + j\omega L)(G + j\omega C)V \quad (2.9)$$

Similarly, if we differentiate Eqn (2.6) with respect to x and substitute for dV/dx from Eqn (2.5), we get

$$\frac{d^2I}{dx^2} = (R + j\omega L)(G + j\omega C)I \quad (2.10)$$

Let us now define a quantity called propagation constant, γ of the transmission line as

$$\gamma^2 = (R + j\omega L)(G + j\omega C) \quad (2.11)$$

The physical interpretation of γ will be discussed later. However, one thing is clear that γ is some characteristic parameter of the line similar to the primary constants except that it depends upon the operating frequency also. With the introduction of γ the Eqns (2.9) and (2.10) can be compactly written as

$$\frac{d^2V}{dx^2} = \gamma^2 V \quad (2.12)$$

$$\frac{d^2I}{dx^2} = \gamma^2 I \quad (2.13)$$

It is interesting to note that both voltage and current are governed by the same differential equation.

EXAMPLE 2.2 For a transmission line the per unit length parameters are 0.1Ω/m, 0.2 μH/m, 10 pF/m and 0.02 ℧/m. Find the complex propagation constant at (a) 1 MHz (b) 1 GHz.

Solution:

(a) The propagation constant

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$$

$$R = 0.1\ \Omega/\text{m}, \qquad L = 0.2 \times 10^{-6}\text{H/m}$$

$$G = 0.02\ ℧/\text{m}, \qquad C = 10 \times 10^{-12}\text{F/m}$$

$$\omega = 2\pi f = 2\pi \times 10^{6}\text{rad/sec}$$

$$\Rightarrow \gamma = \sqrt{(0.1 + j2\pi \times 10^{6} \times 0.2 \times 16^{-6})(0.02 + j2\pi \times 10^{6} \times 10^{-11})}$$

$$= 0.117 + j0.108\,/\text{m}$$

(b) At 1 GHz, $\omega = 2\pi \times 10^{9}$ rad/s,

$$\gamma = 1.4 + j9\,/\text{m}$$

Since, for a given operating frequency γ is constant, Eqns (2.12) and (2.13) are homogeneous equations with constant coefficients, and their solutions can be written as

$$V = V^{+}e^{-\gamma x} + V^{-}e^{+\gamma x} \tag{2.14}$$

$$I = I^{+}e^{-\gamma x} + I^{-}e^{+\gamma x} \tag{2.15}$$

Where, $V^{+}, V^{-}, I^{+}, I^{-}$ are arbitrary constants which are to be evaluated by using appropriate boundry conditions. These constants in general are complex and their phases represent the temporal phases from some reference time. In the above equations the time harmonic function is implicit. If we want to write the instantaneous values of the voltage and current we have to multiply Eqns (2.14) and (2.15) by $e^{j\omega t}$ to arrive at

$$V(t) = V^{+}e^{j\omega t - \gamma x} + V^{-}e^{j\omega t + \gamma x} \tag{2.16}$$

$$I(t) = I^{+}e^{j\omega t - \gamma x} + I^{-}e^{j\omega t + \gamma x} \tag{2.17}$$

Now, from Eqn (2.11) we see that the propagation constant γ is a complex quantity and we can write,

$$\gamma = \alpha + j\beta \tag{2.18}$$

Substituting for γ from Eqn (2.18) into Eqns (2.16) and (2.17) we obtain,

$$V(t) = V^{+}e^{-\alpha x}e^{j\omega t - j\beta x} + V^{-}e^{\alpha x}e^{j\omega t + \beta x} \tag{2.19}$$

$$I(t) = I^{+}e^{-\alpha x}e^{j\omega t - j\beta x} + I^{-}e^{\alpha x}e^{j\omega t + \beta x} \tag{2.20}$$

It can be noted that all the terms in Eqns (2.19) and (2.20) are complex in general and their phases are functions of both space (x) and time (t). For example, take the first term of $V(t)$. The phase of this term is $\phi = \omega t - \beta x + phase(V^{+})$. We can then see the contribution which the space and time make to the phase of this term as follows.

If we freeze time by making $t = constant$, the phase ϕ changes linearly as a function of distance x. It means if we simultaneously look at the whole transmission line, the phase will vary linearly from one end of the line to the other end of the line. On the other hand, if we freeze space, it means observe the phase at a given point on the transmission line, the phase will linearly vary as a function of time t. If we, therefore, plot the function $Re\left\{e^{(j\omega t - j\beta x)}\right\} = cos(\omega t - \beta x)$ in space and time, the function will appear like a sinusoidally corrugated surface as shown in Fig. 2.5. Let us consider a point P on the surface at $t = 0$. As we now increase t, the point moves towards right in space, that is towards $+x$ direction. Since, this is true for every point on the surface, the whole pattern apears to move in the $+x$ direction as a function of time. This behavior is nothing but a 'travelling wave'. The first term on the RHS of Eqn (2.19) hence represents a voltage wave travelling in the $+x$ direction.

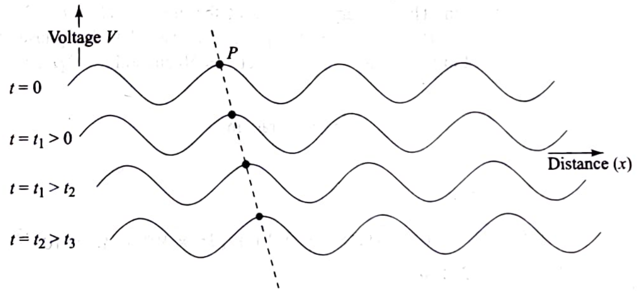


Fig. 2.5 *Voltage as a Function of x for different time*

Similarly, if we plot the $Re\left\{e^{j\omega t + j\beta x}\right\} = \cos(\omega t + \beta x)$ corresponding to the second term on the RHS of Eqn (2.19), we will get a corrugated surface similar to that in the previous case, but now a point P will move towards the left that is towards $-x$ direction as a function of time as shown in Fig. 2.6. In other words, this term represents a voltage wave travelling in $-x$ direction.

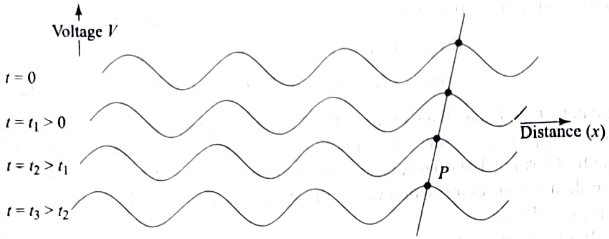


Fig. 2.6 *Voltage as a function of x for different time.*

On identical lines one can show that the first term of $I(t)$ represents a current wave travelling in the $+x$ direction, and the second term represents a current wave travelling in $-x$ direction.

It is clear from the above discussion that as the effect of finite transit time is introduced, that is, as we make use of the distributed elements, the natural solution which we get for voltage and current is a wave type solution. In general, we can therefore say that in an electric circuit the time varying voltage and current exist in the form of waves.

EXAMPLE 2.3 A voltage wave at 1 GHz is travelling on a transmission line in $+x$ direction. The primary constants of the line are, $R = 0.5\ \Omega/\text{m}$, $L = 0.2\ \mu\text{H/m}$, $G = 0.1\ \mho/\text{m}$, $C = 100\ \text{pF/m}$. The wave has 30° phase at $t = 0$ and $x = 0$. Find the phase of the wave at $x = 50$ cm and $t = 1\mu$sec.

Solution:

$$\omega = 2\pi \times 10^9 \text{rad/m}$$

The propagation constant:

$$\gamma = \sqrt{(R + j\omega l)(G + j\omega C)}$$
$$= \sqrt{(0.5 + j2\pi \times 10^9 \times 0.2 \times 10^{-6})(0.1 + j2\pi \times 10^9 \times 100 \times 10^{-12})}$$
$$\gamma = 2.23 + j28.2\ \ 1/\text{m}$$

$$\beta = 28.2\ \text{rad/m}$$

$$\text{Phase of the wave} = \text{Initial Phase} + \omega t - \beta x$$
$$= 30^\circ + 2\pi \times 10^9 \times 10^{-6}\text{rad} - 28.2 \times 0.5\ \text{rad}$$
$$= 6269.608\ \text{rad} = 35922.129^\circ$$

2.3.1 Phase and Attenuation Constants

Having understood the nature of the voltage and current on a transmission line, we can now assign some physical meaning to the propagation constant γ. Since βx is the phase of the wave as a function of x, β represents the phase change per unit length of the transmission line for a travelling wave. Since β defines the phase variation along the transmission line, it is called the 'phase constant' of the transmission line, and it has the unit of *radians per meter*. Now, since for a wave the phase change over a wavelength λ is 2π, by definition we have, $\beta\lambda = 2\pi$. The wavelength of a wave on the transmission line therefore is $\lambda = 2\pi/\beta$. Note, γ and therefore β depends upon the primary constants, R, L, G, C and the frequency, and consequently the wavelength of the wave on the transmission line is also a function of the primary constants and the frequency, in general.

Let us now look at the amplitudes of the two terms of $V(t)$. Amplitude of the first term, i.e. $|V^+|e^{-\alpha x}$ represents the amplitude of the wave travelling in the $+x$ direction at location x on the line. For a positive α, the amplitude exponentially decreases as a function of x. The quantity α then represents attenuation of the wave on the transmission line, and consequently called the 'attenuation constant' of the line. The unit of α is nepers per meter. It can be easily seen that if a line has an attenuation constant of 1 neper/m, a unit amplitude voltage wave will reduce to e^{-1} over a distance of 1 meter. The attenuation of the wave in dB is $-20\log\ (e^{-1}) = 8.68$ dB. We can then say that 1 neper is equivalent to 8.68 dB. The attenuation constant (α) of a line therefore can be given either in neper/m or dB/m. It should however be remembered that the use of α in any of the voltage or current equations, has to be in the unit of neper/m.

EXAMPLE 2.4 For the wave in Example 2.3, calculate attenuation constant in nepers/m and dB/m. If the voltage of the forward travelling wave at $t = 0$ and $x = 0$ is 8.66 V, find the voltage at $x = 1$m at time $t = 100$ nsec. What is the peak voltage at $x = 1$m?

Solution:

From Example 2.3, $\gamma = 2.23 + j28.2$ per meter.

$$\Rightarrow \qquad \alpha = 2.23\ \text{nepers/m} = 8.68 \times 2.23 = 19.356\ \text{dB/m}$$

The voltage for the forward travelling wave is

$$v(t) = Re\{V^+ e^{-\alpha x} e^{j(\omega t - \beta x)}\}$$

Taking $V^+ = |V^+|e^{j\phi}$, where ϕ is the initial phase, we get

$$v(t) = |V^+|e^{-\alpha x}\cos(\phi + \omega t - \beta x)$$

At $x = 0$ and $t = 0$, it is given that $v(t) = 8.66V$. Therefore

$$8.66 = |V^+| \cos 30^o$$

$$\Rightarrow \quad |V^+| = 10\text{V}$$

The instantaneous voltage at $x = 1$m and $t = 100$ nsec is

$$v(t) = 10e^{(-2.23\times 1)} \cos(30^\circ + 2\pi \times 10^9 \times 100 \times 10^{-9}(\text{rad}) - 28.2 \times 1(\text{rad}))$$

$$= -0.8887\text{V}$$

Peak voltage at $x = 1$ m is

$$V_p = 10\, e^{-2.23} = 1.0753\text{V}$$

EXAMPLE 2.5 If the wave in Example 2.4 is travelling in $-x$ direction with all other things same, find the instantaneous voltage at the same time and at the same location.

Solution:

The voltage for a wave travelling in $-x$ direction is given as

$$v(t) = Re\{V^- e^{\alpha x} e^{j(\omega t + \beta x)}\}$$

Taking $V^- = |V^-|e^{j\phi}$, we have

$$v(t) = |V^-|e^{\alpha x} \cos(\phi + \omega t + \beta x)$$

At $x = 0$ and $t = 0$, $v(t) = 8.66$ V (given). Therefore

$$8.66 = |V^-| \cos 30^\circ$$

$$\Rightarrow \quad |V^-| = 10 \text{ V}$$

The voltage at $x = 1$ m and $t = 100$ nsec will be

$$v(t) = 10\, e^{(2.23\times 1)} \cos(30^o + 2\pi \times 10^9 \times 100 \times 10^{-9}(\text{rad}) + 28.2 \times 1(\text{rad}))$$

$$= -83.77 \text{ V}$$

2.3.2 Evaluation of Arbitrary Constants

To evaluate the arbitrary constants V^+, V^-, I^+, I^- in Eqns (2.19) and (2.20), we have to apply appropriate boundary conditions. However, before we do that, we can reduce the number of arbitrary constants by making use of the fact that the voltage and current given by Eqns (2.19) and (2.20) must satisfy the basic differential Eqns (2.5) and (2.6) at every point on the transmission line.

Substituting for V and I from Eqns (2.19) and (2.20) in (2.5) we get

$$\frac{d}{dx}\{V^+e^{-\gamma x} + V^-e^{\gamma x}\} = -(R + j\omega L)\{I^+e^{-\gamma x} + I^-e^{\gamma x}\} \quad (2.21)$$

$$\Rightarrow -\gamma V^+e^{-\gamma x} + \gamma V^-e^{\gamma x} = -(R + j\omega L)\{I^+e^{-\gamma x} + I^-e^{\gamma x}\} \quad (2.22)$$

Although, Eqn (2.22) appears as one equation, there are two equations embedded in it. As seen earlier, the two terms $e^{-\gamma x}$ and $e^{\gamma x}$ represent two travelling waves in opposite directions (one in $+x$ direction and one in $-x$ direction). Therefore, if Eqn (2.22) is to be satisfied at every point on the line, that is, for every value of x, the individual waves must satisfy the equation. In other words, the coefficients of $e^{-\gamma x}$ and $e^{\gamma x}$ on the two sides of the equality sign must be separately equated. We, therefore, get

$$\text{Coefficient of } e^{-\gamma x}: \quad -\gamma V^+ = -(R + j\omega L)I^+ \quad (2.23)$$

$$\text{Coefficient of } e^{\gamma x}: \quad \gamma V^- = -(R + j\omega L)I^- \quad (2.24)$$

Since, $\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$ (see Eqn (2.11)), we get from (Eqns 2.23) and (2.24) respectively

$$\frac{V^+}{I^+} = \frac{R + j\omega L}{\gamma} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad (2.25)$$

$$\frac{V^-}{I^-} = -\frac{R + j\omega L}{\gamma} = -\sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad (2.26)$$

It can be seen easily that since the quantity $\sqrt{(R + j\omega L)/(G + j\omega C)}$ has the dimensions of impedance (it is a ratio of voltage and current) and is a function of the primary constants of the line and the operating frequency, it is a characteristic parameter of the line. It is therefore called the 'characteristic impedance' of the transmission line and is normally denoted by Z_0.

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad (2.27)$$

The parameters γ and Z_0 are called secondary parameters of the transmission line. We will subsequently see that for the analysis of a transmission line, the knowledge of secondary parameters is adequate and generally there is no necessity of going to the primary constants.

EXAMPLE 2.6 A transmission line has primary constants $R = 0.1\Omega$/m, $G = 0.01$ ℧/m, $L = 0.01\mu$H/m, $C = 100$ pF/m. Find the characteristic impedance of the line at 2 GHz.

Solution:

$$\omega = 2\pi \times 2 \times 10^9 = 4\pi \times 10^9 \text{ rad/sec}$$

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

$$= \sqrt{\frac{0.1 + j4\pi \times 19^9 \times 0.01 \times 10^{-6}}{0.01 + j4\pi \times 10^9 \times 100 \times 10^{-12}}}$$

$$= 10 + j0.0358\ \Omega$$

From Eqns (2.25), (2.26) we note that the ratio of V^+ and I^+ is Z_0 whereas the ratio of V^- and I^- is $-Z_0$. Recalling that the V^+ and I^+ are the respective amplitudes of the voltage and current waves travelling in $+x$ direction, we can conclude that the ratio of the voltage and current for a forward travelling wave at every point on the line is constant and is equal to the characteristic impedance Z_0. Said differently, a forward travelling wave always sees the characteristic impedance irrespective of the other boundary conditions on the transmission line. Similarly, we can conclude that the backward travelling wave (the wave travelling in $-x$ direction) always sees an impedance which is $-Z_0$. It should be pointed out that this $-Z_0$ does not represent a negative resistance in the conventional sense but is rather a manifestation of the direction of the wave travel.

Substituting for I^+ and I^- from Eqns (2.25) and (2.26) into Eqns (2.14) and (2.15) we get

$$V = V^+ e^{-\gamma x} + V^- e^{\gamma x} \tag{2.28}$$

$$I = \frac{V^+}{Z_0} e^{-\gamma x} - \frac{V^-}{Z_0} e^{\gamma x} \tag{2.29}$$

Equations (2.28) and (2.29) represent the generalized voltage and current on a transmission line which has propagation constant γ and characteristic impedance Z_0.

EXAMPLE 2.7 For the transmission line in Example 2.6, there are two waves travelling in opposite directions. At $x = 0$ and $t = 0$, the phase of the forward wave is zero and its amplitude is 2 V, whereas the phase of the backward wave is $\pi/3$ and its amplitude is 0.5 V. (i) What is the instantaneous voltage and current at $x = 50$ cm and $t = 1$ nsec (ii) What is the peak voltage and peak current at $x = 1$ m.

Solution:

$$\text{Angular frequency} \quad \omega = 4\pi \times 10^9 \quad \text{rad/sec}$$

(i) The voltage of the forward travelling wave is

$$v_f(t) = |V^+| e^{-\alpha x} \cos(\phi^+ + \omega t - \beta x)$$

Given: $\phi^+ = 0$ and $v_f(t) = 2$ V at $x = 0$, $t = 0$. Therefore we get $|V^+| = 2$

Similarly for the backward travelling wave

$$v_b(t) = |V^-| e^{\alpha x} \cos(\phi^- + \omega t + \beta x)$$

Given: $\phi^- = \pi/3$, and $v_b(t) = 0.5$ V at $x = 0$ and $t = 0$. Therefore we have

$$0.5 = |V^-| \cos \pi/3$$

$$\Rightarrow \quad |V^-| = 1.0 \text{ V}$$

Now, the propagation constant of the line is

$$\gamma = \sqrt{(0.1 + j\omega 0.01 \times 10^{-6})(0.01 + j\omega \times 10^{-10})}$$

$$= 0.055 + 12.566 \quad \text{per meter}$$

$$\Rightarrow \alpha = 0.055 \quad \text{nepers/m}$$

$$\beta = 12.566 \quad \text{rad/m}$$

Voltage on the line is superposition of the forward and backward wave voltages giving

$$v(t) = Re\left\{|V^+| e^{j\phi^+} e^{-\alpha x} e^{j(\omega t - \beta x)} + |V^-| e^{j\phi^-} e^{\alpha x} e^{j(\omega t + \beta x)}\right\}$$

$$= Re\left\{2e^{-0.055x} e^{j(\omega t - \beta x)} + 1e^{j\pi/3} e^{0.055x} e^{j(\omega t + \beta x)}\right\}$$

The current on the line is

$$i(t) = Re\left\{\frac{|V^+| e^{j\phi^+}}{Z_0} e^{-\alpha x} e^{j(\omega t - \beta x)} - \frac{|V^-| e^{j\phi^-}}{Z_0} e^{\alpha x} e^{j(\omega t + \beta x)}\right\}$$

$$= Re\left\{\frac{2}{Z_0} e^{-0.055x} e^{j(\omega t - \beta x)} - \frac{1e^{j\pi/3}}{Z_0} e^{0.055x} e^{j(\omega t + \beta x)}\right\}$$

Therefore, at $x = 50$ cm $= 0.5$ m, and $t = 10$ nsec $= 10^{-9}$ sec, we get

$$v(t) = 2.42 \quad \text{V}$$

$$i(t) = 0.148 \quad \text{A}$$

(ii) Finding the maximum of $|v(t)|$ the peak voltage at $x = 0.5$ m will be $V_p = 2.6$ V, and finding the maximum of the current $i(t)$ the peak current will be 0.168 Amp.

Note that the maximum voltage occurs when $\omega t \approx 2.8$ rad, whereas the maximum current occurs when $\omega t \approx 0.5$ rad.

2.4 STANDING WAVES AND IMPEDANCE TRANSFORMATION

We have seen earlier that on a transmission line the voltage and current are represented by superposition of two waves travelling in the opposite directions.

Since the two waves have same frequency and the propagation constant, the resultant is a stationary wave or a standing wave as shown in Fig. 2.7. Depending on the relative amplitudes of the two waves one can either get a full standing wave or a partial standing wave.

To obtain the voltage and current distribution on the line we have to supply the boundary conditions on the line. To evaluate both the arbitary constants V^+ and V^- we need two boundary conditions. However, if we want to analyse only the standing wave behavior on the line one boundary condition is adequate. This boundary condition is represented in terms of the impedance at any point on the transmission line. Generally this 'any point' is taken as the load end of the transmission and all distances on the line are measured from this point towards the generator and are denoted by say 'l'. Then the positive direction of l is opposite to that of x. Replacing x by $-l$ in Eqns (2.28) and (2.29) we get voltage and current on the line as

$$V(l) = V^+e^{\gamma l} + V^-e^{-\gamma l} \tag{2.30}$$

$$I(l) = \frac{V^+}{Z_0}e^{\gamma l} - \frac{V^-}{Z_0}e^{-\gamma l} \tag{2.31}$$

From Fig. 2.7 it is clear that the energy source is connected to the line at AA′ and therefore a voltage or current wave going away from the source, i.e. in $+x$ direction is understandable. But what is the origin of the wave which is travelling in the $-x$ direction? There is no energy source at the other end of the line at BB′. One can then argue that if at all there is a wave travelling in the $-x$ direction , it must be because of the reflection of the forward wave from the load point. It is then important to investigate how the terminating load impedance Z_L (Fig. 2.7) affects the reflection of the wave. Before we do that, however, let us first define a parameter which gives the relative amplitudes of the two waves at any point on the line. This parameter is called the 'Reflection Coefficient' and is defined as

$$\Gamma(l) = \frac{V^-e^{-\gamma l}}{V^+e^{\gamma l}} = \frac{V^-}{V^+}e^{-2\gamma l} \tag{2.32}$$

Γ in general is a complex quantity. As the name suggests, $\Gamma(l)$ defines the complex relation between the reflected voltage wave and the forward or incident voltage wave at any location l on the line. Higher value of Γ indicates more reflection from the load end.

Substituting from Eqn (2.32) into Eqns (2.30) and (2.31) the voltage and current can be written as

$$V(l) = V^+e^{\gamma l}[1 + \Gamma(l)] \tag{2.33}$$

$$I(l) = \frac{V^+}{Z_0}e^{\gamma l}[1 - \Gamma(l)] \tag{2.34}$$

The impedance seen at any point on the line then is

$$Z(l) \equiv \frac{V(l)}{I(l)} = Z_0\frac{1 + \Gamma(l)}{1 - \Gamma(l)} \tag{2.35}$$

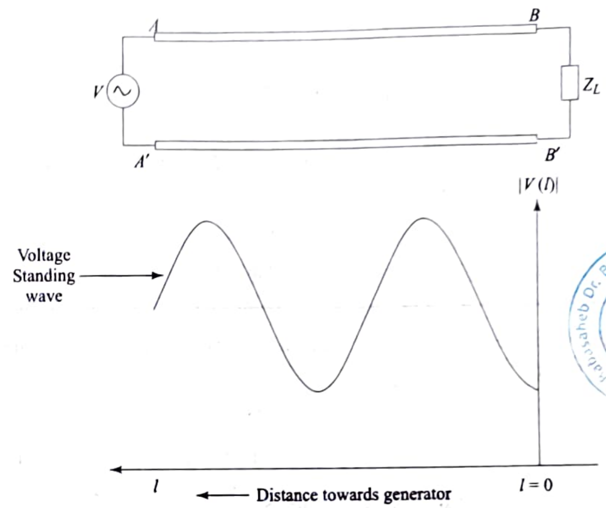


Fig. 2.7 *Standing wave on a transmission line.*

Inverting Eqn (2.35) we get

$$\Gamma(l) = \frac{Z(l) - Z_0}{Z(l) + Z_0} \tag{2.36}$$

Without losing generality let us now take the load end of the line as reference point and measure all distances from this point. Then at $l = 0$, the impedance $Z(l)$ is equal to the load impedance, Z_L.

Substituting $Z(l) = Z_L$ in Eqn (2.36) we get

$$\Gamma(0) \equiv \Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2.37}$$

Γ_L denotes the reflection coefficient at the load-end and from Eqn (2.32) we have

$$\Gamma_L = \frac{V^-}{V^+} = \frac{\text{Reflected voltage at the load-end}}{\text{Incident voltage at the load-end}} \tag{2.38}$$

EXAMPLE 2.8 The transmission line in Example 2.6 is connected to a load impedance $10 + j20\ \Omega$ at 2 GHz. Find the reflection coefficient (i) at the load-end of the line (ii) at a distance of 20 cm from the load.

Solution:

(i) From Example 2.6 we have $Z_0 = 10 + j0.0358\ \Omega$.

The reflection coefficient at the load-end of the line is

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{(10 + j20) - (10 + j0.0358)}{(10 + j20) + (10 + j0.0358)}$$
$$= 0.499 + 0.498j$$

(ii) As solved in Example 2.7, $\gamma = 0.055 + j12.566$ per meter. Therefore

$$\Gamma(l = 20\text{cm}) = \Gamma_L e^{-2\gamma l}$$
$$= (0.499 + j0.498)e^{-2(0.055 + j12.566)0.2}$$
$$= -0.3127 + 0.6149j$$

The voltage and current at any location on transmission line can then finally be written as

$$V(l) = V^{+}e^{\gamma l}\{1 + \Gamma_L e^{-2\gamma l}\} \tag{2.39}$$

$$I(l) = \frac{V^{+}}{Z_0}e^{\gamma l}\{1 - \Gamma_L e^{-2\gamma l}\} \tag{2.40}$$

and the impedance at any point on the line is

$$Z(l) = Z_0\left[\frac{1 + \Gamma_L e^{-2\gamma l}}{1 - \Gamma_L e^{-2\gamma l}}\right] \tag{2.41}$$

Substituting for Γ_L from Eqn (2.37) into Eqn (2.41) and taking $e^{-\gamma l}$ common from numerator and denominator, we get

$$Z(l) = Z_0\left[\frac{e^{\gamma l} + \frac{Z_L - Z_0}{Z_L + Z_0}e^{-\gamma l}}{e^{\gamma l} - \frac{Z_L - Z_0}{Z_L + Z_0}e^{-\gamma l}}\right] \tag{2.42}$$

$$= Z_0\left[\frac{(Z_L + Z_0)e^{\gamma l} + (Z_L - Z_0)e^{-\gamma l}}{(Z_L + Z_0)e^{\gamma l} - (Z_L - Z_0)e^{-\gamma l}}\right] \tag{2.43}$$

Rearranging terms of Z_L and Z_0, we get

$$Z(l) = Z_0\left[\frac{Z_L(e^{\gamma l} + e^{-\gamma l}) + Z_0(e^{\gamma l} - e^{-\gamma l})}{Z_L(e^{\gamma l} - e^{-\gamma l}) + Z_0(e^{\gamma l} + e^{-\gamma l})}\right] \tag{2.44}$$

Since $(e^{x} + e^{-x})/2 = \cosh x$ and $(e^{x} - e^{-x})/2 = \sinh x$, the impedance at a distance l from the load can finally be written as

$$Z(l) = Z_0\left[\frac{Z_L \cosh\gamma l + Z_0 \sinh\gamma l}{Z_L \sinh\gamma l + Z_0 \cosh\gamma l}\right] \tag{2.45}$$

Equation (2.45) is the impedance transformation equation. It shows that if a line is terminated in an impedance Z_L, the impedance seen at a distance l from it is not Z_L but is $Z(l)$. Or in other words, a length l of the transmission line, transforms the impedance from Z_L to $Z(l)$.

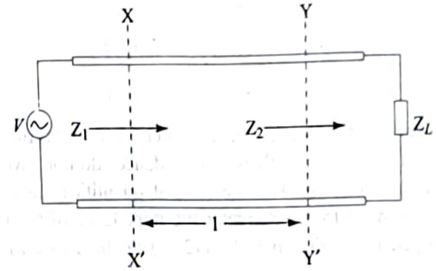


Fig. 2.8 *Impedance transformation on a transmission line.*

It should be noted here that although Eqn (2.45) gives the transformation of the load impedance, there is nothing special about the load impedance. We obtained transformation of load impedance because we defined $l = 0$ at the load end. If we define $l = 0$ at some other location on the line we get transformation of the impedance from that point. We can therefore generalise the impedance transformation equation for any two points on the transmission line.

Let the impedances measured at two locations on the line be Z_1 and Z_2 respectively as shown in Fig. 2.8.

Then from Eqn (2.45) we get

$$Z_1 = Z_0\left[\frac{Z_2 \cosh\gamma l + Z_0 \sinh\gamma l}{Z_2 \sinh\gamma l + Z_0 \cosh\gamma l}\right] \tag{2.46}$$

If we invert Eqn (2.46) we get

$$Z_2 = Z_0\left[\frac{Z_1 \cosh\gamma l - Z_0 \sinh\gamma l}{-Z_1 \sinh\gamma l + Z_0 \cosh\gamma l}\right] \tag{2.47}$$

Since, $\sinh(-\gamma l) = -\sinh\gamma l$ and $\cosh(-\gamma l) = \cosh\gamma l$, Eqn (2.47) can be re-written as

$$Z_2 = Z_0\left[\frac{Z_1 \cosh(-\gamma l) + Z_0 \sinh(-\gamma l)}{Z_1 \sinh(-\gamma l) + Z_0 \cosh(-\gamma l)}\right] \tag{2.48}$$

From Eqns (2.46) and (2.48) it is evident that the two expressions are identical except the sign of l. In Eqn (2.46), l has positive sign whereas in Eqn (2.48) it has negative sign. Therefore it is clear that if the impedance at YY′ (Z_2) is known we can compute the impedance at XX′ by taking l positive in Eqn (2.46) and if the impedance at XX′ (Z_1) is known we can calculate impedance at YY′ by again using Eqn (2.46) but with l negative. Equation (2.46) hence represents a generalized impedance transformation from any point on the line to any other point on the line with appropriate use of the sign for the length l. It should be remembered that all lengths measured towards the generator are positive and all lengths measured away from generator are negative. For analysis of transmission lines, therefore, the knowledge of location of the generator is crucial.

From Eqn (2.46) we can make one more important observation.

Let us re-arrange Eqn (2.46) to get

$$\frac{Z_1}{Z_0} = \frac{\frac{Z_2}{Z_0}\cosh\gamma l + \sinh\gamma l}{\frac{Z_1}{Z_0}\sinh\gamma l + \cosh\gamma l} \tag{2.49}$$

In Eqn (2.49) every impedance has been written as the ratio of the impedance and Z_0. It means that absolute values of impedances do not have any meaning in impedance transformation. The meaningful quantities are the normalized impedances with respect to the characteristic impedance of the line. Denoting a normalized impedance with a 'bar', Eqn (2.49) can be finally written as

$$\bar{Z}_1 = \frac{\bar{Z}_2\cosh\gamma l + \sinh\gamma l}{\bar{Z}_1\sinh\gamma l + \cosh\gamma l} \tag{2.50}$$

EXAMPLE 2.9 A transmission line has the propagation constant $\gamma = 0.1 + j10$/m, and characteristic impedance $Z_0 = 50 + j5\ \Omega$. The line is terminated in an impedance $100 - j30\ \Omega$. Find the impedance at a distance of 1.5 m from the load.

Solution:

The impedance can be obtained using Eqn (2.45). For $l = 1.5$ m,

$$\cosh\gamma l = \cosh\{(0.1 + j10)1.5\} = -0.7683 + 0.0979j$$

$$\sinh\gamma l = \sinh\{(0.1 + j10)1.5\} = -0.1144 + 0.6576j$$

and the impedance is given by

$$Z(l) = Z_0\left[\frac{Z_L\cosh\gamma l + Z_0\sinh\gamma l}{Z_L\sinh\gamma l + Z_0\cosh\gamma l}\right]$$
$$= 57.3175 + 38.6619j\quad \Omega$$

EXAMPLE 2.10 At some location on a long transmission line, the impedance is $100 + j50\ \Omega$. Find the impedance at 50 cm on either side of this location. The transmission line has $Z_0 = 50\ \Omega$ (almost real), and $\gamma = 0.1 + j20$/m.

Solution:

Assume that the generator is on the left side of the line. Let the location at which the impedance is given, is denoted by O, the location 50 cm towards the generator be denoted by A and the location away from the generator be denoted by B. Then by sign convention for the distance $l = +50$ cm for A and $l = -50$cm for B. The impedance at A and B can be obtained using Eqn (2.45) as

$$\gamma l = 0.05 + 10j$$

$$Z(l) = Z_0\left[\frac{Z_L\cosh\gamma l + Z_0\sinh\gamma l}{Z_L\sinh\gamma l + Z_0\cosh\gamma l}\right]$$

$$\Rightarrow \quad Z_A = Z(l = +50\text{ cm}) = 77.73 - 47.70j\ \Omega$$

$$\Rightarrow \quad Z_B = Z(l = -50\text{ cm}) = 29.36 + 38.37j\ \Omega$$

2.5 LOSS-LESS AND LOW-LOSS TRANSMISSION LINES

In the previous sections we developed basic voltage, current and impedance relations for a general tranmission line. However, in practice since the primary use of a tranmission line is to send signals efficiently from one point to another, every effort is made to minimize power loss on the line. Since the power is lost in the resistance of the two conductors and the conductance of the dielectric separating the two conductors, a loss-less tranmission line implies $R = 0$ and $G = 0$. Therefore, for a loss-less transmission line we get (from Eqns (2.11) and (2.27)).

$$\gamma = \sqrt{j\omega L\, j\omega C} = j\omega\sqrt{LC} \tag{2.51}$$

$$Z_0 = \sqrt{\frac{j\omega L}{j\omega C}} = \sqrt{\frac{L}{C}} \tag{2.52}$$

It should be noted that for a loss-less line, γ is purely imaginary and the characteristic impedance is purely real. Since $\gamma = \alpha + j\beta$, pure imaginary γ means $\alpha = 0$ which by definition is the condition for a loss-less line. The phase constant for a loss-less line is $\beta = \omega\sqrt{LC}$.

In practice we do not find a loss-less transmission line. What we however get is a low-loss transmission line. A line is called low-loss provided it has $R << \omega L$ and $G << \omega C$ at the frequency of operation. The line parameter γ for a low-loss line can be written as

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{2.53}$$

$$= \sqrt{j\omega L\left\{1 - j\frac{R}{\omega L}\right\}\, j\omega C\left\{1 - j\frac{G}{\omega C}\right\}} \tag{2.54}$$

$$= j\omega\sqrt{LC}\left\{1 - j\frac{R}{\omega L}\right\}^{1/2}\left\{1 - j\frac{G}{\omega C}\right\}^{1/2} \tag{2.55}$$

Since $R/\omega L$ and $G/\omega C$ are $<< 1$, expanding the brackets binomially and retaining only first order terms, we get,

$$\gamma = \alpha + j\beta = j\omega\sqrt{LC}\left\{1 - j\frac{R}{2\omega L}\right\}\left\{1 - j\frac{G}{2\omega C}\right\} \tag{2.56}$$

Again the product $(R/2\omega L)(G/2\omega C)$ is a second order term and therefore can be neglected, giving

$$\gamma = \alpha + j\beta = j\omega\sqrt{LC}\left\{1 - j\frac{R}{2\omega L} - j\frac{G}{2\omega C}\right\} \tag{2.57}$$

Separating real and imaginary parts we get

$$\text{Attenuation constant: } \alpha = \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} \tag{2.58}$$

$$\text{Phase constant: } \beta = \omega\sqrt{LC} \tag{2.59}$$

From Eqns (2.51) and (2.59) it can be observed that the phase constant of a low-loss line is same as that of the loss-less line for given L and C. A low-loss transmission line therefore can be analyzed like a loss-less transmission line. The use of α is made only in those cases where specifically power loss calculations are to be carried out. In practice if the condition $\alpha << \beta$ is satisfied, we treat the line as a loss-less line. The condition $\alpha << \beta$ implies negligible reduction in the wave amplitude over one wavelength distance on the tranmission line. Until and unless it is specified one can take the liberty of choosing $\alpha = 0$ in the analysis of low-loss transmission lines. In the following examples we investigate the characteristics of a loss-less tranmission line.

EXAMPLE 2.11 A transmission line has $L = 0.25\mu$H/m, $C = 100$pF/m and $G = 0$. What should be the value of R for the line so that the line can be treated as low-loss line? The frequency of operation is 100 MHz.

Solution:

The phase constant of the low-loss line is

$$\beta \approx \omega\sqrt{LC} = 2\pi\sqrt{0.25 \times 10^{-6} \times 100 \times 10^{-12}} \times 10^{8}$$
$$= \pi$$

The attenuation constant from Eqn (2.58) is

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}} \qquad (\text{ since } G = 0\ )$$

$$= \frac{R}{2}\sqrt{\frac{10^{-10}}{0.25 \times 10^{-6}}} = 0.01R \text{ nepers/m}$$

For low-loss line we should have $\beta >> \alpha$. Taking $\alpha < 1\%$ of β, we get

$$0.01R < \frac{\beta}{100}$$

$$\Rightarrow \qquad R < \pi = 3.14\ \Omega/\text{m}.$$

Substituting $\gamma = j\beta$ in Eqns (2.30) and (2.31) we obtain the voltage and current on a loss-less transmission line as

$$V(l) = V^+e^{j\beta l} + V^-e^{-j\beta l} \tag{2.60}$$

$$I(l) = V^+e^{j\beta l} - V^-e^{-j\beta l} \tag{2.61}$$

The reflection coefficient from Eqn (2.32) will be

$$\Gamma(l) = \frac{V^-}{V^+}e^{-j2\beta l} \tag{2.62}$$

From Eqn (2.38) $V^-/V^+ = \Gamma_L$ and we get,

$$\Gamma(l) = \Gamma_L e^{-j2\beta l} \tag{2.63}$$

From Eqn (2.63) we can note that $|\Gamma(l)| = |\Gamma_L|$. It means that the magnitude of the reflection coefficient remains the same at every point on the line and only its phase changes as one traverses the transmission line.
Now let us write the voltage and current on the line in terms of the reflection coefficient as has been done in Eqns (2.39) and (2.40) as

$$V(l) = V^+e^{j\beta l}\left\{1 + \Gamma_L e^{-j2\beta l}\right\} \tag{2.64}$$

$$I(l) = \frac{V^+}{Z_0}e^{j\beta l}\left\{1 - \Gamma_L e^{-j2\beta l}\right\} \tag{2.65}$$

Since Γ_L in general is complex (refer Eqn (2.32)), let us assume that the magnitude of Γ_L is denoted by $|\Gamma_L|$ and its phase is denoted by ϕ. Then by definition

$$\Gamma_L = |\Gamma_L|e^{j\phi} \tag{2.66}$$

Substituting for Γ_L in Eqns (2.64) and (2.65) we get

$$V(l) = V^+e^{j\beta l}\left\{1 + |\Gamma_L|e^{j(\phi - 2\beta l)}\right\} \tag{2.67}$$

$$I(l) = \frac{V^+}{Z_0}e^{j\beta l}\left\{1 - |\Gamma_L|e^{j(\phi - 2\beta l)}\right\} \tag{2.68}$$

As one moves along the transmission line, the length l and therefore the phase $(\phi - 2\beta l)$ changes. Since the second term inside the bracket represents reflected wave, the phase of the reflected wave changes relative to the incident wave as a function of l. At locations where the phase $(\phi - 2\beta l)$ equals multiples of π the exponential term becomes ± 1. At this location the two terms in the bracket are real and they add or subtract directly. When $\exp(j(\phi - 2\beta l)) = +1$ the terms in Eqn (2.67) add whereas they subtract in Eqn (2.68) giving voltage maximum but current minimum. Similarly, when $\exp(j(\phi - 2\beta l)) = -1$ the terms in voltage expression subtract and they add in the current expression giving current maximum but voltage minimum. So we find that on a transmission line when the voltage is maximum the current is minimum and vice versa as shown in Fig. 2.9.

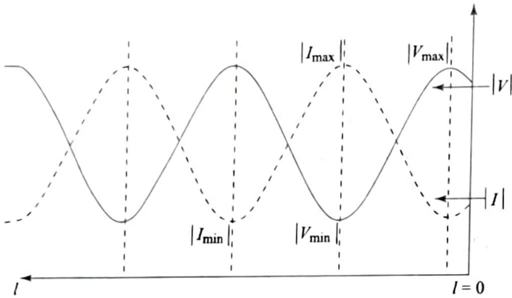


Fig. 2.9 *Voltage and current standing waves.*

The maximum and minimum magnitudes of the voltage are

$$|V|_{max} = |V^+|(1 + |\Gamma_L|) \tag{2.69}$$

$$|V|_{min} = |V^+|(1 - |\Gamma_L|) \tag{2.70}$$

Similarly, the maximum and minimum magnitudes of the current are

$$|I|_{max} = |\frac{V^+}{Z_0}|(1 + |\Gamma_L|) = \frac{|V|_{max}}{Z_0} \tag{2.71}$$

$$|I|_{min} = |\frac{V^+}{Z_0}|(1 - |\Gamma_L|) = \frac{|V|_{min}}{Z_0} \tag{2.72}$$

Remember here that for a loss-less line the characteristic impedance Z_0 is real and hence, $|Z_0| = Z_0$.

At this stage, we can make one more important observation from Eqns (2.71) and (2.72) and that is, at points where voltage or current is maximum or minimum, since the quantity inside the bracket is real, the phase difference between $V(l)$ and $I(l)$ is zero irrespective of the phase of V^+. It should be noted that this phase difference is temporal at a specific location in space. Therefore, the ratio $V(l)/I(l)$ which is the impedance at location 'l', is real. In other words, we can say that at locations where voltage or current is maximum or minimum, the impedance measured on the line is purely resistive (real impedance).

The plot of $|V|$ or $|I|$ as a function of distance l is called the voltage or current standing wave pattern. A characteristic parameter of the standing wave pattern is called the *voltage standing wave ratio* (written in short as VSWR) and is denoted by ρ. The VSWR is defined as

$$\rho = \frac{|V|_{max}}{|V|_{min}} \tag{2.73}$$

VSWR is an easily measurable parameter as it does not require any measurement of phase. It is worthwhile to mention here that the measurement of phase at high frequency is rather a tedious task and at times becomes unreliable.

Substituting from Eqns (2.69) and (2.70) into Eqn (2.73) we have

$$\rho = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|} \tag{2.74}$$

$$\text{or } |\Gamma_L| = \frac{\rho - 1}{\rho + 1} \tag{2.75}$$

The VSWR is an accurate indicator of the reflections on the transmission line. By definition, since $|V|_{max}$ is greater than $|V|_{min}$, ρ is always greater than 1. It could be as high as ∞ when $|V|_{min}$ goes to zero. Since from Eqn (2.75) $\rho = 1$ corresponds to $|\Gamma_L| = 0$, it represents 'no-reflection' condition. Similarly $\rho = \infty$ corresponds to $|\Gamma_L| = 1$ meaning amplitude of reflected wave is equal to that of the incident wave, i.e. full reflection. Since reflected wave carries some power backward, full incident power does not get delivered to the load in the presence of reflection. For efficient power delivery to the load therefore $|\Gamma_L|$ and hence ρ should be as small as possible. VSWR of 1 corresponds to the maximum power transfer efficiency whereas, VSWR of ∞ represents no power delivery to the load.

EXAMPLE 2.12 A loss-less transmission line has 75 Ω characteristic impedance. The line is terminated in a load impedance of $50 - j100$ Ω. The maximum voltage measured on the line is 100 V. Find the maximum and minimum current, and the minimum voltage on the line. At what distance from the load the voltage and current are maximum?

Solution:

The reflection coefficient at the load is

$$\Gamma_L = \frac{50 - j100 - 75}{50 - j100 + 75}$$
$$= 0.2683 - 0.5854j$$
$$\Rightarrow \qquad |\Gamma_L| = 0.644$$

The maximum voltage on the line $|V_{max}| = 100V$ (given). Therefore,

$$100 = |V^+|(1 + |\Gamma_L|)$$

$$\Rightarrow \qquad |V^+| = \frac{100}{1 + |\Gamma_L|} = \frac{100}{1 + 0.644} = 60.83$$

Maximum current $|I_{max}| = |V_{max}|/Z_0 = 100/50 = 2$ A.

Minimum current $|I_{min}| = \frac{|V^+|}{Z_0}(1 - |\Gamma_L|) = 0.433$ A.

Minimum voltage $|V_{min}| = |I_{min}|Z_0 = 21.66$ V.

Maximum voltage occurs when

$$\phi - 2\beta l = \pm 2m\pi \qquad m = 0, 1, 2, 3 ...$$

We choose appropriate sign so that l is positive. Substituting $\beta = 2\pi/\lambda$ we get

$$2\beta l = 2m\pi + \phi$$

i.e. $$l = \frac{(2m\pi + \phi)\lambda}{4\pi} \qquad m = 0, 1, 2, 3...$$

Therefore the voltage maxima occur at

$$l = 0.4\lambda, 0.9\lambda, 1.4\lambda \ldots$$

Now, the voltage maxima and the current minima occur at the same locations. Therefore, the current minima also occur at $l = 0.4\lambda, 0.9\lambda, 1.4\lambda$ ... The voltage minima (also current maxima) are staggered by $\lambda/4$ with respect to the voltage maxima (also current minima). Hence, the voltage minima (current maxima) occur at

$$l + \frac{\lambda}{4} = 0.16\lambda, 0.66\lambda, 1.16\lambda...$$

2.5.1 Impedance Variation on Loss-less Transmission Line

The impedance at any point on the transmission line is

$$Z(l) = \frac{V(l)}{I(l)} = Z_0\left[\frac{1 + \Gamma_L e^{-j2\beta l}}{1 - \Gamma_L e^{-j2\beta l}}\right] \tag{2.76}$$

Substituting for Γ_L from Eqn (2.37) we can write

$$Z(l) = \frac{V(l)}{I(l)} = Z_0\left[\frac{1 + \frac{Z_L - Z_0}{Z_L + Z_0}e^{-j2\beta l}}{1 - \frac{Z_L - Z_0}{Z_L + Z_0}e^{-j2\beta l}}\right] \tag{2.77}$$

Rearranging terms of Z_L and Z_0 and noting that $e^{j\beta l} + e^{-j\beta l} = 2\cos\beta l$ and $e^{j\beta l} - e^{-j\beta l} = 2j\sin\beta l$ we get

$$Z(l) = Z_0\left[\frac{Z_L\cos\beta l + jZ_0\sin\beta l}{Z_0\cos\beta l + jZ_L\sin\beta l}\right] \tag{2.78}$$

Or in terms of normalized impedances,

$$\bar{Z}(l) = \left[\frac{\bar{Z}_L\cos\beta l + j\sin\beta l}{\cos\beta l + j\bar{Z}_L\sin\beta l}\right] \tag{2.79}$$

Where, $\bar{Z}(l) = Z(l)/Z_0$, and $\bar{Z}_L = Z_L/Z_0$.

The maximum impedance occurs where the voltage is maximum and current is minimum and its value is

$$[Z(l)]_{\max} = \frac{V_{\max}}{I_{\min}} = Z_0\left[\frac{1 + |\Gamma_L|}{1 - |\Gamma_L|}\right] = R_{\max}(\text{say}) \tag{2.80}$$

Noting that the quantity inside the square bracket is the VSWR, we get

$$R_{\max} = Z_0\rho \tag{2.81}$$

Similarly, the minimum impedance occurs at a location where the voltage is minimum and the current is maximum, and its value is

$$[Z(l)]_{\min} = \frac{V_{\min}}{I_{\max}} = Z_0\left[\frac{1 - |\Gamma_L|}{1 + |\Gamma_L|}\right] = R_{\min}(\text{say}) \tag{2.82}$$

$$\Rightarrow R_{\min} = Z_0/\rho \tag{2.83}$$

The impedance on a line therefore varies between Z_0/ρ and $Z_0\rho$. This means the impedance value at some point on the line is greater than Z_0 and at other point it is less than Z_0.

EXAMPLE 2.13 A 50 Ω transmission line is connected to a parallel combination of a 100 Ω resistance and a 1 nF capacitance. Find the VSWR on the line at a frequency of 2 MHz. Also find the maximum and minimum resistance seen on the line.

Solution:

The load impedance Z_L is a parralel combination of $R = 100\Omega$ and $C = 1$ nF. Therefore,

$$Z_L = \frac{R(1/j\omega C)}{R + (1/j\omega C)} = \frac{R}{1 + j\omega RC}$$

$$= \frac{100}{1 + j2\pi \times 2 \times 10^6 \times 100 \times 10^{-9}}$$

$$= \frac{100}{1 + 0.4\pi j} = 38.77 - j48.72$$

The reflection coefficient at the load-end of the line is

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} = 0.134 - j0.475$$

$$|\Gamma_L| = 0.494$$

The VSWR $\rho = \dfrac{1 + |\Gamma_L|}{1 - |\Gamma_L|} = 2.95$

The maximim resistance on the line is $R_{\max} = \rho Z_0 = 147.53$ kΩ

The minimum resistance on the line is $R_{\min} = Z_0/\rho = 16.94$ Ω

2.5.2 Important Characteristics of a Loss-less Line

A loss-less transmission line exhibits following important characteristics.

1. Line characteristics repeat every λ/2 As we have seen earlier, the propagation constant $\beta = 2\pi/\lambda$. Let the impedance at some point l on the line be $Z(l)$. The impedance at a distance of $\lambda/2$ from this location from Eqn (2.78) will be

$$Z(l+\lambda/2) = Z_0\left[\frac{Z_L\cos\frac{2\pi}{\lambda}(l+\lambda/2) + jZ_0\sin\frac{2\pi}{\lambda}(l+\lambda/2)}{Z_0\cos\frac{2\pi}{\lambda}(l+\lambda/2) + jZ_L\sin\frac{2\pi}{\lambda}(l+\lambda/2)}\right] \quad (2.84)$$

$$= Z_0\left[\frac{-Z_L\cos\beta l - jZ_0\sin\beta l}{-Z_0\cos\beta l - jZ_L\sin\beta l}\right] \quad (2.85)$$

$$\Rightarrow \quad Z(l+\lambda/2) = Z(l) \quad (2.86)$$

Equation (2.86) tells us that the impedance on a line repeats over every $\lambda/2$ distance. In other words, a study of a $\lambda/2$ section of a line is sufficient to understand the behavior of the whole line.

2. Normalized impedance inverts every $\lambda/4$ Suppose the normalized impedance at a location l is $\overline{Z}(l)$. Then the normalized impedance at a distance of $\lambda/4$ from it from Eqn (2.79) will be

$$\overline{Z}(l+\lambda/4) = \frac{\overline{Z}_L\cos\frac{2\pi}{\lambda}(l+\lambda/4) + j\sin\frac{2\pi}{\lambda}(l+\lambda/4)}{\cos\frac{2\pi}{\lambda}(l+\lambda/4) + j\overline{Z}_L\sin\frac{2\pi}{\lambda}(l+\lambda/4)} \quad (2.87)$$

$$= \frac{-\overline{Z}_L\sin\beta l + j\cos\beta l}{-\sin\beta l + j\overline{Z}_L\cos\beta l} \quad (2.88)$$

$$\Rightarrow \quad \overline{Z}(l+\lambda/4) = \frac{1}{\overline{Z}(l)} \quad (2.89)$$

The normalized impedance on the line therefore inverts every $\lambda/4$ distance. Note that it is the normalized impedance and not the actual impedance. The actual impedance at location l will be $Z_0\overline{Z}(l)$ and at $(l+\lambda/4)$ it will be $Z_0/\overline{Z}(l)$.

3. For load impedance $Z_L = Z_0$, the impedance at any point on the line is Z_0 Suppose the line is terminated in an impedance equal to the characteristic impedance, i.e. $Z_L = Z_0$. Then the impedance at a distance l from the load is

$$Z(l) = Z_0\left[\frac{Z_0\cos\beta l + jZ_0\sin\beta l}{Z_0\cos\beta l + jZ_0\sin\beta l}\right] \quad (2.90)$$

$$= Z_0 \quad (2.91)$$

This indicates that if a line is terminated in the characteristic impedance Z_0, the impedance at every point on the line is also Z_0. This condition is called the 'matched load' condition. If the load impedance is matched to the characteristic impedance i.e. $Z_L = Z_0$, the reflection coefficient Γ is zero everywhere on the line (see Eqn (2.37)). In other words, for the matched load condition there is only forward wave (and no reflected wave) on the line. As discussed earlier, the forward wave always sees the characteristic impedance, and consequently the impedance at every location on the line is Z_0. We can now use this property to give a physical meaning to the characteristic impedance.

Definition of the Characteristic Impedance: *The characteristic impedance of a line is that impedance with which when the line is terminated, the impedance measured on any point on the line is same as the terminating impedance.*

2.6 POWER TRANSFER ON A TRANSMISSION LINE

Consider a loss-less transmission line with characteristic impedance Z_0. Let the line be terminated in a complex load impedance $Z = R + jX \neq Z_0$. Since the load impedance is not equal to the characteristic impedance, there is reflection on the line, and the voltage and the current on the line can be given as

$$V(l) = V^+e^{j\beta l}\{1+\Gamma_L e^{-j2\beta l}\} \quad (2.92)$$

$$I(l) = \frac{V^+}{Z_0}e^{j\beta l}\{1-\Gamma_L e^{-j2\beta l}\} \quad (2.93)$$

Since, the reference point $l = 0$ is at the load end, the power delivered to the load is

$$P_L = \frac{1}{2}\,\mathrm{Re}(VI^*) \text{ at } l = 0 \quad (2.94)$$

$$= \frac{1}{2}\,\mathrm{Re}\left\{[V^+(1+\Gamma_L)]\left[\frac{V^+}{Z_0}(1-\Gamma_L)\right]^*\right\} \quad (2.95)$$

$$= \frac{1}{2}\,\mathrm{Re}\left\{\frac{|V^+|^2}{Z_0}[1-|\Gamma_L|^2+(\Gamma_L-\Gamma_L^*)]\right\} \quad (2.96)$$

Since, the difference of any complex number and its conjugate is purely imaginary, $(\Gamma_L - \Gamma_L^*)$ is a purely imaginary quantity. Therefore, the power delivered to the load is

$$P_L = \frac{|V^+|^2}{2Z_0}\{1-|\Gamma_L|^2\} \quad (2.97)$$

One could have arrived at the same expression using a different argument. The wave which travels towards the load has an amplitude V^+. As discussed earlier, a travelling wave always sees the characteristic impedance irrespective of the terminating load. The power carried by the wave travelling towards the load (also called the incident power) is

$$P_{inc} = \frac{1}{2}Re\{V^+(I^+)^*\} = \frac{1}{2}Re\{V^+(\frac{V^+}{Z_0})^*\} \quad (2.98)$$

$$= \frac{|V^+|^2}{2Z_0} \quad \text{Note: } Z_0 \text{ is real for a loss-less line} \quad (2.99)$$

Now there is a reflected wave on the line, and its amplitude is $V^- = \Gamma_L V^+$. The reflected wave also sees characteristic impedance and therefore the power taken back by the reflected wave is

$$P_{ref} = \frac{|\Gamma_L V^+|^2}{2Z_0} \tag{2.100}$$

$$= \frac{|V^+|^2}{2Z_0}|\Gamma_L|^2 \tag{2.101}$$

Since P_{inc} is the power travelling towards the load, and P_{ref} is the power travelling back from the load, the difference of the two powers is the power delivered to the load. We therefore, have

$$P_L = P_{inc} - P_{ref} \tag{2.102}$$

$$P_L = \frac{|V^+|^2}{2Z_0}\{1 - |\Gamma_L|^2\} \tag{2.103}$$

Equation (2.103) is same as Eqn (2.97). We hence see that one can either use circuit concept or wave concept to calculate the power delivered to the load. Further, it is interesting to note, that since the line is loss-less, the resistive power calculated at any point on the line must be same as that delivered to the load. Let us verify this.

The complex power at any point on the line is

$$P(l) = \frac{1}{2}\{V(l)[I(l)]^*\} \tag{2.104}$$

Substituting for $V(l)$ and $I(l)$ from Eqns (2.92) and (2.93) we get

$$P(l) = \frac{1}{2}\left\{V^+e^{j\beta l}(1 + \Gamma_L e^{-j2\beta l})\right\}\left\{\frac{V^+}{Z_0}e^{j\beta l}(1 - \Gamma_L e^{-j2\beta l})\right\}^* \tag{2.105}$$

$$= \frac{|V^+|^2}{2Z_0}\{1 - |\Gamma_L|^2 + \text{Im}(\Gamma_L e^{-j2\beta l})\} \tag{2.106}$$

Separating real and imaginary parts, we get

$$\text{Re}\{P(l)\} = \frac{|V^+|^2}{2Z_0}\{1 - |\Gamma_L|^2\} = P_L \tag{2.107}$$

$$\text{Im}\{P(l)\} = \frac{|V^+|^2}{2Z_0}Im(\Gamma_L e^{-j2\beta l}) \tag{2.108}$$

It should be noted that the real power is independent of l and is equal to the power delivered to the load, whereas, the imaginary (reactive) power is a function of l and therefore varies along the transmission line. This variation is due to the variation of voltage and current along the line which consequently varies the reactive fields and hence the energy stored along the line. It is important that we clearly differentiate between the energy flow and the energy storage at different locations on the line. The energy flow is same everywhere on the line whereas the energy stored is not.

EXAMPLE 2.14 A 50Ω loss-less transmission line is connected to a load of $50 + j50$ Ω. The maximum voltage measured on the line is 50 V. Find the power delivered to the load and the peak voltage at the load-end of the line.

Solution:

The magnitude of the reflection coefficient

$$|\Gamma_L| = \left|\frac{Z_L - Z_0}{Z_L + Z_0}\right| = \left|\frac{50 + j50 - 50}{50 + j50 + 50}\right|$$

$$= |0.2 + 0.4j| = 0.4472$$

The VSWR $\rho = \frac{1+|\Gamma_L|}{1-|\Gamma_L|} = 2.618$

Since the line is loss-less, the power loss at any point on the line is same as the power delivered to the load. We can, therefore, conviniently choose a point on the line where the voltage is maximum (so the current is minimum). The impedance at this location is real and its value is $\rho Z_0 = 130.9$. The power loss at that point (which is same as power delivered to the load) is

$$P_L = \frac{1}{2}\frac{|V_{\max}|^2}{\rho Z_0} = 9.55 \quad \text{W}$$

Now, the power loss in the load can also be written in terms of the voltage across the load as

$$P_L = \frac{1}{2}\frac{|V_L|^2}{R_L}$$

$$\Rightarrow \quad \text{Peak voltage at the load} |V_L| = 30.9 \quad \text{V}$$

2.6.1 Evaluation of V^+

In the earlier section, we have derived expressions for voltage, current and power in terms of the parameter V^+, the amplitude of the incident voltage wave. In other words, we assumed that the amplitude of the incident voltage wave is known a priori. The question is—how to find the incident wave amplitude. In practice, we get a line which is connected to a voltage or a current generator at one end, and to a load at the other, and we will have to find out the complex V^+ in terms of the generator voltage/current, the line parameters and the load impedance. Let us consider a transmission line of character impedance Z_0 and length l. Let the line be connected to a voltage source, having peak voltage V_s and internal impedance Z_s, as shown in Fig. 2.10(a). Let the line be connected to a load impedance Z_L at other end.

It is clear by now, that the voltage source does not see Z_L connected across it but sees the transformed impedance between terminals AA′. If the transformed

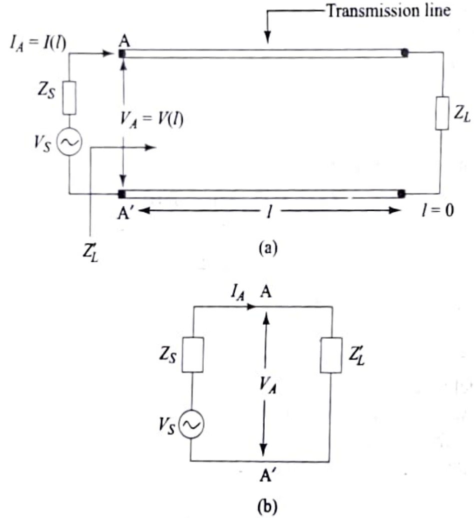


Fig. 2.10 *(a) Transmission line connected to a generator and a load (b) Lumped circuit at the generator-end of the line.*

impedance is denoted by Z_L', we have

$$Z_L' = Z_0 \left[\frac{Z_L \cos \beta l + jZ_0 \sin \beta l}{Z_0 \cos \beta l + jZ_L \sin \beta l} \right] \tag{2.109}$$

The lumped circuit at AA′ will be as shown in Fig. 2.10(b), and the current and voltage at terminals AA′ can be obtained as

$$I_A = \frac{V_s}{Z_S + Z_L'} \tag{2.110}$$

$$V_A = Z_s' I_s = \frac{Z_L' V_s}{Z_S + Z_L'} \tag{2.111}$$

Let us now write the expressions for voltage and current at AA′ in terms of V^+ using transmission line equations, as done in the previous sections.

$$V(l) = V^+ e^{j\beta l} \{1 + \Gamma_L e^{-j2\beta l}\} \tag{2.112}$$

$$I(l) = \frac{V^+}{Z_0} e^{j\beta l} \{1 - \Gamma_L e^{-j2\beta l}\} \tag{2.113}$$

where, $\Gamma_L = (Z_L - Z_0)/(Z_L + Z_0)$.

From Fig. 2.10 (a), it is clear that V_A and $V(l)$ represent the same voltage. Similarly, I_A and $I(l)$ represent the same current. Equating (2.111) and (2.112) we get

$$V^+ e^{j\beta l} \{1 + \Gamma_L e^{-j2\beta l}\} = \frac{Z_L' V_s}{Z_S + Z_L'} \tag{2.114}$$

Solving for V^+ we can obtain the amplitude of the incident voltage wave as

$$V^+ = \frac{Z_L' V_s e^{-j\beta l}}{(Z_s + Z_L')(1 + \Gamma_L e^{-j2\beta l})} \tag{2.115}$$

One can verify that the Eqn (2.115) could have been obtained by equating (2.110) and (2.113) as well.

Since the line is loss-less, the power delivered to the load Z_L is same as the power delivered to the transformed load $Z_L' \equiv R' + jX'$, and is given by

$$P = \frac{1}{2} \mathrm{Re}(V_A I_A^*) \tag{2.116}$$

$$= \frac{1}{2} R' \left| \frac{V_s}{Z_L' + Z_s} \right|^2 \tag{2.117}$$

For maximum power transfer to take place from the source to the load, obviously, we should have $Z_s = Z_L'^*$. Since Z_L' is a function of l and the frequency f, for a given load and source impedance, the maximum power transfer takes place only for a specific frequency and length of the transmission line. The only exception is when both source and load impedances are equal to the characteristic impedance Z_0, in which case there is maximum power transfer irrespective of the length of the line. Since, neither the load nor the source impedance is in our control, usually we use the so called 'matching networks' in front of source or load impedances. The matching networks transform the source and load impedances to the characteristic impedance and the the power transfer becomes independent of the length of the transmission line and the frequency.

It is clear from the above discussion that in high frequency circuits, the characteristic impedance plays a very important role. It is, therefore, desirable to have some standardization of the characteristic impedance to make different circuits compatible with each other. For a co-axial transmission line the standard characteristic impedances are 50 Ω and 75 Ω, and for parallel wire lines (flat ribbon cables) the standard characteristic impedances are 300 Ω and 600 Ω.

EXAMPLE 2.15 A 2.5 m long 75Ω co-axial cable is connected to a generator at one end, and to a load of $75 + j25$ Ω at the other end. The generator has an open-circuit rms voltage of 10 V, and an internal resistance of 50 Ω. Find the power delivered to the load. The frequency of operation is 150 MHz and the velocity of the wave on the cable is 2×10^8 m/sec.

Solution:

Here we have $V_s = 10$ V, and $Z_s = 50$ Ω.

The wavelength on the cable is

$$\lambda = \frac{v}{f} = \frac{2 \times 10^8}{150 \times 10^6} = 1.333 \text{ m}$$

$$\Rightarrow \qquad \beta = \frac{2\pi}{\lambda} = 1.5\pi \text{ rad/m}$$

$$\Rightarrow \qquad \beta l = 1.5\pi \times 2.5 = 11.781 \text{ rad}$$

The transformed impedance at the generator-end is

$$Z_L' = Zo\left(\frac{Z_L \cos\beta l + jZ_o \sin\beta l}{Z_o \cos\beta l + jZ_L \sin\beta l}\right)$$
$$= 54 + 2.96j$$

Now, from the equivalent lumped circuit at the generator-end we can calculate the power supplied to the input of the line. However, since the line is loss-less, power supplied by the generator to the line, i.e. the power delivered to the impedance Z_L', is same as the power delivered to the load. Using Eqn 2.117 we get

$$P_L = Re\{V_A I_A^*\} = 0.4988 \text{ W}$$

(Note that there is no factor $1/2$, since V_A and I_A are rms values.)

2.7 ANALYSIS OF TRANSMISSION LINE IN TERMS OF ADMITTANCES

Many times, especially while analyzing transmission lines connected in parralel, analysis in terms of admittance becomes simpler compared to the analysis in terms of impedance. Therefore we here analyze transmission line characteristics in terms of admittances.

To start with, we define the characteristic admittance Y_0 which is the reciprocal of the characteristic impedance Z_0, i.e.

$$Y_0 = \frac{1}{Z_0} \tag{2.118}$$

The characteristic admittance, therefore, is a ratio of I^+ and V^+ by definition. For a loss-less line since the characteristic impedance Z_0 is real, so is the characteristic admittance Y_0.

Now converting every impedance to the corresponding admittance i.e. $Y_L = 1/Z_L$, $Y(l) = 1/Z(l)$, etc. we can re-write Eqn (2.36) to get reflection coefficient on any location on the line as

$$\Gamma(l) = \frac{1/Y(l) - 1/Y_0}{1/Y(l) + 1/Y_0} = \frac{Y_0 - Y(l)}{Y_0 + Y(l)} \tag{2.119}$$

Similarly, we can write the admittance at any point on the line (using Eqn (2.78)) as

$$Y(l) = Y_0\left[\frac{Y_L \cos\beta l + jY_0 \sin\beta l}{Y_0 \cos\beta l + jY_L \sin\beta l}\right] \tag{2.120}$$

Note that Eqn (2.120) is identical to Eqn (2.78). That is to say that the impedances and admittances are governed by the same transformation relation.

EXAMPLE 2.16 A line of 300Ω charateristic impedance is terminated in an admittance $0.01 + j0.02f$. Find (i) The reflection coefficient at the load-end (ii) reflection coefficient at a distance of 0.2λ from the load-end (iii) impedance at a distance of 0.2λ from the load-end.

Solution:

The characteristic impedance of the line $Z_0 = 300\Omega$. Therefore the characteristic admittance is

$$Y_0 = \frac{1}{Z_0} = \frac{1}{300} = 3.33 \times 10^{-3}\ ℧$$

(i) The reflection coefficient at the load-end is

$$\Gamma_L = \frac{Y_0 - Y_L}{Y_0 + Y_L} = -0.8462 - 0.2308j$$

(ii) Reflection coefficient at a distance of 0.2λ towards the generator is

$$\Gamma(l) = \Gamma_L e^{-j2\beta l} = \Gamma_L e^{-j2\times\frac{2\pi}{\lambda}\times 0.2\lambda}$$
$$= 0.5489 + 0.684j$$

(iii) Impedance at location $l = 0.2\lambda$ on the line is

$$Z(l) = Z_0 \frac{1 + \Gamma(l)}{1 - \Gamma(l)}$$
$$= 103.11 + 611.3j\ \Omega$$

EXAMPLE 2.17 A 50 Ω co-axial cable terminated in a 50 Ω resistance, forms the bus of a network. A network peripheral unit having impedance 75 Ω is connected to the bus at some location through a 50 Ω cable of length 0.3 λ. Find the impedance at a distance of 0.2 λ from the junction. Find the VSWR on the cable.

Solution:

In Fig. 2.11 at the junction we have two impedances Z_1 and Z_2 in parallel. Since the cable has been terminated in its characteristic impedance, Z_1 will be same as the characteristic impedance 50 Ω. Z_2 however will be transformed version of 75 Ω impedance. Hence, we have

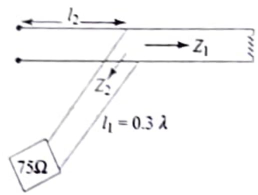


Fig. 2.11 *Bus network.*

$$Z_2 = Z(l_1) = Z_0 \frac{75\cos\beta l_1 + j50\sin\beta l_1}{50\cos\beta l_1 + j75\sin\beta l_1}$$

and $\beta l_1 = 2\pi/\lambda(0.3\lambda) = 0.6\pi = 108^\circ$, giving

$$Z_2 = Z(l_1) = 50\frac{75\cos 108^\circ + j50\sin 108^\circ}{50\cos 108^\circ + j75\sin 108^\circ}$$
$$= 35.2008 + 8.621j\ \Omega$$

Since, at the junction, the two impedances are connected in parallel, the impedance Z is

$$Z = \frac{Z_1 Z_2}{Z_1 + Z_2} = 20.9549 + 2.9389j\ \Omega$$

The impedance at a distance of l_2 from the junction is

$$Z(l_2) = Z_0\frac{Z\cos\beta l_2 + j50\sin\beta l_2}{50\cos\beta l_2 + jZ\sin\beta l_2}$$

and $\beta l_2 = 2\pi/\lambda(0.2\lambda) = 0.4\pi = 72^\circ$, we get

$$Z(l_2) = 50\frac{(20.9549 + 2.9389j)\cos 72^\circ + j50\sin 72^\circ}{50\cos 72^\circ + j(20.9549 + 2.9389j)\sin 72^\circ}$$
$$= 94 + j43.47\ \Omega$$

The magnitude of the reflection coefficient on the line is

$$|\Gamma| = \left|\frac{Z - Z_0}{Z + Z_0}\right| = 0.41$$

$$\Rightarrow \qquad \text{VSWR on the line, } \rho = \frac{1 + |\Gamma|}{1 - |\Gamma|} = 2.396$$

2.8 GRAPHICAL REPRESENTATION OF A TRANSMISSION LINE

In the previous sections we derived analytical expressions for voltage, current, reflection coefficient, impedance, admittance, etc. It is well known that an

image or a graph creates a long lasting impression on mind than text or an equation. A graphical representation of a transmission line helps in pictorial visualization of some of the basic concepts. One can use the graphical means for solving transmission line problems and at times they are also easier as compared to analytical means. However, that is not the whole purpose of graphical representation. A graphical representation provides a good account of tranmission line characteristics in a compact manner. Even while solving transmission line problems analytically, a qualitative cross-check with the graphical model is always helpful in avoiding conceptual mistakes.

2.9 IMPEDANCE SMITH CHART

The graphical representation describes the impedance/admittance characterities of a transmission line. For the time being let us confine our analysis to passive impedances only. Later we will extend it to the admittances. As we have seen earlier, all impedance expressions can be written in terms of normalized impedances. Let us therefore carry out the analysis in terms of

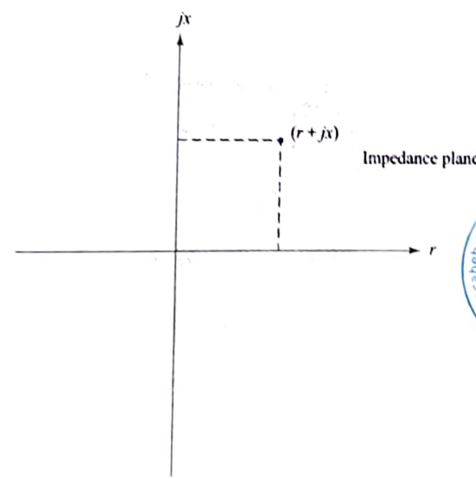


Fig. 2.12 *Complex impedance plane.*

normalized impedances. An impedance $Z = R + jX$ when normalized with the characteristic impedance Z_0, is denoted by $\overline{Z} = r + jx$ where $\overline{Z} = Z/Z_0$, $r = R/Z_0$ and $x = X/Z_0$. The reflection coefficient for a normalized impedance $\overline{Z}$ is then given as

$$\Gamma = \frac{Z - Z_0}{Z + Z_0} = \frac{\overline{Z} - 1}{\overline{Z} + 1} \tag{2.121}$$

$$= \frac{r + jx - 1}{r + jx + 1} = \frac{(r - 1) + jx}{(r + 1) + jx} \tag{2.122}$$

For passive loads, r lies between 0 and ∞ and x lies between $-\infty$ and $+\infty$. A passive load $r + jx$ therefore can be represented by a point in the right half of the complex plane including the imaginary axis as shown in Fig. 2.12.

The reflection coefficient Γ also is complex in general and can be written as

$$\Gamma \equiv u + jv \equiv Re^{j\theta} \tag{2.123}$$

We can note from Eqn (2.121) that there is one-to-one correspondance between $\overline{Z}$ and Γ, and for a normalized impedance $r + jx$ we obtain a unique complex reflection coefficient Γ. Moreover, the magnitude of the reflection coefficient $|\Gamma|$ is always less than or equal to unity. The possible values of Γ are therefore confined within the unit circle in the complex Γ plane as shown in Fig. 2.13.

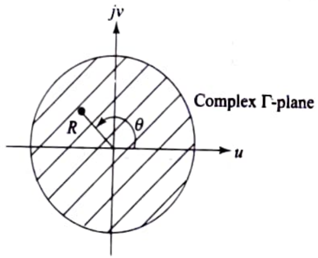


Fig. 2.13 *Complex reflection coefficient plane.*

We, therefore, see that the semi-infinite impedance plane is mapped to the area within the unit circle in the Γ-plane with one to one correspondence between the points in the two planes. One can show that the transformation from $\overline{Z}$-plane to the Γ- plane and vice-versa is a conformal transformation. Let us map a point $\overline{Z} = r + jx$ onto the $\Gamma = u + jv$ plane. Inverting Eqn (2.121) we have

$$\overline{Z} = \frac{1 + \Gamma}{1 - \Gamma} \tag{2.124}$$

$$\Rightarrow \qquad r + jx = \frac{1 + (u + jv)}{1 - (u + jv)} \tag{2.125}$$

Separating the real and the imaginary parts we get two equations of u and v, one in terms of r and other in terms of x, as follows.

$$u^2 - 2\left(\frac{r}{r+1}\right)u + v^2 + \left(\frac{r-1}{r+1}\right) = 0 \tag{2.126}$$

$$u^2 + v^2 - 2u - \left(\frac{2}{x}\right)v + 1 = 0 \tag{2.127}$$

It is interesting to note that both Eqns (2.126) and (2.127) represent circles in the Γ-plane. Since r = constant represents a vertical line in the Z-plane, Eqn (2.126) transforms vertical lines in the Z-plane into circles in the Γ-plane. These circles are called constant resistance circles. Similarly the lines x = constant map into circles in the Γ-plane as given by Eqn (2.127). These are called the constant reactance circles. Remember that only those portions of the circles are of relevance which lie within the unit circle in the Γ-plane. Let us now analyse the characteristics of the transformed circles.

2.9.1 Constant Resistance Circles

The constant resistance circles have their centres at $(\frac{r}{r+1}, 0)$ and radii $(\frac{1}{r+1})$. Figure 2.14 shows the constant resistance circles for different values of r ranging between 0 and ∞.

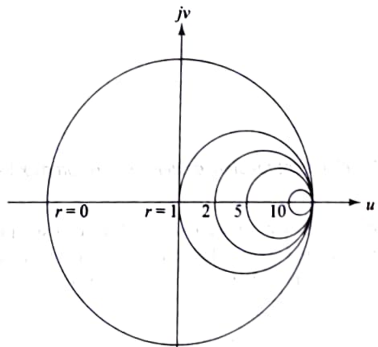


Fig. 2.14 *Constant resistance circles in the complex Γ-plane.*

We can note following things about the constant resistance circles.

(a) The circles always have centres on the real Γ-axis (u-axis).

(b) All circles pass through the point(1, 0) in the complex Γ plane.

(c) For $r = 0$ the center lies at the origin of the Γ plane and it shifts to the right as r increases.

(d) As r increases the radius of the circle goes on reducing and for $r \to \infty$ the radius approaches zero, i.e. the circle reduces to a point.

(e) The outermost circle with center (0, 0) and radius unity, corresponds to $r = 0$ or in other words represents reactive loads only.

(f) The right most point on the unit circle $(1, 0)$ represents $r = 0$ as well as $r = \infty$.

2.9.2 Constant Reactance Circles

The constant reactance circles have their centers at $(1, \frac{1}{x})$ and radii $(\frac{1}{x})$. The centres for these circles lie on a vertical line passing through $(1, 0)$ point in the Γ-plane. The constant reactance circles are shown in Fig. 2.15 for different values of x.

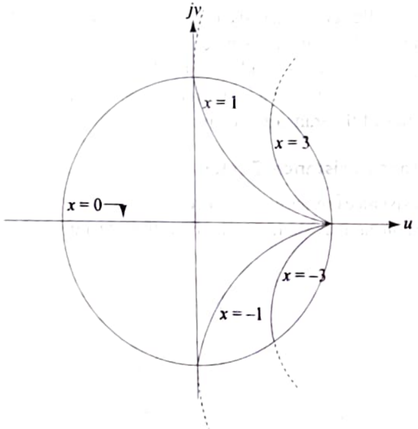


Fig. 2.15 *Constant reactance circles in the complex Γ-plane.*

Note again that only those portions of the circles are of significance which lie within the unit circle in the Γ-plane. The curves shown dotted do not correspond to any passive load impedance. We can note the following points about constant reactance circles:

(a) These circles have their centers on a vertical line passing through point $(1, 0)$.

(b) For positive x the center lies above the real Γ-axis and for negative x, the center lies below the real Γ-axis.

(c) For $x = 0$, the center is at $(1, \pm\infty)$ and radius is ∞. This circle therefore represents a straight line.

(d) As the magnitude of the reactance increases the center moves towards the real Γ-axis and it lies on the real Γ-axis at (1,0) for $x = \pm\infty$.

(e) As the magnitude of the reactance increases, the radius of the circle $(\frac{1}{x})$ decreases and it approaches zero as $x \to \pm\infty$.

(f) All circles pass through the point $(1, 0)$.

(g) The real Γ-axis (u-axis) corresponds to $x = 0$ and therefore represents real load impedances, i.e. purely resistive impedances.

(h) The right most point on the unit circle, $(1, 0)$, corresponds to $x = 0$ as well as $x = \pm\infty$.

2.9.3 The Smith Chart

The Smith chart is a graphical figure which is obtained by superposing the constant resistance and the constant reactance circles within the unit circle in the complex Γ-plane. A Smith chart is shown in Fig. 2.16. Since we have mapped here the impedances to the Γ-plane, let us call this the Impedance Smith chart.

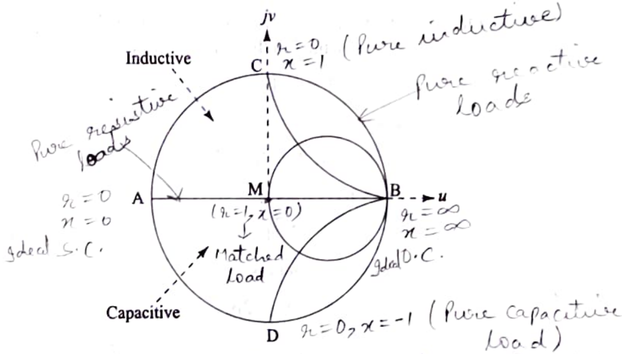


Fig. 2.16 *Smith chart: Superposition of constant resistance and constant reactance circles in the complex Γ-plane.*

Generally the u, v axes are not drawn on the Smith chart. However one should not forget that the Smith chart is a figure which is drawn on the complex Γ-plane with its center as origin. The intersection of constant resistance and constant reactance circles uniquely defines a complex load impedance in the Γ-plane. Let us identify some special points on the Smith Chart.

(a) The left most point A on the Smith chart corresponds to $r = 0, x = 0$ and therefore represents ideal short-circuit load.

(b) The right most point B on the Smith chart corresponds to $r = \infty, x = \infty$ and therefore represents ideal open circuit load.

(c) The center of the Smith chart M , corresponds to $r = 1, x = 0$ and hence represents the matched load.

(d) Line AB represents pure resistive loads and the outermost circle passing through A and B represents pure reactive loads.

(e) The upper most point C represents a pure inductive load of unity reactance and the lower most point D represents a pure capacitive load of unity reactance.

(f) In general the upper half of the Impedance Smith chart represents the complex inductive loads and the lower half represents the complex capacitive loads.

2.9.4 Constant VSWR Circles

We have seen earlier that the voltage reflection coefficient at any location l from the load is given as (see Eqn (2.63))

$$\Gamma(l) = \Gamma_L e^{-j2\beta l} \tag{2.128}$$

where Γ_L is the complex reflection coefficient at the load and is given by Eqn (2.37). Let the Γ_L be represented in the polar form as

$$\Gamma_L = |\Gamma_L| e^{j\theta_L} \tag{2.129}$$

Then the reflection coefficient $\Gamma(l)$ will be

$$\Gamma(l) \equiv Re^{j\theta} = |\Gamma_L| e^{j(\theta_L - 2\beta l)} \tag{2.130}$$

The magnitude of the reflection coefficient is

$$R \equiv |\Gamma(l)| = |\Gamma_L| \tag{2.131}$$

and the phase of the reflection coefficient is

$$\theta \equiv \theta_L - 2\beta l \tag{2.132}$$

As we change the value of l, i.e. as we move along the tranmission line the magnitude of $\Gamma(l)$ remains same (R = constant) but its phase varies linearly. As l increases, i.e. as we move towards the generator, the phase of the reflection coefficient $\Gamma(l)$ becomes more negative. In the complex Γ-plane therefore a point $Re^{j\theta}$ moves clockwise on a circle with center at (0,0) and radius $|\Gamma_L|$ as shown in Fig. 2.17.

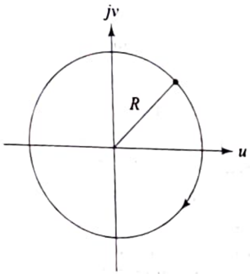


Fig. 2.17

All points on this circle have same $|\Gamma| = |\Gamma_L|$. Now since the VSWR on the line is

$$\rho = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|} \tag{2.133}$$

all the points on this circle have same VSWR. Hence, these circles are called the 'constant VSWR circles'. The constant VSWR circles are shown in Fig 2.18.

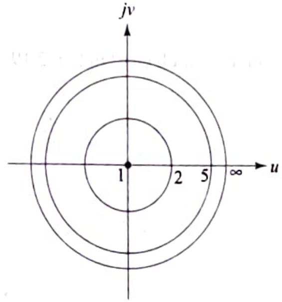


Fig. 2.18 *Constant VSWR circles drawn in the complex Γ-plane.*

We can make following observations about the constant VSWR circles:

(a) All the circles have same center, the origin of the complex Γ-plane.

(b) The origin in the Γ-plane represents $|\Gamma_L| = 0$ or $\rho = 1$. As we move radially outwards the $|\Gamma_L|$ and hence ρ increases monotonically and for the outermost unity circle, $|\Gamma_L| = 1$ and $\rho = \infty$.

(c) The origin corresponds to the condition $|\Gamma_L| = 0$, i.e. no reflection on the line. This point represents the best matching of the load as there is no reflected power on the line. For the outer most circle, since $|\Gamma_L| = 1$, we get the worst impedance matching as the entire power is reflected on the transmission line. We can therefore make a general statement that closer is the point to the origin of the Γ-plane, (i.e. the center of the Smith chart) better is the impedance matching.

(d) As we have defined earlier, the $+l$ indicates a distance towards the generator. As l becomes more positive, θ decreases and the point moves clockwise on the constant VSWR circle. If we move away from the generator, l becomes negative and then the point on the circle moves in the anticlockwise direction.

For analysing a transmission line problem graphically, the constant VSWR circles are to be superimposed or drawn on the Smith chart. For the sake of visual clarity the constant VSWR circles are not permanently drawn on the Smith chart. As and when required the user draws an appropriate constant VSWR circle on the Smith chart.

EXAMPLE 2.18 Draw following on the Smith chart. The normalizing impedance is 50Ω. (a) $50 + j75\Omega$ (b) $10 + j0\Omega$ (c) $0 - j80\Omega$ (d) $\Gamma = 0.3\angle 60^o$ (e) Constant VSWR circle for $\rho = 2.5$ (f) Minimum resistance point on the constant VSWR circle for $\rho = 1.5$.

Solution:

The points are marked on the Smith chart in Fig. 2.19.

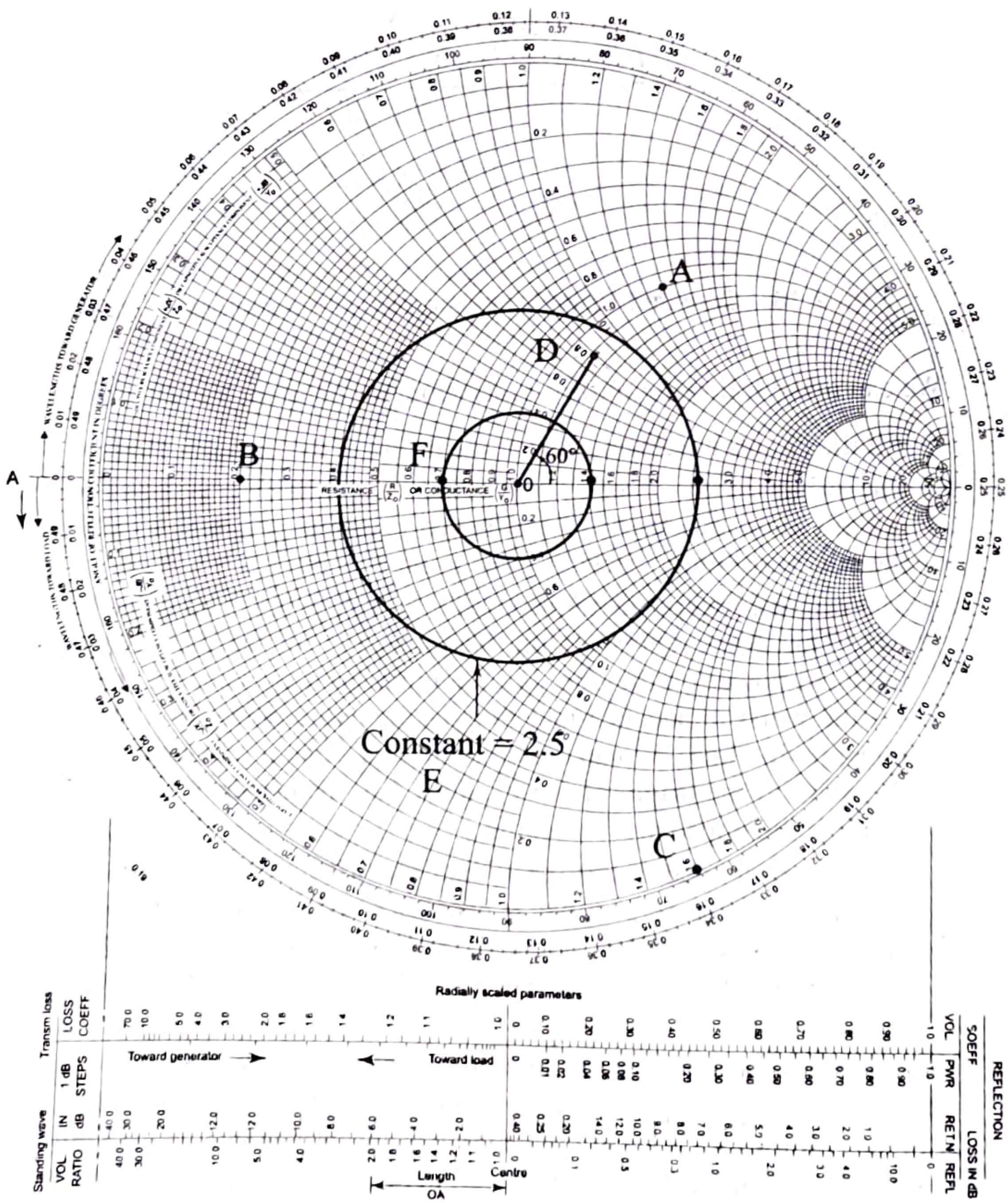


Fig. 2.19 *Smith chart for Example 2.18*

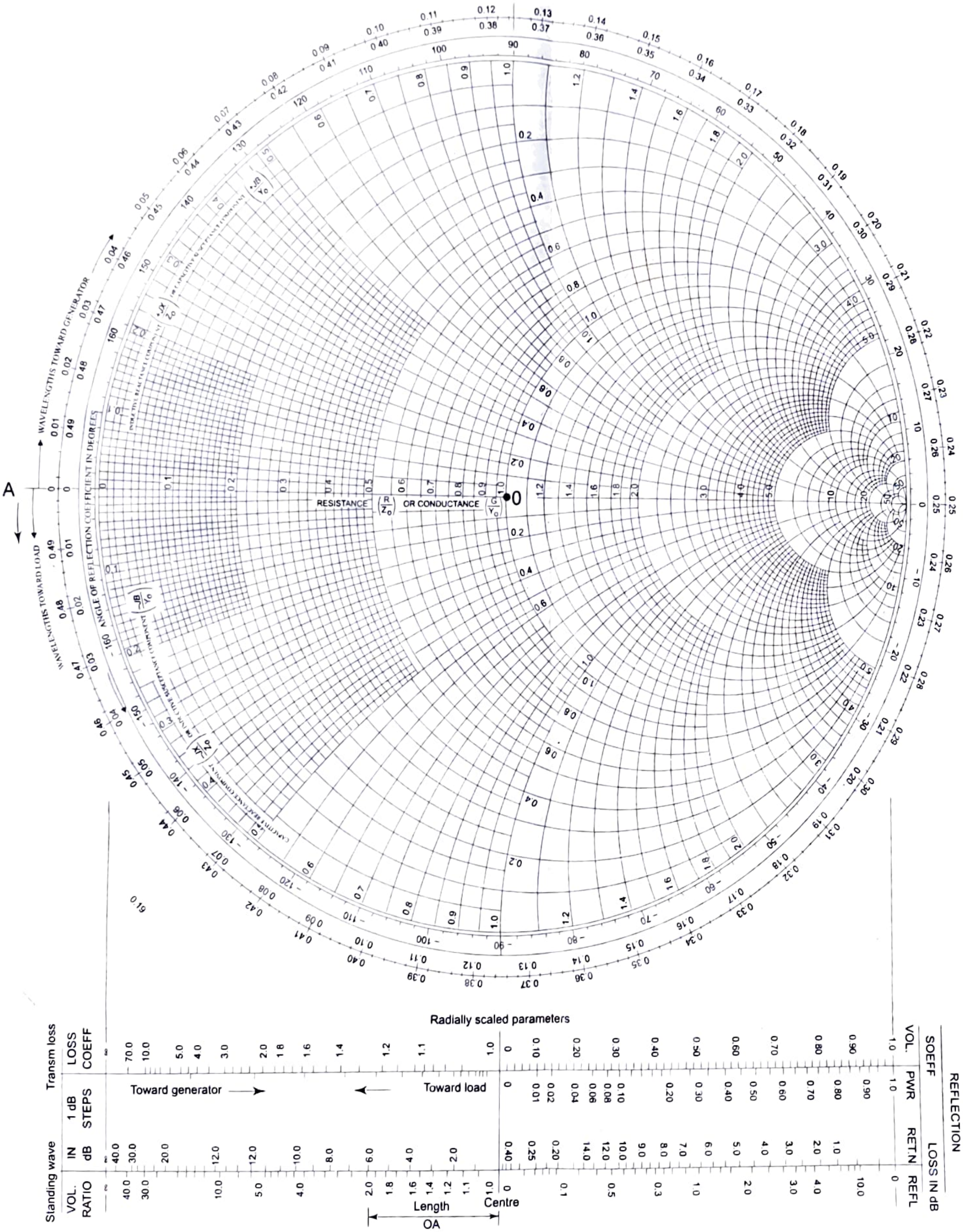


Fig. 2.20 *Standard printed Smith chart*

2.10 TRANSMISSION LINE CALCULATIONS WITH THE HELP OF THE SMITH CHART

The Smith chart is readily availaible in the printed form without constant VSWR circles drawn on it (see Fig. 2.20 as an Extended sheet). Generally the circumference of the Smith chart is marked with the length normalized to the wavelength. Arrows indicating 'towards generator' and 'towards load' are also normally indicated. However, even if this information is not printed, one can workout the direction from the first principles. As stated in the previous section, a clockwise rotation on the Smith chart indicates movement 'towards generator' and an anti-clockwise rotation indicates a movement 'away from the generator'. One complete rotation around the Smith chart corresponds to a phase change of $2\beta l = 2\pi$. Since $\beta = \frac{2\pi}{\lambda}$, one rotation is equal to a distance of $\frac{\lambda}{2}$, i.e. half the wavelength. One can verify the transmission line characteristic which states that impedance characteristics repeat every $\frac{\lambda}{2}$ distance on the line, because after one rotation on the Smith chart the point retraces the same circle and the impedance values repeat.

Smith chart is a very useful tool for solving transmission line problems. A variety of calculations can be carried out using the Smith chart without getting into complex computations. The Smith chart finds application in obtaining following quantities on a transmission line.

(a) Load Reflection Coefficient Let us find the reflection coefficient for a load impedance $R + jX$. First normalize the impedance with the characteristic impedance Z_0 to get $r + jx \equiv \frac{R+jX}{Z_0}$. Identify the constant resistance and the constant reactance circles corresponding to r and x respectively. Intersection of the two circles, marks the load impedance $r + jx$ on the Smith chart as point P (see Fig. 2.21). Measure the radial distance of P from the centre of the chart M. This is the magnitude of the load reflection coefficient $|\Gamma_L|$. The angle which the radius vector MP makes with the $+u$-axis is the phase of the load reflection coefficient θ_L. Note here that the Smith chart should be placed in such a way that the most clusterred portion of the chart lies on the right side. The horizontal line towards right then indicates the real $+u$-axis.

(b) Reflection Coefficient at a Distance from the Load Let the transmission line be terminated in a load impedance Z_L. Let us find the reflection coefficient at a distance l from the load using the Smith chart.

First, mark the load impedance as described in (a). Draw the constant VSWR circle passing through P. Rotate the radius vector MP by an angle $2\beta l$ in the clockwise direction to get point Q (see Fig. 2.22). Radial distance MQ gives the magnitude of the reflection coefficient, $|\Gamma(l)|$. Angle which the radius vector MQ makes with the $+u$-axis gives the phase of $\Gamma(l)$, θ.

(c) Transformed Impedance at a Distance from the Load Suppose, we have to obtain the transformed impedance at a distance l from the load, carry out the steps in (a) and (b) to get to point Q. Now instead of measuring reflection coefficient, identify the constant resistance and the constant reactance

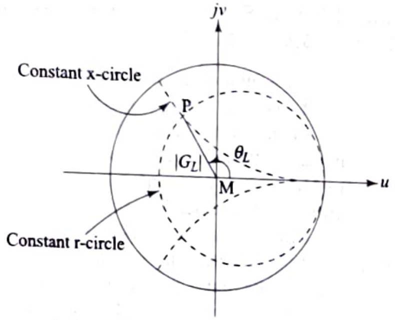


Fig. 2.21 *Smith chart as point P*

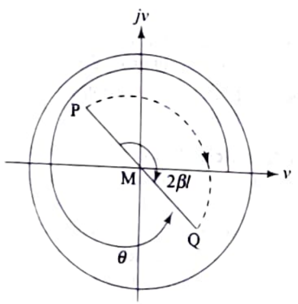


Fig. 2.22

circles passing through Q. They provide the transformed normalized resistance and reactance $r(l)$ and $x(l)$ respectively. Multiplying by Z_0 we get the transformed impedance $Z(l) = Z_0(r(l) + jx(l))$. The same procedure is used for transforming impedance from any point on the line to any other point. However if the distance l is away from the generator, it should be treated negative and hence the rotation of the radius vector must be by $2\beta l$ in the anti-clockwise direction.

(d) VSWR on the Line As we have seen earlier if the load impedance Z_L is not equal to Z_0 there is a standing wave on the transmission line. We also note from Eqn (2.81) that the maximum impedance seen on the line $R_{max} = Z_0\rho$. This means that the maximum normalized impedance $r_{max}(= R_{max}/Z_0)$ measured on the line is nothing but the VSWR, ρ. The task of finding ρ is then very simple. Mark the normalized load impedance $\overline{Z}_L = r + jx$ on the Smith chart. Draw the constant VSWR circle passing through the marked point. The impedance corresponding to the intersection point of the constant VSWR circle and the $+u$-axis is $r_{max} = \rho$

(see point T in Fig. 2.23). So, just by reading the value r_{max} from the chart we obtain the VSWR on the line.

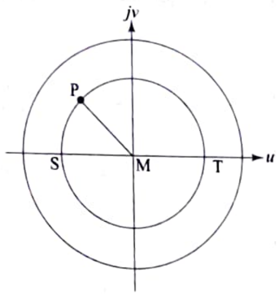


Fig. 2.23

(e) Location of Voltage Maximum or Minimum As we have discussed earlier, at the location of voltage maximum the impedance is maximum (R_{max}), and at the location of voltage minimum the impedance is minimum (R_{min}). Hence point T indicates the location of voltage maximum and point S indicates the location of voltage minimum in Fig. 2.23. To find distance of these points from the load, measure the angle between the load point P and points T and S respectively in the clockwise direction.The angle PMT in the clockwise direction when divided by 2β gives the distance of the voltage maximum from the load, l_{max}. Similarly, one can obtain distance of voltage minimum, l_{min} by measuring angle PMS in clockwise direction from P to S. Alternatively one can make use of the fact that the maximum and minimum voltage are seperated by a distance of $\lambda/4$.

(f) Identifying the Type of Load From the above discussion we can make one more observation. If the load is inductive, point P lies in the upper half of the Smith chart. Then while moving clockwise on the constant-VSWR circle, we first meet point T and then S. In other words, for inductive loads, voltage maximum is closer to the load point than the voltage minimum. Exactly the opposite occurs for capacitive loads i.e. the voltage minimum is closer to the load than the maximum. We can therefore quickly identify the load looking at the standing wave pattern. If the pattern is like the one given in Fig. 2.24(a), that is the voltage drops towards the load, the load is inductive. Similarly if the pattern is like that in Fig. 2.24(b), that is, the voltage rises towards load, the load is capacitive. Consequently if the voltage is maximum or minimum at the load, the load impedance is purely resistive (Fig. 2.24(c)). For the purely reactive loads the point P in Fig. 2.3 will lie on the outermost circle making $|\Gamma_L| = 1$ and $V_{min} = 0$. The pattern for pure inductive load will be like that in Fig. 2.24(d) and that for the pure capacative load will be like that in Fig. 2.24(e).

The above explanation clearly demonstrates the effective use of the Smith chart for the transmission line calculations. The important thing is, most of the impedances and the standing wave quantities can be calculated without any major computation.

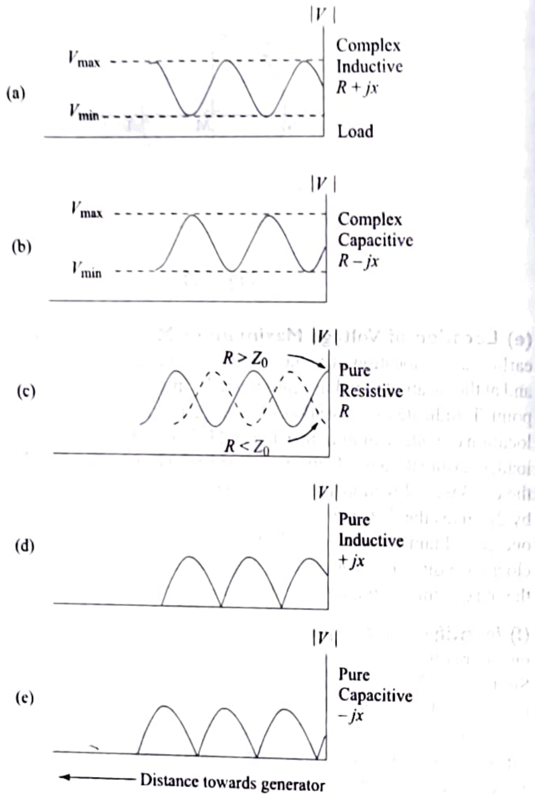


Fig. 2.24 *Voltage standing wave patterns for various types of load impedances.*

EXAMPLE 2.19 A 50Ω line is terminated in a load impedance 25 + j35Ω. With the help of the Smith chart find (i) Reflection coefficient in cartesian and polar form (ii) Reflection coefficient and impedance at a distance of 0.2λ from the load-end of the line (iii) VSWR on the line.

Solution:

(i) The normalized impedance is

$$\overline{Z} = r + jx = \frac{25 + j35}{50} = 0.5 + j0.7$$

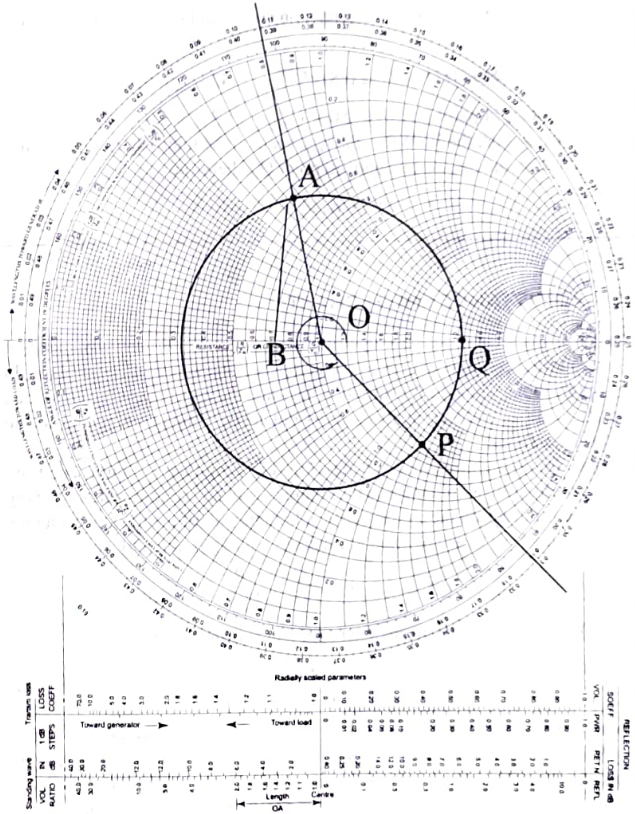


Fig. 2.25

Refer to Fig. 2.25. $\overline{Z}$ is denoted by point A on the chart. All distances on the Smith chart are normalized with respect to the radius of the chart. $|\Gamma_L|$ = distance OA = 0.52.

The complex reflection coefficient is

$$\text{Polar form: } \Gamma_L = |OA|e^{j\theta_L} = 0.52e^{j100.52^\circ}$$

$$\text{Cartesian form: } \Gamma_L = OB + jAB = -0.095 + j0.51$$

(ii) Draw constant VSWR circle passing through point A. The distance 0.2λ towards the generator corresponds to $\theta = 2\beta l = 2\frac{2\pi}{\lambda}0.2\lambda = 0.8\pi = 144^\circ$ in the clockwise direction. Moving on the constant VSWR circle by 144° from point A, we reach to point P. The reflection coefficient corresponding to point P is

$$\Gamma = |OP|e^{j\theta_l} = 0.51e^{j326^\circ} = 0.42 - j0.28$$

The impedance at point P is, $Z = 50 \times (1.4 - j1.37) = 70 - j68.5$
(iii) We know that, the VSWR is the r_{max} seen on the line. The r_{max} and hence the VSWR on the line corresponds to point Q. Therefore

$$\text{VSWR}, \rho = 3.1$$

2.11 ADMITTANCE SMITH CHART

In the previous section we developed the impedance Smith chart by mapping complex normalized impedances to the complex reflection coefficient plane. In many applications when the transmission lines and the impedances are connected in parallel, the admittance analysis turns out to be simpler compared to the impedance analysis.

In the following sections, we develop the admittance Smith chart by mapping normalized admittances to the complex Γ-plane. Normalization of every admittance is done with the characteristic admittance Y_0 of the transmission line. An admittance $Y = G + jB$ when normalized with Y_0 is noted by

$$\overline{Y} = g + jb = \frac{G}{Y_0} + j\frac{B}{Y_0}$$

Writing Eqn (2.121) in the normalized admittance form we get

$$\Gamma = \frac{1/Y - 1/Y_0}{1/Y + 1/Y_0} = \frac{1 - \overline{Y}}{1 + \overline{Y}} \tag{2.134}$$

To obtain the admittance Smith chart we can carryout the mapping of points $g + jb$ to the Γ-plane on the line identical to that given in the previous section. However, without going through the same algebraic steps, let us use the basic concepts to derive the admittance Smith chart from the impedance Smith chart.

Equation (2.134) can be written as

$$\Gamma = -\frac{\overline{Y} - 1}{\overline{Y} + 1} = \frac{\overline{Y} - 1}{\overline{Y} + 1}\angle\pi \tag{2.135}$$

Now, if we take normalized impedance $\overline{Z}$ equal to $\overline{Y}$ i.e. $r = g$ and $x = b$, we get Γ for $\overline{Z}$ which is 180^o out of phase with respect to the Γ for $\overline{Y}$. It means

that for same numerical values, if the normalized load is impedance we get some point P on the Γ plane and if the load is admittance we get point P′ which is diagonally opposite to P on the Γ-plane (see Fig. 2.26). P′ is obtained by rotating P by 180° around the origin of the Γ plane. Since, this is true for every $\overline{Z}$ and $\overline{Y}$, all constant resistance and constant reactance circles when rotated by 180° around the origin of the Γ-plane give corresponding constant conductance (constant-g) and constant susceptance (constant-b) circles respectively.

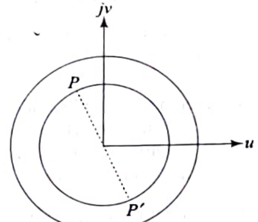


Fig. 2.26 *Impedance to admittance conversion.*

The Admittance Smith chart therefore appears as shown in Fig. 2.27.

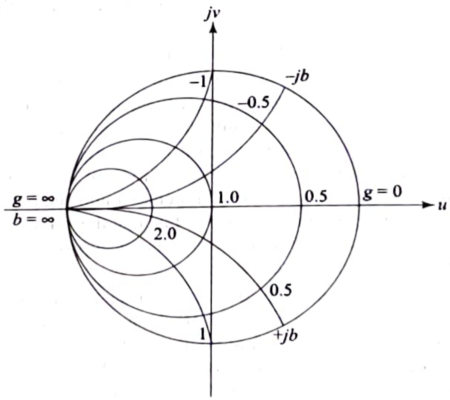


Fig. 2.27 *Constant conductance and constant susceptance circles drawn on the complex Γ-plane.*

The admittance Smith chart therefore is obtained by rotating the impedance Smith chart by 180° and replacing r by g and x by b. Since it is just a matter of rotation, there is no need to have separate Smith charts for impedance and admittance. Generally the Smith chart is used with orientation as shown in Fig. 2.20, and while using it as admittance chart, the u and v axes are rotated by 180°. In Fig. 2.18 therefore the u-axis will point to left and jv-axis will point downwards if the chart is used as the admittance chart. The admittance chart then would be as shown in Fig. 2.28.

Although Figs 2.20 and 2.28 appear identical, the following points should be kept in mind while making their use for transmission line calculations.

(a) While calculating phase of the reflection coefficient from the admittance Smith chart the phase must be measured from the rotated u-axis as shown in Fig. 2.28.

(b) Although r and x can be interchanged with g and b respectively and a point (r, x) and (g, b) will have the same spatial location on the Smith chart for $r = g$ and $x = b$, physical conditions corresponding to the two will not be identical. Let us analyse some specific examples.

- Upper half of the Smith chart with $+jx$ represents inductive loads whereas $+jb$ represents capacitive loads.
- Point A in Fig. 2.16 is $r = 0, x = 0$ as well as $g = 0, b = 0$. But $r = 0, x = 0$, represents short circuit load whereas, $g = 0, b = 0$, represents an open circuit load. The point A therefore represents the short circuit in the impedance chart whereas it represents the open circuit in the admittance chart.
- Similarly Point B in Fig. 2.16 represents the open circuit for the impedance chart and in admittance chart it represents the short circuit.
- In Fig. 2.23, point T corresponds to the voltage maximum if the chart is the impedance chart, and a voltage minimum if the chart is the admittance chart. The opposite is true for point S. Now, since the voltage maximum coincides with the current minimum and vice versa, the point T in admittance Smith chart represents the location of the current maximum and point S represents location of the current minimum. So we find that the voltage standing wave pattern and the impedance have the same relationship as the current standing wave pattern and the admittance.
- As we have seen, the reflection coefficients for same normalized impedance and admittance values are 180° out of phase. Therefore any normalized impedance can be converted to normalized admittance and vice-versa by taking a diagonally opposite point on the constant VSWR circle. In Fig. 2.26, P′ gives normalized admittance corresponding to the normalized impedance at P. We can therefore switch between admittance and impedance Smith charts freely without any additional computation.

All transmission calculations using admittance Smith chart are identical to

that with the impedance Smith chart (described in the previous section) except the modifications mentioned earlier.

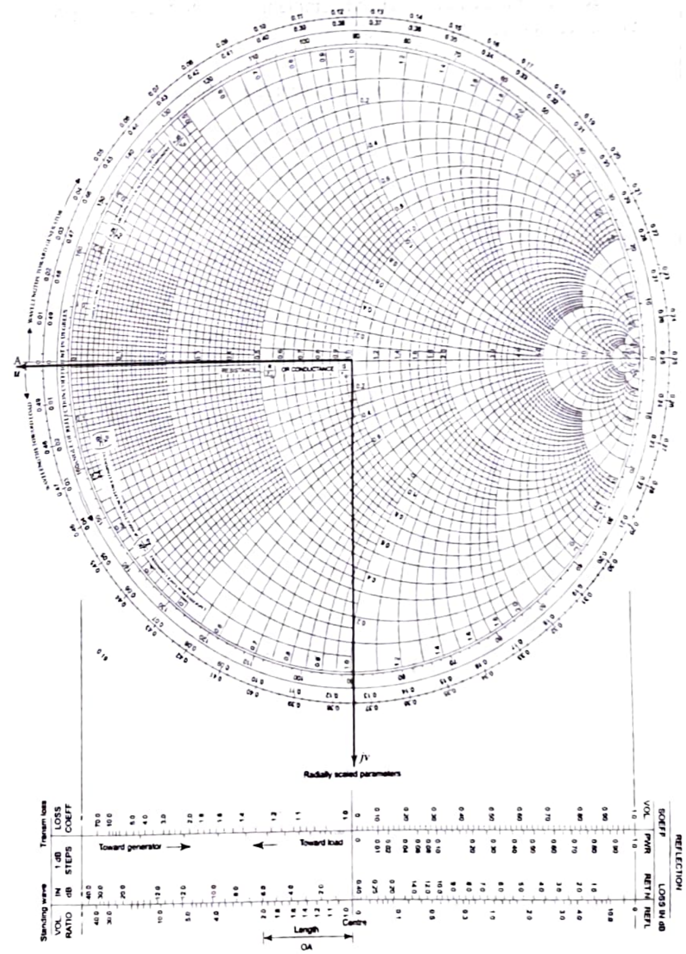


Fig. 2.28 *For Admittance Smith chart, the coordinate axes of the complex Γ-plane is rotated by 180°.*

EXAMPLE 2.20 A line is terminated in a normalized admittance $0.2 - j0.5$. Find the location of the voltage maximum from the load-end. Also find the reflection coefficient, normalized admittance, and normalized impedance at a distance of $0.12\,\lambda$ from the load.

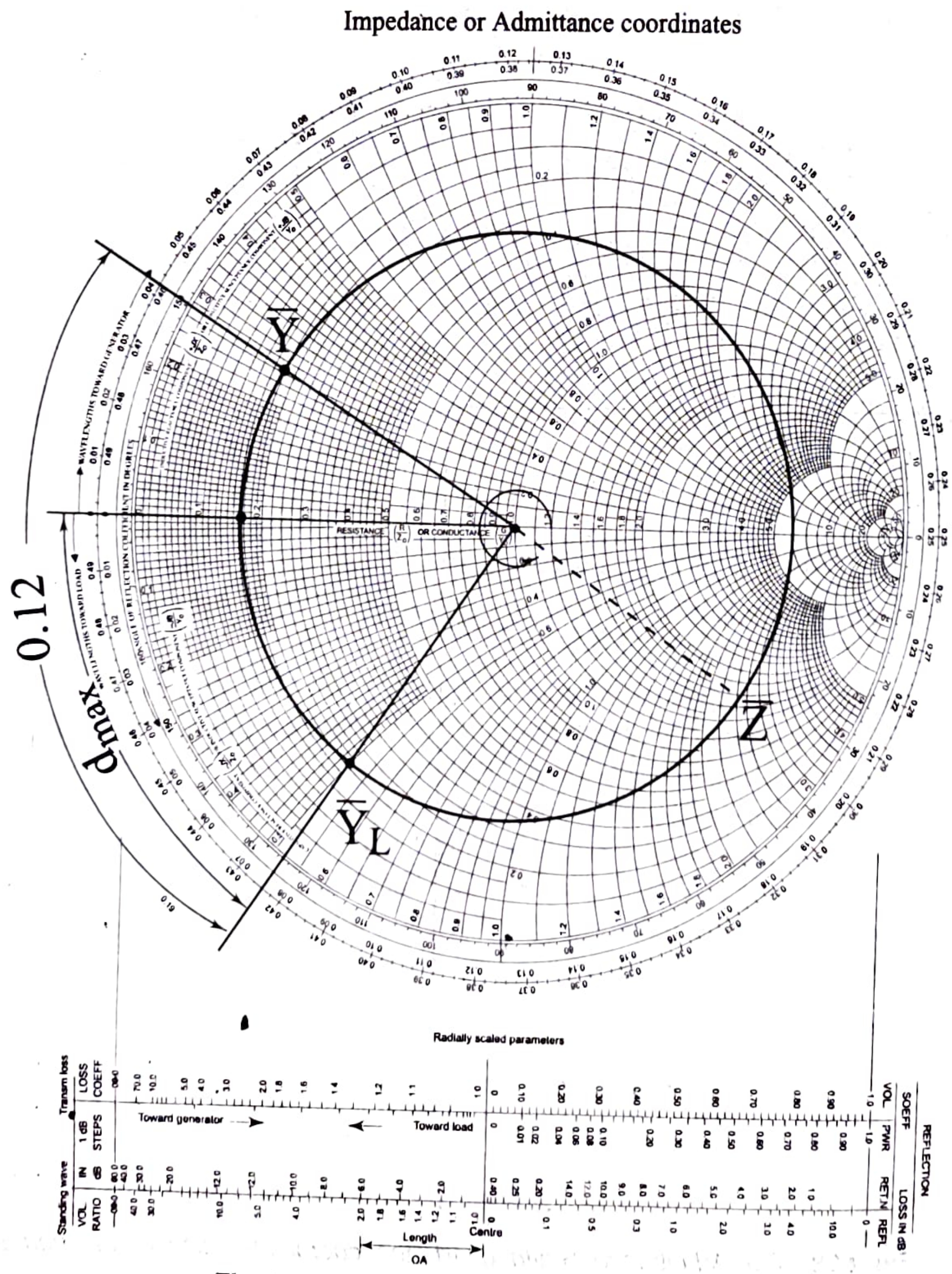


Fig. 2.29 *Smith chart for Example 2.20*

Solution:

Refer to the Smith chart in Fig. 2.29.

First take the Smith chart as the admittance chart. Mark the load admittance $\overline{Y}_L = 0.2 - j0.5$ on the chart. Draw the constant VSWR-circle passing through $\overline{Y}_L$.

Note that the voltage maximum occurs where the admittance is minimum. The left most point on the constant VSWR-circle corresponds to the voltage maximum. The distance of the voltage maximum is $d_{\max} = 0.076\ \lambda$.

Following the procedure identical to that in Example 2.19 and using the chart as the admittance chart, we get at a distance of 0.12 λ,

$$\overline{Y} = 0.18 + j0.28$$

$$\Gamma = 0.706\angle 328°$$

The normalized impedance $\overline{Z}$ is the diagonally oppsite point to $\overline{Y}$ on the constant VSWR circle. Hence, the normalized impedance is

$$\overline{Z} = 1.6 - j2.6$$

2.12 APPLICATIONS OF TRANSMISSION LINE

In the previous sections, we developed the understanding of basic characteristics of a transmission line. The impedance transformation property which a transmission line has, can be utilized in a variety of applications. The transmission line is therefore, not merely used for transporting power from one point to another but has some important applications as discussed in the following sections.

2.12.1 Measurement of Unknown Impedance

At high frequencies the measurement of an impedance is rather a difficult task. This is due to the fact that as the frequency increases, direct measurement of the signal phase becomes more and more difficult and at times impossible. A transmission line can be of great use in this situation. As we have seen earlier, the characteristics of the load impedance are uniquely manifested in the standing wave pattern on the line. By measuring the standing wave pattern, (which is only a measurement of the signal amplitude) one can indirectly obtain the phase of the complex load impedance.

For measuring the standing wave pattern, a special type of transmission line called the slotted transmission line is used. In this transmission line there is a voltage probe which moves along the length of the line and measures the magnitude of the voltage. A plot of probe output as a function of distance gives the standing wave pattern.

The unknown impedance which is to be measured is connected at the end of the transmission line as shown in Fig. 2.30. The transmission line is excited with a source of desired frequency ω. From the standing wave pattern three quantities, namely the maximum voltage $|V|_{max}$, minimum voltage $|V|_{min}$, and the distance of the voltage minimum from the load is measured. The ratio of $|V|_{max}$ and $|V|_{min}$ gives the VSWR on the line.

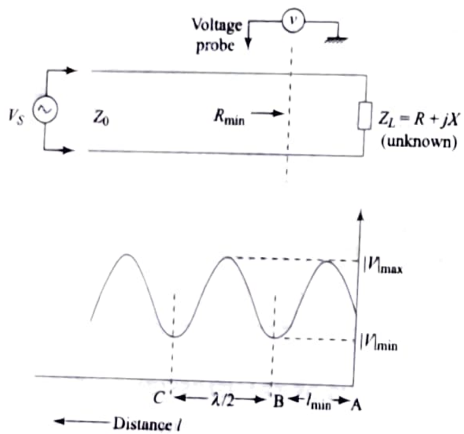


Fig. 2.30 *Transmission line with voltage measuring probe and the voltage standing wave pattern.*

$$\rho = \frac{|V|_{max}}{|V|_{min}} \tag{2.136}$$

We know that at point B on the transmission line where the voltage is minimum, the impedance is real and its value is $R_{min} = Z_0/\rho$. The impedance R_{min} is nothing but the transformed value of the load impedance Z_L. We can therefore obtain the unknown impedance Z_L by transforming back R_{min} from point B to point A. Let the distance of the voltage minimum from the load be l_{min}. Since the transformation from B to A is away from the generator, the distance BA is negative. The unknown impedance therefore is

$$Z_L \equiv R + jX = Z_0\left[\frac{R_{min}\cos\beta(-l_{min}) + jZ_0\sin\beta(-l_{min})}{Z_0\cos\beta(-l_{min}) + jR_{min}\sin\beta(-l_{min})}\right] \tag{2.137}$$

Substituting for $R_{min} = Z_0/\rho$, we get

$$Z_L \equiv R + jX = Z_0\left[\frac{\frac{Z_0}{\rho}\cos\beta l_{min} - jZ_0\sin\beta l_{min}}{Z_0\cos\beta l_{min} - j\frac{Z_0}{\rho}\sin\beta l_{min}}\right] \tag{2.138}$$

$$= Z_0\left[\frac{1 - j\rho\tan\beta l_{min}}{\rho - j\tan\beta l_{min}}\right] \tag{2.139}$$

Separating real and imaginary parts we get

$$R = \frac{\rho(1 + \tan^2\beta l_{min})}{\rho^2 + \tan^2\beta l_{min}} \tag{2.140}$$

$$X = \frac{(1-\rho^2)\tan\beta l_{min}}{\rho^2 + \tan^2\beta l_{min}} \tag{2.141}$$

So, we see that the complex impedance is measured without directly measuring phase of any quantity. Generally, the value of β which is required in Eqns (2.140) and (2.141) is not supplied in advance and one has to obtain it from the standing wave pattern itself. Since, the distance between any two consecutive voltage minima or maxima is equal to $\lambda/2$, and β is equal to $2\pi/\lambda$, measurement of β from the standing wave pattern is a trivial task. Generally, voltage minima are preferred because they can be located with higher precision compared to the voltage maxima. While practically implementing the above scheme one would also notice that invariably the location of unknown impedance Z_L is not precisely defined. As a result, the measurement of l_{min} may have some error which in turn will result into an error in the load impedance.

To overcome this problem the measurement is carried out in two steps. First, the standing wave pattern is obtained with the unknown load as explained earlier. Now replace the unknown impedance by an ideal short-circuit and obtain the standing wave pattern again. The two standing wave patterns are shown in Fig. 2.31.

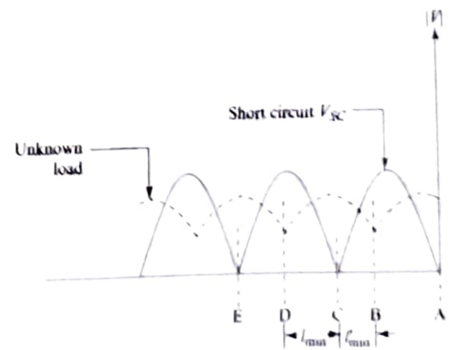


Fig. 2.31 *Voltage standing wave patterns for a short circuit and an unknown load.*

At the short circuit point (which is also the location of the unknown impedance) the voltage V_{sc} is zero. The voltage is also zero at points which are multiple of $\lambda/2$ away from it, i.e. at points C and E, etc. The points C and E, etc. represent impedance conditions identical to that at A, that is, the impedance at C or E is equal to the unknown impedance. The unknown impedance, therefore, can be obtained by transforming impedance Z_0/ρ at B or D to point C. If impedance is transformed from D to C the distance $l_{\min}$ is negative, whereas, if the transformation is made from B to C the distance $l'_{\min}$ is positive. The unknown impedance, therefore, can be evaluated as

$$Z_L = R + jX = Z_0\left[\frac{1 - j\rho\tan\beta l'_{\min}}{\rho - j\tan\beta l'_{\min}}\right] = Z_0\left[\frac{1 + j\rho\tan\beta l_{\min}}{\rho + j\tan\beta l_{\min}}\right] \quad (2.142)$$

One can note here that, $l_{\min} + l'_{\min} = \lambda/2$. In the impedance calculation either of l_{min} or l'_{min} can be used. As long as the sign of the distance is taken correctly it does not matter which of the minima is taken for impedance transformation.

EXAMPLE 2.21 A 50 Ω line is terminated in an unknown impedance Z. The distance of the voltage maximm from the load is 6 cm, and the distance between two consecutive maxima is 30 cm. The VSWR on the line is measured to be 3. Find the unknown impedance Z.

Solution:

The distance between two consecutive maxima $= \lambda/2 = 30$ cm i.e., $\lambda = 60$ cm. The phase constant on the line is

$$\beta = \frac{2\pi}{\lambda} = \frac{\pi}{30} \text{ rad/cm}$$

Refer Fig. 2.32. Draw VSWR = 3 circle and mark $V_{\min}$ point.
Since the distance between adjacent voltage maximum and minimum is $\lambda/4 =$ 15 cm, the distance of the voltage minimum from the load is

$$l_{\min} = 15 + 6 = 21 \text{ cm}$$

$$\beta l_{\min} = \frac{\pi}{30} \times 21 \text{ rad} = 126^\circ$$

Move on constant VSWR circle in anticlockwise direction to reach to point Z. Read the impedance value at Z.

$$Z = 50 \times (1.15 - j1.23) = 57.5 - j61.5\ \Omega$$

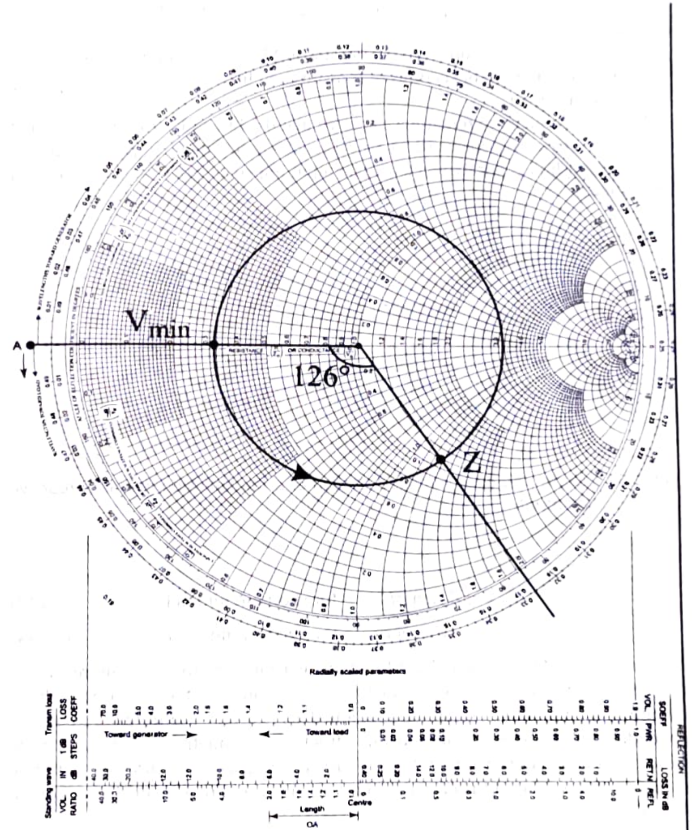


Fig. 2.32

2.12.2 Transmission Line as a Circuit Element

As the frequency increases, the realization of lumped reactive elements becomes more and more difficult. For example, if we wind a coil to get a certain inductance, the distributed capacitance of the coil may be so large that the coil, instead of showing inductive reactance may show capacitive reactance. Similarly, at times the lead inductance of a capacitor may be large enough to nullify the capacitive reactance.

At frequencies of hundreds and thousands of MHz where lumped elements are hard to realize, the use of sections of transmission line as reactive elements may be more convenient.

From the impedance relation we can see that if a line of length l is terminated in a short circuit or open circuit (see Fig. 2.33) the input impedance of the transmission line is purely reactive. The input impedance of a loss-less line can be written as

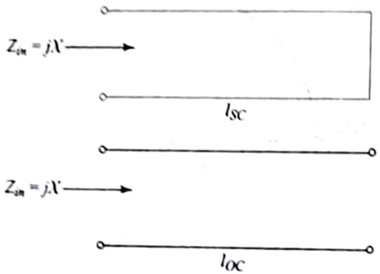


Fig. 2.33 *Short and open circuited section of a transmission line as a reactive element.*

$$Z_{in} = jZ_0 \tan \beta l \qquad \text{for short circuit load} \qquad (2.143)$$

$$= -jZ_0 \cot \beta l \qquad \text{for open circuit load} \qquad (2.144)$$

Since, the range of tan and cot functions is from $-\infty$ to $+\infty$, any reactance can be realized by proper choice of l. Moreover, any reactance can be realized by either open or short circuit termination. This is a very useful feature because depending upon the transmission line structure, terminating one way may be easier than other. For example, for a microstrip type line (see in later sections), realizing an open circuit is easier as short circuit would require drilling a hole in the substrate.

Now, if a reactance X is to be realized in a high frequency circuit one can use a short circuited line of length l_{sc} or an open circuited line of length l_{oc} where l_{sc} and l_{oc} can be obtained by inverting Eqns (2.143) and (2.144) to give

$$l_{sc} = \frac{1}{\beta} \tan^{-1}\left(\frac{X}{Z_0}\right) \qquad (2.145)$$

$$l_{oc} = \frac{1}{\beta} \cot^{-1}\left(\frac{-X}{Z_0}\right) \qquad (2.146)$$

Alternatively, one can use the Smith chart to find l_{sc} or l_{oc} as follows:

We know that the pure reactive impedances lie on the outermost circle of the Smith chart. Mark the reactance X to be realized on the Smith chart to get point 'X' in Fig. 2.34. Now, the length l_{sc} is the distance of the short circuit point from point X, away from the generator. If we, therefore, move in anticlockwise direction from point X to the short circuit (SC) point on the Smith chart we get l_{sc} (see Fig. 2.34). Similarly, l_{oc} is the distance of open circuit (OC) from X in the anticlockwise direction as indicated in Fig. 2.34. Note here that, instead of reactance if we had to realize a susceptance B, the procedure is identical except that SC and OC points are interchanged.

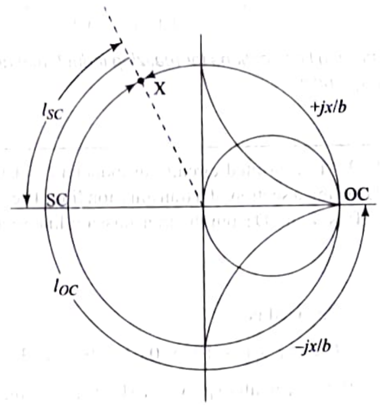
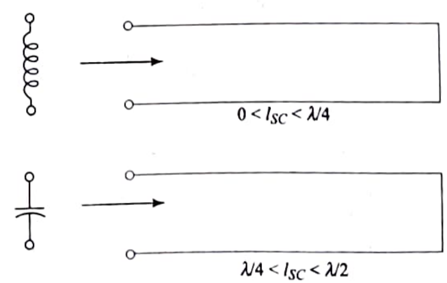


Fig. 2.34 *Calculation of length of a transmission line section for realizing a reactance using the Smith chart.*

Figure 2.35 shows the range of tranmission line lengths and the corresponding reactances which can be realized at the input terminals of the line.



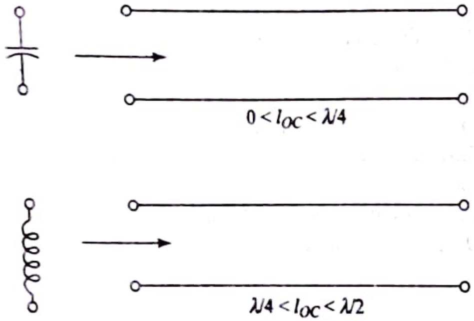


Fig. 2.35 *Transmission line sections for realizing inductance and capacitance at high frequencies.*

EXAMPLE 2.22 In a printed circuit, an inductance of 0.01 μH is to be realized at 6 GHz using a section of a transmission line. The wavelength of the signal on the PCB is 4 cm. Design the transmission line section as a reactive element.

Solution:

The reactance to be realized is

$$X = \omega L = 2\pi \times 6 \times 10^9 \times 0.01 \times 10^{-6} = 377 \ \Omega$$

Generally, in a PCB it is difficult to provide a short circuit since a hole has to be drilled through the substrate. So, it is convinient to use open circuited section of a transmission line. The length of line, therefore, is (see Eqn (2.146))

$$l_{oc} = \frac{1}{\beta}\cot^{-1}\left(\frac{-X}{Z_0}\right) = \frac{\lambda}{2\pi}\cot^{-1}\left(\frac{-X}{Z_0}\right)$$

Now, first we have to choose suitable Z_0 for the line. Let us take $Z_0 = 150\ \Omega$. (Choose Z_0 such that the ratio X/Z_0 is not too large. The $\cot x$-function is very steep around $x = 0, \pi/2$). We then get

$$l_{oc} = \frac{4.0}{2\pi}\cot^{-1}\left(\frac{-377}{150}\right) = 1.76 \text{ cm}$$

2.12.3 Transmission Lines as Resonant Circuits

We have seen in the previous section that a short circuited line behaves like an inductor if $0 < l_{sc} < \lambda/4$, and it behaves like a capacitor if $\lambda/4 < l_{sc} < \lambda/2$. If the length is exact multiple of $\lambda/4$ the imput impedance of the line is zero or ∞. Let us plot the input impedance as a function of frequency 'f', for a given length of transmission l and a given termination (short circuit or open circuit).

Figure 2.36 shows the variation of reactance as a function of frequency for open and short circuited sections of a transmission line. It is clear that around frequencies f_1, f_2, f_3, f_4.., for which the length l is an integer multilple of $\lambda/4$, the impedance variation is identical to an L-C resonant circuit. The impedance characteristics of a series and a parallel resonant circuit are shown in Fig. 2.37.

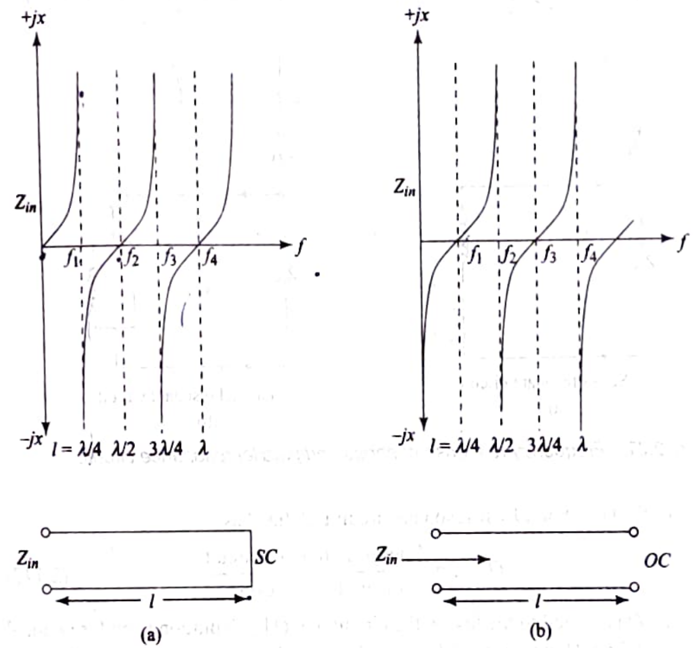


Fig. 2.36 *Variation of input impedance of a transmission line section as function of frequency.*

Comparing Fig. 2.36 with Fig. 2.37, one can observe that a short circuited line behaves like a parallel resonant circuit around frequencies f_1 and f_3, whereas around f_2 and f_4 its behavior is like a series resonant circuit. In general then we can say that a short circuited section of a line having length equal to odd multiples of $\lambda/4$ (i.e. $\lambda/4$, $3\lambda/4$, $5\lambda/4$, etc) is equivalent to a parallel resonant circuit. Similarly, if the length is equal to even multiples of $\lambda/4$ (i.e. $\lambda/2$, λ, $3\lambda/2$, etc.) the line is equivalent to a series resonant circuit. A converse is true for an open circuited section of a line i.e. if the length of the line is equal to odd multiples of $\lambda/4$, the line behaves like a series resonant circuit, and if the length of the line is equal to even multiple of $\lambda/4$, the line behaves like a parallel resonant circuit.

Since the section of transmission line having length equal to integer multiples of $\lambda/4$ is equivalent to a resonant circuit, it is worthwhile to ask what is its quality

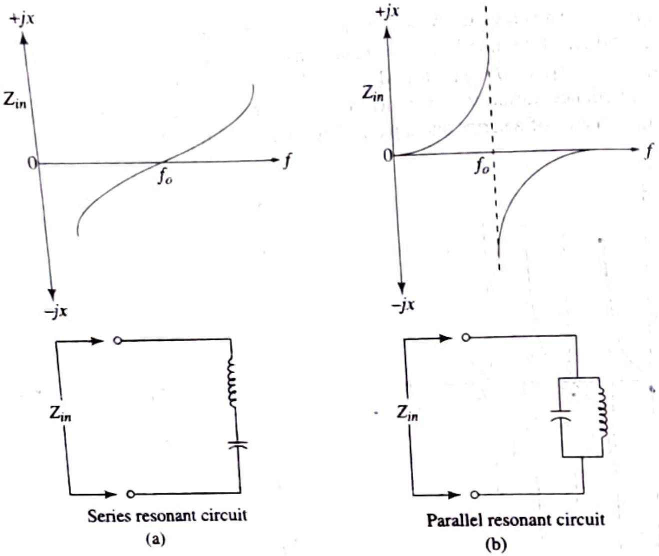


Fig. 2.37 *Frequency response of series and parallel resonance circuit.*

factor Q. The Q of a LCR resonant circuit is defined as

$$Q = 2\pi \frac{\text{Energy stored in a circuit}}{\text{Enegy lost per cycle}} \tag{2.147}$$

As the Q is related to the loss in the circuit, the Q by definition is ∞ for an ideal loss-less line. However, for a low-loss line the Q is finite. In this particular case where one is explicitly investigating the loss, the assumption which we have been making so far that is, a low-loss line can be treated as loss-less line, is not justified. No matter how small the loss is, the propagation constant γ can not be taken to be purely imaginary. Taking γ to be complex, i.e. $\gamma = \alpha + j\beta$ ($\alpha << \beta$ for low loss lines) one has to re-investigate the input impedance of a resonant line. The input impedance of a short or open circuited line having propagation constant γ can be obtained from Eqn (2.45) as

$$Z_{sc} = Z_0 \tanh \gamma l \quad \text{for short circuit} \tag{2.148}$$

$$Z_{oc} = Z_0 \coth \gamma l \quad \text{for open circuit} \tag{2.149}$$

Note that although γ has been taken complex for a low-loss transmission line, Z_0 is almost real. Substituting for $\gamma = \alpha + j\beta$, we get

$$Z_{sc} = Z_0 \tanh(\alpha + j\beta)l = Z_0 \left[\frac{\tanh \alpha l + \tanh (j\beta)l}{1 + \tanh \alpha l \tanh (j\beta)l}\right] \tag{2.150}$$

For a low-loss line, taking $\alpha l << 1$, we have $\tanh \alpha l \approx \alpha l$. Also, $\tanh(j\beta)l = j\tan\beta l$. Hence, we get

$$Z_{sc} \approx Z_0 \left[\frac{\alpha l + j\tan\beta l}{1 + j\alpha l \tan\beta l}\right] \tag{2.151}$$

Similarly, for an open circuited line we get

$$Z_{oc} \approx Z_0 \left[\frac{1 + j\alpha l \tan\beta l}{\alpha l + j\tan\beta l}\right] \tag{2.152}$$

For resonant lines, l is integer multiples of $\lambda/4$ i.e. $\beta l (= 2\pi l/\lambda)$ is integer multiples of $\pi/2$. If, we take odd multiples of $\lambda/4$, $\tan\beta l = \infty$, and we get

$$Z_{sc} \approx \frac{Z_0}{\alpha l} \quad \text{Parallel resonance} \tag{2.153}$$

$$Z_{oc} \approx Z_0 \alpha l \quad \text{Series resonance} \tag{2.154}$$

On the other hand, if we take even multiples of $\lambda/4$, $\tan\beta l = 0$, giving

$$Z_{sc} \approx Z_0 \alpha l \quad \text{Series resonance} \tag{2.155}$$

$$Z_{oc} \approx \frac{Z_0}{\alpha l} \quad \text{Parallel resonance} \tag{2.156}$$

It is now clear that a short-circuited line having length equal to odd multiples of $\lambda/4$, and an open circuited line having length equal to even multiples of $\lambda/4$ is equivalent to a parallel resonant circuit. Similarly, a short-circuited line having length equal to even multiples of $\lambda/4$, and an open-circuited line of length equal to odd multiples of $\lambda/4$ is equivalent to a series resonant circuit. From Eqns (2.153) to (2.156) we conclude that a parallel resonant section of a line has an impedance $Z_0/\alpha l$ and a series resonant section has an impedance $Z_0 \alpha l$. One can cross-check the result with that of an ideal loss-less line. In the absence of any loss, the parallel resonant circuit shows infinite impedance and a series resonant circuit shows zero impedance at the resonance. The above result is indeed consistent with this.

For calculation of Q let us re-write Eqn (2.147) as

$$Q = 2\pi f_0 \frac{\text{Energy stored in the line}}{\text{Energy lost per second}} \tag{2.157}$$

Here, 'Energy lost per cycle' is written as 'Energy lost per second divided by the resonant frequency f_0'.

Let us now consider a short circuited section of a line having length equal to odd multiples of $\lambda/4$. This line is equivalent to a parallel resonant circuit. Let the line be applied with a voltage V_0 between its input terminals as shown in Fig. 2.38.

The voltage and current standing wave patterns on the line are shown in Fig. 2.38(b,c). The voltage is zero at the short-circuit-end of the line and is maximum at the input end of the line. Similarly, current is maximum at the short-circuit end and minimum at the input end of the line.

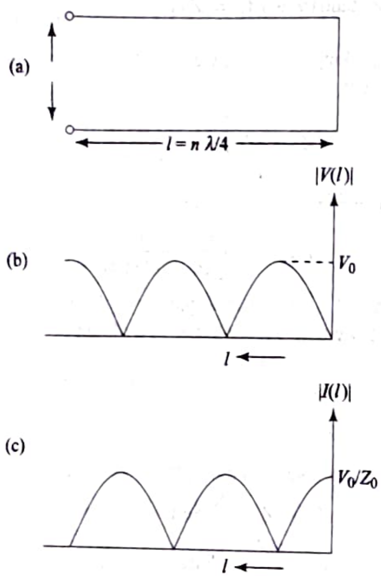


Fig. 2.38 *Voltage and current variation on a resonant section of transmission line.*

The maximum value of the voltage on the line is V_0 and maximum value of current is V_0/Z_0 (see Eqn (2.71)). For a short-circuited line the voltage and current on the line are given as

$$V(l) = |V_0 \sin \beta l| \tag{2.158}$$

$$I(l) = |\frac{V_0}{Z_0} \cos \beta l| \tag{2.159}$$

The energy stored in a $n\lambda/4$ long section of the line is

$$U = \frac{1}{2}C\int_0^{n\lambda/4} [V(l)]^2 dl + \frac{1}{2}L\int_0^{n\lambda/4} [I(l)]^2 dl \tag{2.160}$$

$$= \frac{1}{2}C\int_0^{n\lambda/4} [V_0 \sin \beta l]^2 dl + \frac{1}{2}L\int_0^{n\lambda/4} \left[\frac{V_0}{Z_0}\cos \beta l\right]^2 dl \tag{2.161}$$

$$= \frac{1}{4}CV_0^2\left(\frac{n\lambda}{4}\right) + \frac{1}{4}L\frac{V_0^2}{Z_0^2}\left(\frac{n\lambda}{4}\right) \tag{2.162}$$

Since $Z_0 = \sqrt{L/C}$, we have $L/Z_0^2 = C$. Substituting in Eqn (2.162) we note that the two terms on RHS of Eqn (2.162) are equal, i.e. the inductive and capacitive

energies are equal, and the total energy is

$$U = \frac{1}{2}CV_0^2\left(\frac{n\lambda}{4}\right) \tag{2.163}$$

The energy lost per second is nothing but the power loss in the line. At resonance the line effectively appears like a resistance of value $Z_0/\alpha l$ (see Eqn (2.153)). The power loss in the line therefore is ($l = n\lambda/4$)

$$P_{loss} = \frac{V_0^2}{(Z_0/\alpha l)} = \frac{V_0^2}{Z_0}.\alpha.\frac{n\lambda}{4} \tag{2.164}$$

Substituting from Eqns (2.163) and (2.164) in Eqn (2.157) we get the quality factor of the line as

$$Q = 2\pi f_0 \frac{Z_0 C}{2\alpha} \tag{2.165}$$

Again noting that $Z_0 C = (\sqrt{L/C})C = \sqrt{LC}$, and $2\pi f_0 = \omega$, the numerator in Eqn (2.165) is equal to $\omega\sqrt{LC}$, which is nothing but β (refer Eqn (2.59)). We therefore get

$$Q = \frac{\beta}{2\alpha} \tag{2.166}$$

One can note here that Q is independant of the length of the line as long as the loss is small. Equation (2.166) also suggests that to achieve high Q, α should be $<< \beta$. In practice generally the lines have loss low enough to give a Q of few hundred very easily. Since, the 3 dB-bandwidth of a resonant circuit is f_0/Q, higher value of Q implies highly tuned circuits. The transmission line sections therefore act as excellent frequency selective circuits at high frequencies.

EXAMPLE 2.23 A 75 Ω low-loss transmission line has a loss of 1.5 dB/m. The velocity of the voltage wave on the line is 2×10^8 m/sec. A section of the line is used to make a series resonant circuit at 1 GHz. Find the input impedance of the line, its quality factor and the 3 dB bandwidth of the resonant circuit.

Solution:

The wavelength on the line is

$$\lambda = \frac{v}{f} = \frac{2 \times 10^8}{10^9} = 0.2 \text{ m}$$

The phase constant is

$$\beta = \frac{2\pi}{\lambda} = 10\pi \text{ rad/m}$$

If we take open circuited section of the line, for series resonance, the length of the section would be

$$l_{oc} = \frac{\lambda}{4}, \frac{3\lambda}{4}, \ldots$$
$$= 0.05, 0.15, \ldots\text{m}$$

If we take short circuited section of the line, for series resonance, the length of the section would be

$$l_{sc} = \frac{\lambda}{2}, \frac{2\lambda}{2}, \ldots$$
$$= 0.1, 0.2, \ldots\text{m}$$

The loss of the line is $\alpha = 1.5$ dB/m $= 1.5/8.68 = 0.173$ nepers/m
For series resonance, the input impedance is

$$Z_{in} = Z_0 \alpha l = 12.95\, l$$

where $l = l_{oc}$ or l_{sc}. The quality factor is

$$Q = \frac{\beta}{2\alpha} = \frac{10\,\pi}{2 \times 0.173} = 90.95$$

The 3-dB bandwidth is

$$BW = \frac{f_0}{Q} = \frac{1\text{ GHz}}{90.95} = 10.99 \text{ MHz}$$

2.12.4 Voltage or Current Step-up Transformer

In the previous application, we studied the impedance behavior of a resonant transmission line. Here, let us investigate the voltage and current that exist on a resonant section of a transmission line. Let us take a resonant transmission line of length $\lambda/4$. The line is open circuited at one end and short circuited at the other as shown in Fig. 2.39.

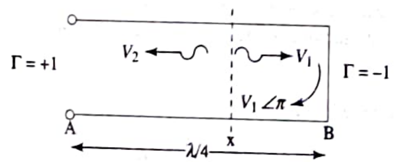


Fig. 2.39 *Growth of voltage waves on a resonant section of a transmission line. Transmission line as a step-up transformer.*

Let us say there is a voltage source which induces a voltage in the line at some point X. This induced voltage will send two voltage waves V_1 and V_2 with equal amplitudes. Consider now one of the waves, say V_1. This wave travels upto point B to encounter a short circuit. Since the reflection coefficient for a short circuit is -1, the wave gets fully reflected with a phase reversal. The wave after travelling a distance BA reachs to the open circuited end of the line and again gets fully reflected but with no phase reversal as $\Gamma = +1$ for the open circuit. After one round trip, therefore, when the wave V_1 reaches point X, its amplitude is same as its original value but its phase is changed by 2π, π due to reflection at point B and π due to propagation of a round trip distance of $\lambda/2$. This wave, therefore, adds up with the induced voltage in phase and the added up wave travels on the transmission line. The process is regenerative and the amplitude of the voltage wave V_1 goes on increasing. Exactly identical thing happens with the other wave V_2. Since, the two waves travelling in the opposite directions identically grow in amplitude, the result is a continiously growing standing wave on the line with appropriate voltage maximum at A and voltage minimum at B. If the coupling of voltage is sustained, and the line is loss-less, there is no limit on the voltage and current and the voltage and current eventually would grow to ∞. However, if the line has a loss (no matter how small), then of course the voltage and current stabilize at some finite values. As the voltage/current increases, the ohmic loss also increases and when the power lost in the line just equals the power supplied by the coupling source, the voltage/current stabilizes. It should be noted, however, that the maximum stabilized voltage or current on the line could be much higher than the coupling voltage or current. This suggests a possibility of using a resonant transmission line as a step-up voltage or current transformer.

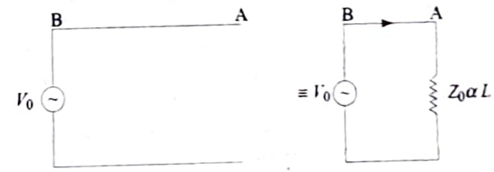


Fig. 2.40 *Resonant section of a low-loss transmission line.*

As an illustrative example let us take a resonant section of a line same as that in Fig. 2.39, and instead of putting a short circuit put an ideal voltage source at point B as shown in Fig. 2.40. The open circuit at point A appears as almost short (for a low loss line) at point B. The impedance seen by the voltage source is $Z_0\alpha l$ and a current $V_0/Z_0\alpha l$ flows in terminal B. Since point B is a voltage minimum and a current $V_0/Z_0\alpha l$ is equal to the maximum current current maximum, the source current $V_0/Z_0\alpha l$ is equal to the maximum current on the line I_{max}. The maximum voltage on the line then is $Z_0 I_{max}$ and it appears at point A. We therefore have

$$V_A = Z_0 \frac{V_0}{Z_0 \alpha l} = \frac{V_0}{\alpha l} \tag{2.167}$$

Since, $\alpha l << 1$ for a low-loss line, we get $V_A >> V_0$. That is, the voltage at the open-circuited end of the resonant line is much higher compared to the excitation voltage V_0. A resonant section of a transmission line therefore can be used as

ansformer. It should be kept in mind, however, that a large step-up
sible only in the no-load condition. Any loading at point A reduces the
$_{ax}$, i.e. the voltage amplification ratio.
ltage step-up ratio is

$$\frac{V_A}{V_0} = \frac{1}{\alpha l} \tag{2.168}$$

ng $l = n\lambda/4$ (where n is an odd integer), and substituting $\lambda = 2\pi/\beta$, the
step-up ratio can be written as

$$\frac{V_A}{V_0} = \frac{2\beta}{n\pi\alpha} \tag{2.169}$$

$$\Rightarrow \quad \text{Volatge Amplification Ratio} = \frac{4Q}{n\pi} \tag{2.170}$$

Since, Q is typically few hundreds for a low-loss transmission line, a voltage
nplification of few hundreds may be expected in a resonant line.

.13 IMPEDANCE MATCHING USING TRANSMISSION LINES

While discussing characteristics of a transmission line, we have seen that if a line is terminated in the characteristic impedance Z_0, the impedance at every point on the line becomes Z_0. Also, in this case there is no reflection on the line and the power is maximally transferred to the load. In practice, however, it is not always possible to design a circuit whose input or output impedance is matched to the adjacent circuits. For example, there may be a circuit which is to be connected to a signal generator of 50 Ω output impedance but the input impedance of the circuit is not 50 Ω. When this connection is made, not only the maximum power is not being transferred by the generator to the load, but the reflected wave might enter the generator and may alter its characteristics like frequency, etc. It is, therefore, essential to devise a technique which can avoid reflections from the circuits.

Transmission lines can be used for matching two impedances. Due to low-loss the transmission line provides impedance matching with negligible loss of power. In the following sections we discuss various impedance matching techniques using transmission line sections.

2.13.1 Quarter-Wavelength Transformer

This technique is generally used for matching two resistive loads, or for matching a resistive load to a transmission line, or for matching two transmission lines with unequal characteristic impedances (see Fig. 2.41). All cases are identical in principle as all require matching between two purely resistive impedances.

The principle here is very simple. We introduce a section of a transmission line (transformer) between two resistances to be matched, such that the transformed impedances perfectly match at either end of the transformer section. That is,

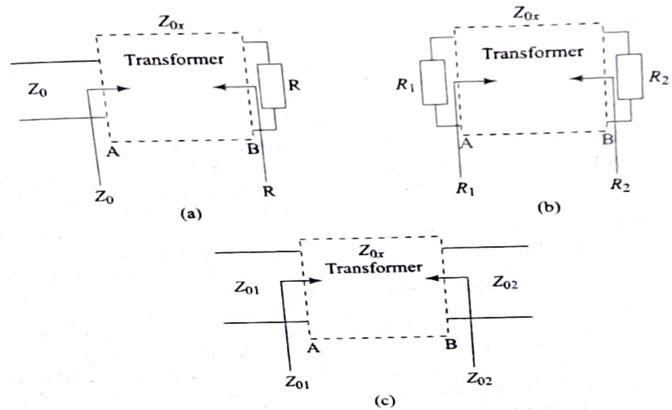


Fig. 2.41 *Impedance matching using quarter wavelength transformer. (a) Matching of a resistance to a line (b) Matching of two resistances (c) Matching between two long lines.*

in Fig. 2.41(a) say, the impedance seen towards right at A appears to be Z_0, and impedance seen towards left at B appears to be R. So, when seen from transmission line side it appears to be terminated in Z_0, and when seen from load resistance side it appears to be connected to a conjugately matched load R. Similar is true for Figs 2.41(b,c).

Since, pure resistances are to be matched here, the impedance transformation in the transformer has to be from a resistive impedance to a resistive impedance. This is possible only in the following two cases.

(a) when length of the transforming line is $\lambda/2$
(b) when the length of the transforming line is $\lambda/4$

A $\lambda/2$ section of a transmission line transforms an impedance into itself and hence does not serve any purpose for matching. So, the only possibility is that the transformer must be $\lambda/4$ long.

Let us assume that the characteristic impedance of the transformer section is Z_{0x}. For $\lambda/4$ length, the transformer inverts the normalized impedance (see Eqn (2.89)). Therefore, the impedance seen at A towards right in Fig. 2.41(a) would be

$$Z_A = \frac{1}{(R/Z_{0x})}.Z_{0x} = \frac{Z_{0x}^2}{R} \tag{2.171}$$

For matching at A, Z_A should be equal to Z_0, i.e.

$$\frac{Z_{0x}^2}{R} = Z_0 \tag{2.172}$$

$$\Rightarrow Z_{0x} = \sqrt{RZ_0} \tag{2.173}$$

For Fig. 2.41(b) $Z_{0x} = \sqrt{R_1 R_2}$ and for Fig. 2.41(c), $Z_{0x} = \sqrt{Z_{01} Z_{02}}$.

So, in general we can say that two resistive impedances can be matched by a section of a transmission line which is quarter-wavelength long and has characteristic impedance equal to the geometric mean of the two resistances.

In first look, it appears from the above discussion that a quarter wavelength transformer can be used to match only purely resistive impedances. However, if we see carefully, we find that it is not true. This is due to the fact that we can always transform a complex impedance into a purely real one by adding an appropriate section of a transmission line. Let us consider matching of a complex impedance $R + jX$ to the characteristic impedance of a transmission line Z_0. To transform impedance $R + jX$ to some real value let an extra length 'L' of a transmission line be added between the quarter wavelength transformer and the impedance as shown in Fig. 2.42. The characteristic impedance of the extra length of the line be say Z_{01} (one can use a line of characteristic impedance Z_0 as well). The length L should be chosen such that the transformed impedance Z' seen towards right at B is purely real. This can be done easily by using the Smith chart.

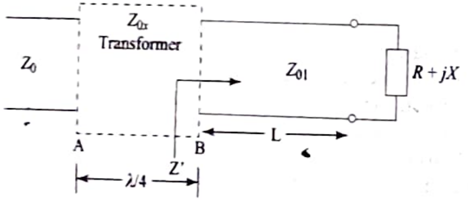


Fig. 2.42 *Matching of a complex impedance to a line using quarter wavelength transformer.*

First the impedance $Z = R + jX$ is normalized with respect to Z_{01} to give $\bar{Z} = (R + jX)/Z_{01} \equiv \bar{r} + j\bar{x}$. Let this point be denoted by P (see Fig. 2.43). Draw the constant VSWR circle through P. The circle will intersect the real axis at points S and T. These points represent location on the transmission line where the impedance is purely resistive. At point T the normalized impedance is r_{max} and at point S the normalized impedance is r_{min}.

If we take $L = L_{max}$ then impedance $Z' = Z_{01} r_{max}$ and then for matching we get

$$Z_{0x} = \sqrt{Z_0 Z_{01} r_{max}} \tag{2.174}$$

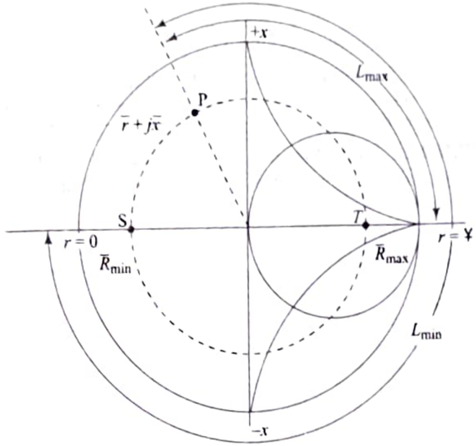


Fig. 2.43 *Matching of complex impedance using quarter wavelength transformer using the Smith chart.*

Similarly, if we take $L = L_{min}$ then $Z' = Z_{01} r_{min}$ and we get

$$Z_{0x} = \sqrt{Z_0 Z_{01} r_{min}} \tag{2.175}$$

Theoretically, both solutions are equally acceptable and only their numerical values will decide which is practically more realizable.

The quarter-wavelength transformer although can match any impedance, has a very serious draw-back. For every impedance to be matched, one needs a line with different Z_{0x}. Since the characteristic impedance of a line is decided by the physical structure of the line like conductor size, dielectric constant, separation between conductors etc. for every impedance Z_{0x}, one would need a special transmission line. Realizing a line with a particular Z_{0x} may not be always possible in practice.

To overcome this drawback, the stub-matching techniques have been proposed. These techniques make use of the standard tranmission line sections for matching arbitrary impedances.

EXAMPLE 2.24 Two very long loss-less cables of characteristic impedances 50Ω and 100Ω respectively are to be joined for reflection-less transmission. Find the suitable matching transformer.

Solution:

The quarter wavelength transformer is appropriate for this case as we have to match two real impedances. The characteristic impedance of the tranformer section is

$$Z_{0x} = \sqrt{50 \times 100} = 70.7\ \Omega$$

The length of the transformer should be odd multiples of $\lambda/4$.

EXAMPLE 2.25 A load impedance $75 - j35\Omega$ is to be matched to 50Ω using quarter wave transformer. Design the matching set-up.

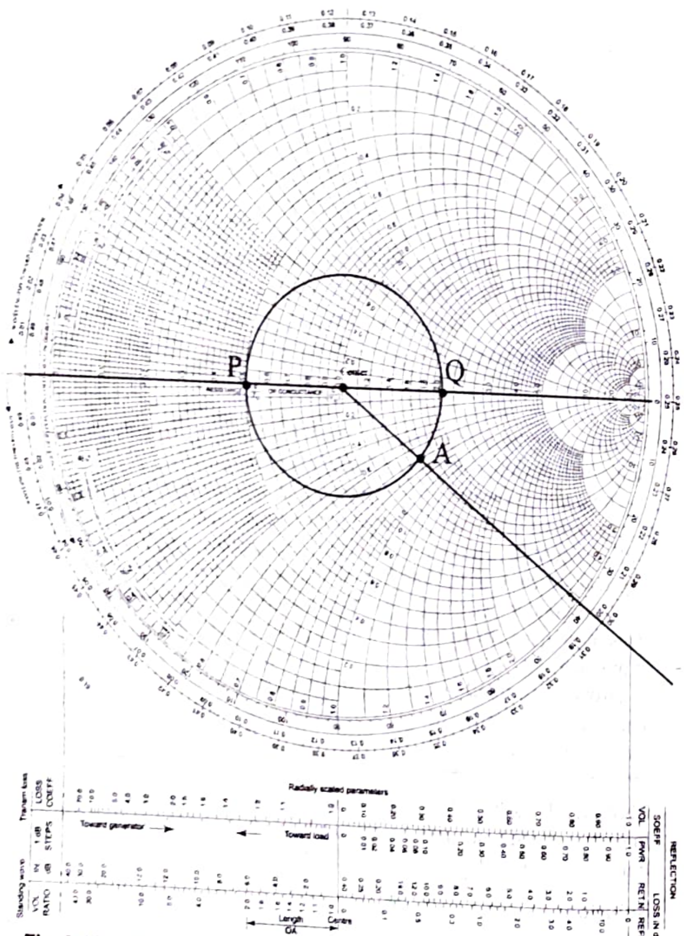


Fig. 2.44 *Impedance matching using quarter wavelength transformer*

Solution:

The normalized load impedance is

$$\overline{Z} = \frac{75 - j35}{50} = 1.5 - j0.7$$

Since, the quarter wave transformer matches real impedances, first connect a cable between the transformer and the load such that the transformer sees a real impedance. Let the cable used be having 50 Ω characteristic impedance. We can find the length of the cable, l, with the help of the Smith chart. The length l would correspond to arc AP or AQ measured clockwise (see Fig. 2.44).

If we take l corresponding to arc AP, we get $Z' = R_P = 25\ \Omega$ and the transformer characteristic impedance will be $Z_{0x} = \sqrt{50R_P} = 35.34\ \Omega$.

If, we, however, take l corresponding to arc AQ, we get $Z' = R_Q = 100\ \Omega$ and the transformer characteristic impedance will be $Z_{0x} = \sqrt{50R_Q} = 70.7\ \Omega$.

Both solutions are valid solutions and only other practical constraints may favour one over the other.

2.13.2 Single-Stub Matching Technique

A stub is a short-circuited section of a transmission line connected in parallel to the main transmission line. A stub of appropriate length is placed at some distance from the load such that the impedance seen beyond the stub is equal to the characteristic impedance.

Suppose, we have a load impedance Z_L connected to a transmission line with characteristic impedance Z_0 (Fig. 2.45). The objective here is that no reflection should be seen by the generator. In other words, even if there are standing waves in the vicinity of the load Z_L, the standing waves must vanish beyond certain distance from the load. Conceptually, this can be achieved by adding a stub to the main line such that the reflected wave from the short-circuit end of the stub and the reflected wave from the load on the main line completely cancel each other at point B to give no net reflected wave beyond point B towards the generator. In terms of impedances, the philosophy of the single stub matching is as follows:

We note that, any movement along the transmission line changes both the real and imaginary parts of the impedance seen on the line. Therefore, first make the real part of the impedance equal to the characteristic impedance by moving along the line. At that location add a reactance to nullify the reactance of the transformed impedance.

In the first look, manipulation of reflected waves might appear a tedious task and infact if one tries to solve the problem by analytical means it is a rather tedious task. However, with the help of the Smith chart the problem can be solved very easily.

As the Smith chart always uses normalized impedances, let us normalize the impedance Z with Z_0 to give $\overline{Z} = r + jx$. In terms of normalized impedances Fig. 2.45 can be re-drawn as Fig. 2.46.

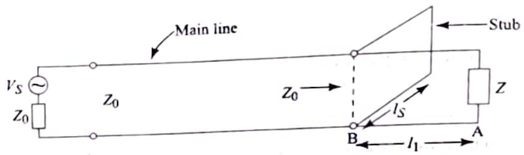


Fig. 2.45 *A stub connected to the main transmission line for matching the load.*

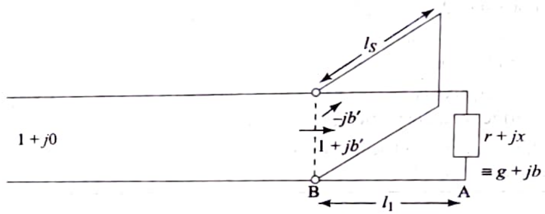


Fig. 2.46 *Main transmission line and the stub with normalized admittances.*

Since, we are having a parallel connection of transmission lines, it is more convenient to solve the problem using admittances rather than impedances. To convert the impedance into admittance also we make use of the Smith chart and avoid any analytical calculation.

First, taking the Smith chart as the impedance chart mark the normalized load impedance $r + jx$ as point P (Fig. 2.47) and draw the constant VSWR circle through P. A diagonally opposite point A then represents the normalized load admittance $g + jb$ corresponding to the load impedance $r + jx$. The admittance seen at point A (Fig. 2.47) is therefore $Y_A = g + jb$. Now onwards treat the Smith chart as the admittance chart.

We can note in Fig. 2.47 that the constant VSWR circle intersects the $g = 1$ circle at two points B and D. This means, if we move along the transmission line by a distance corresponding to arc AB or AD in the clockwise direction, the real part of the transformed admittance would become unity. Let us choose point B. The arc AB corresponds to the distance of the stub from the load, l_1. Let the admittance at B be $Y_B = 1 + jb'$ (say). Now if the stub susceptance at B is $-jb'$, the total admittance at B would be $1 + j0$.

Finding the length of a short circuited line which would give a desired susceptance (reactance) at its input terminals, has already been discussed in the Section 2.12.2. Mark the $-jb'$ point on the Smith chart say Q. Then on the stub if we move away from the generator we should reach the short circuit point. The arc QS in the anticlockwise direction therefore provides the length of the stub, l_s.

One may note that, instead of point B if we had chosen point D, the transformed admittance would have been $Y_D = 1 - jb'$. Then the susceptance of the stub would have been $+jb'$ and the stub length would have been $(\lambda/2 - l_s)$.

The admittance flow diagram is shown by arrows in Fig. 2.47. The admittance moves from A to B and then to M (center of the chart).

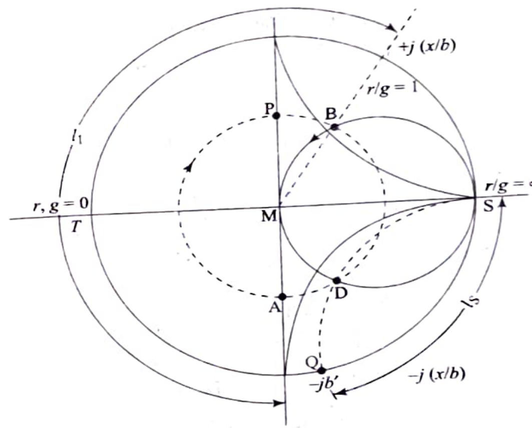


Fig. 2.47 *Single-stub matching using the Smith chart.*

The single-stub matching technique is superior to the quarter wavelength transformer as it makes use of only one type of transmission line for the main line as well as the stub. This technique also in principle is capable of matching any complex load to the characteristic impedance/admittance. The single stub matching technique is quite popular in matching fixed impedances at microwave frequencies.

The single stub matching technique although has overcome the drawback of the earlier technique, it still is not suitable for matching variable impedances. A change in load impedance results in a change in the length as well as the location of the stub. Even if changing length of a stub is a simpler task, changing the location of a stub may not be easy in certain transmission line configurations. For example, if the transmission line is a co-axial cable, the connection of a stub would need drilling of a hole in the outer conductor.

For a variable impedance, the matching is achieved using the double-stub matching technique.

EXAMPLE 2.26 A load impedance $90 - j25$ is to be matched to 50Ω using single stub matching. Find the length and location of stub.

Solution:

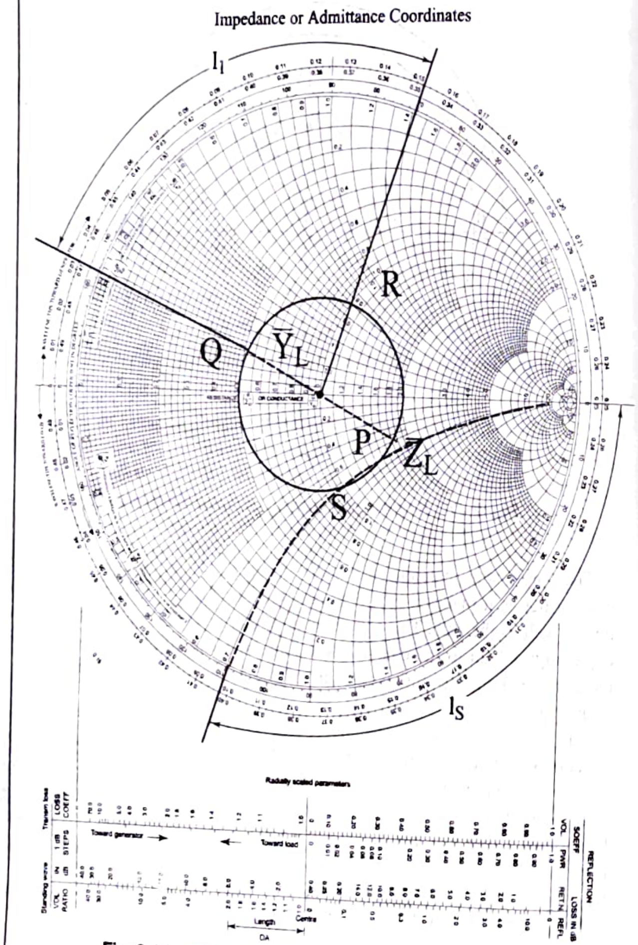


Fig. 2.48 *Single stub matching using the Smith chart.*

$$\bar{Z}_L = \frac{90 - j25}{50} = 1.8 - j0.5$$

Refer to Fig. 2.48 for length and location of the stub.

2.13.3 Double-Stub Matching Technique

To overcome the drawbacks of the single-stub matching technique, the double-stub matching technique is employed. The technique uses two stubs with fixed locations. As the load changes only the lengths of the stubs are adjusted to achieve matching.

Let us assume that a normalized admittance $g + jb$ is to be matched using the double stub matching technique. The first stub is located at a convenient distance from the load say l_1 (Fig. 2.49). The second stub is located at a distance of $3\lambda/8$ from the first stub. Although there is nothing unique about this distance of $3\lambda/8$, we will see later that it has certain advantages. The impedance matching philosophy can be understood as follows:

We have already seen in the single-stub matching technique that at the location where a stub is connected if the admittance seen towards the load is of the form $1 + jb$ then the stub can cancel the reactive part of the admittance and we can get matching. In the present case, if the admittance at point C is of the form $1 + jb$, the matching can be achieved. This means at location B the admittance must be a transformed version of $1 + jb$ by a distance $3\lambda/8$ away from the generator. Since the admittance $1 + jb$ is represented by $g = 1$ constant conductance circle, the impedance at B must lie on a circle which is generated by rotating every point on the $g = 1$ circle by $270° = \beta\frac{3\lambda}{8}$ around the center of the Smith chart in the anticlockwise direction. The first stub essentially helps in bringing the admittance to lie on the rotated circle.

Working backwards now, one can write the steps involved in the double stub matching as follows (see Fig. 2.49):

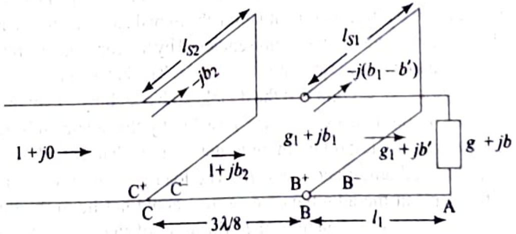


Fig. 2.49 *Double-stub matching configuration.*

1. Mark the admittance $g + jb$ on the Smith chart (Point A).
2. Move on constant VSWR circle passing through A by a distance l_1 to reach B^-. Let the admittance be $g_1 + jb'$.

3. Move along the constant-conductance (constant-g) circle to reach $B^+(g_1 + jb_1)$ (a point on the rotated $g = 1$ circle). Note that a stub at B will change only the reactive part and therefore we move on a circle which keeps the real part of $g_1 + jb'$ same while going from B^- to B^+.
4. Transform admittance $g_1 + jb_1$ at B^+ to C^- by moving a distance of $3\lambda/8$ on a constant VSWR circle passing through B^+. The point C^- must be lying to the $g = 1$ circle. Let the transformed admittance at point C^- be $1 + jb_2$.
5. Add a stub to give susceptance $-jb_2$ at location C so as to move the point C^- to C^+ which is the matched point.
6. To calculate the length of the first stub l_{s1} we note that this stub must provide a susceptance which is the difference between the susceptances at B^+ and B^-. That is, the stub susceptance b_{s1} is equal to $(b_1 - b')$. Mark the susceptance $j(b_1 - b')$ on the chart to get point S_1. Distance from S_1 to S in the anticlockwise direction gives the length l_{s1} of the first stub.
7. The second stub should have a susceptance of $-jb_2$. To get the length l_{s2} of the second stub the procedure is same as that used in the single stub matching. That is, mark $-jb_2$ on the Smith chart to get point S_2. Measure distance S_2S in anti-clockwise direction to give l_{s2}.

From the above discussion, it might appear that the double stub matching is the ultimate solution for impedance matching as it has overcome the drawbacks of the single-stub matching technique. However, if we observe carefully we will note that there is some problem with this technique. The whole matching process relies on the fact that by moving along a constant conductance circle one can go from point B^- to B^+. (B^+ lies on the rotated $g = 1$ circle). If this step is not realizable then the whole matching process is unrealizable. Knowing the nature of the constant-g circles, that they lie one inside another we note that if the point B^- lies in the hatched region (Fig. 2.50), a movement along a constant-g circle can never bring the point outside the hatched region and consequently can never reach the rotated $g = 1$ circle. In this situation, the impedance matching can not be obtained. The hatched region is the region enclosed by the constant-g circle which is externally tangential to the rotated $g = 1$ circle. The hatched region is called the forbidden region. So, we find that the double-stub matching technique cannot match all the load admittances as it was possible by the single-stub matching technique. However, the important thing to note here is that as such there is no constraint on the load admittance, $g + jb$. The load $g + jb$ may lie anywhere on the Smith chart but the admittance $g_1 + jb'$ should not lie in the forbidden region. Since $g_1 + jb'$ is a function of l_1 (location of the first stub), one can always vary l_1 so as not to get $g_1 + jb'$ in the forbidden region. In brief, if l_1 can be varied appropriately the double stub matching also can provide matching for all load admittances. One would then wonder, if the location of the stub, l_1, is to be varied, in what way is this technique superior to the single stub technique! The two situations are not quite identical. In previous case l_1 had to be precisely

adjusted for each load, whereas in the present case l_1 should be chosen so as to take the point B^- out of the forbidden region.

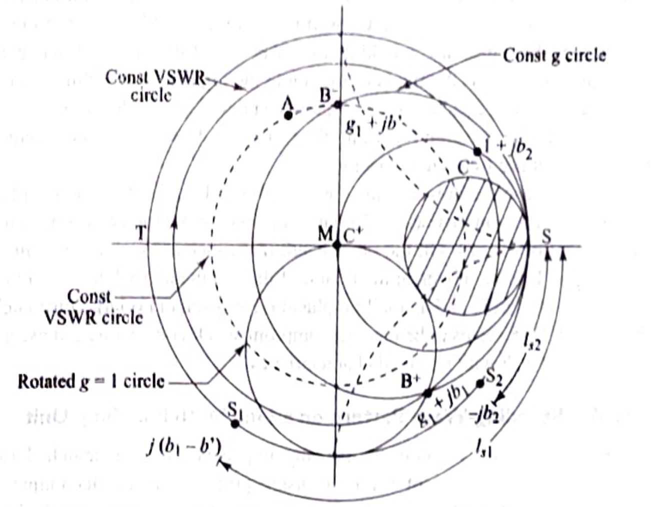


Fig. 2.50 *Double stub matching using the Smith chart.*

As a solution to the above problem, one can use three stubs seperated by $3\lambda/8$ distance, instead of the two stubs (see Fig. 2.51). One still uses only two adjacent stubs (B, C or C, D) at a time for matching purpose. For those loads which have $g_1 + jb'$ outside the forbidden region we use stub B and C as used earlier and stub D is disconnected. The disconnection is not physical and is achieved by simply adjusting the length of the stub to $\lambda/4$. The $\lambda/4$-stub appears open at the junction point D.

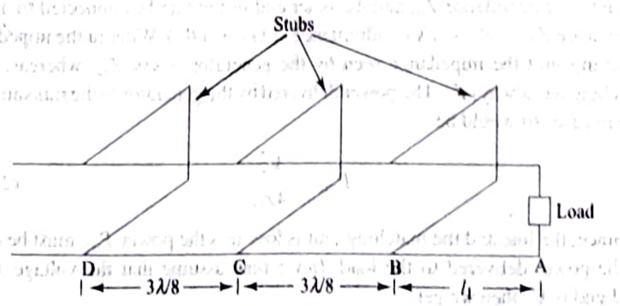


Fig. 2.51 *Triple Stub matching configuration.*

If the admittance at B happens to lie in the forbidden region, disconnect stub B by adjusting its length to $\lambda/4$, and use stubs C and D for double-stub matching. Disconnection of stub B effectively increases l_1 by $3\lambda/8$ and the admittance $g_1 + jb'$ now corresponds to location C. So if the admittance at B is in the forbidden region, the admittance at C would definitely be out of it. A three-stub matching technique therefore, is the final destination in the impedance matching, as it can match all impedances/admittances without changing the stub locations.

Coming to the question regarding the separation between the stubs which is $3\lambda/8$, one can use following reasoning:

If we decrease the stub spacing, the rotated $g = 1$ circle moves leftwards and the forbidden region increases. The largest forbidden region corresponds to the stub separation of $\lambda/4$. On the other hand, if we increase the stub spacing, the rotated $g = 1$ circle moves in the region of the Smith chart where the gradients of g and b are large, and a small graphical error results into large mismatch. A distance of $3\lambda/8$ seems to be the best compromise. Of course, one can use a stub separation of $\lambda/8$ with identical characteristics.

2.13.4 Standing-Wave Pattern on a Line With Matching Unit

In the last section, we noticed that any impedance can be matched to the characteristic impedance of a line by using proper matching technique. It is interesting to note that even highly reactive impedances can be matched to the characteristic impedance which is purely real. After the matching unit every impedance appears as the characteristic impedance and the reflection coefficient becomes zero making the VSWR on the line unity. However, between the matching unit and the load impedance one would observe the standing waves which are similar to the ones observed without the matching unit. Figure 2.52(a) shows the standing waves on the line without the matching unit and Fig. 2.52(b) shows the standing waves with the matching unit. The standing wave pattern beyond B in Fig. 2.52(b) is constant as there is perfect matching of impedances.

Let the input of the line be connected to a voltage source with amplitude V_0 and internal impedance Z_0, and the other end of the line be connected to a load impedance $Z_L = R_L + jX_L$ (admittance $= G_L + jB_L$). Without the impedance matching unit the impedance seen by the generator is say Z_{in}, whereas, after matching it is always Z_0. The power delivered by the generator to the transmission line in case (b) would be

$$P_{out} = \frac{V_0^2}{4Z_0} \tag{2.176}$$

Since, the line and the matching unit is loss-less the power P_{out} must be equal to the power delivered to the load. If we now assume that the voltage at the load-end is V_L' then we get

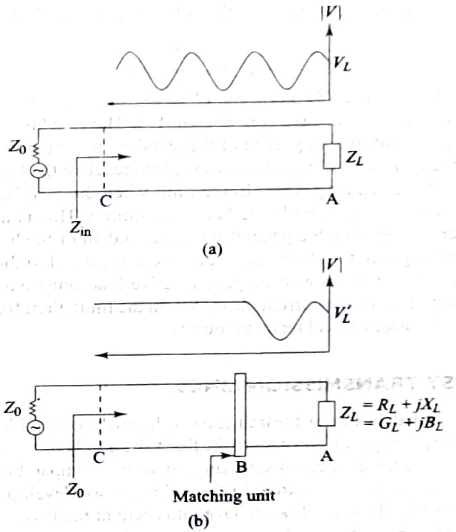


Fig. 2.52 *(a) Voltage standing waves without the impedance matching unit (b) Voltage standing waves with the impedance matching unit.*

$$P_{out} = \frac{V_0^2}{4Z_0} = Re\left[V_L'\left(\frac{V_L'}{Z_L}\right)^*\right] \tag{2.177}$$

$$\Rightarrow \quad |V_L'|^2 = \frac{V_0^2}{4Z_0G_L} \tag{2.178}$$

$$\Rightarrow \quad |V_L'| = \frac{V_0}{2}\sqrt{\frac{1}{Z_0G_L}} \tag{2.179}$$

Suppose, for sake of argument if we assume that the length of the line is integer multiples of $\lambda/2$, then without matching unit the input impedance Z_{in} will be same as Z_L. The magnitude of the voltage at point C and also at point A then will be

$$|V_C| = \left|\frac{V_0 Z_L}{Z_0 + Z_L}\right| \tag{2.180}$$

Since, the voltage at point A is same as the voltage V_L across the load Z_L, we get

$$|V_L| = \frac{V_0 R_L}{|(R_L + Z_0) + jX_L|} \tag{2.181}$$

From Eqns (2.179) and (2.181) one can observe that $|V_L|$ is not equal to $|V_L'|$. One can verify that $|V_L'|$ is always greater than $|V_L|$. The conclusion, therefore, is with the matching unit, the amplitude of the standing wave pattern between the load and the matching unit always increases. More reactive the load is, higher is the amplitude of the standing wave. In the limit when the load becomes purely reactive, the standing wave amplitude becomes infinite. This is due to the fact that, the power delivered by the generator is independent of the load as after the matching unit the generator always sees an impedance equal to the charactristic impedance. To absorb the power a highly reactive load must have a very high voltage across it or very high current through it. In the limit, therefore, the voltage across the purely reactive load tends to infinity.

2.14 LOSSY TRANSMISSION LINES

Till now we assumed the loss on the transmission line to be negligible and therefore we assumed the propagation constant of the line to be purely imaginary ($\gamma = j\beta$). In situations where the attenuation constant α becomes comparable to β, the line characteristics have to be re-visited. In the following we develop modifications to the analysis of the loss-less lines to take into account the loss.

If we take the line to be very lossy, i.e. $R >> \omega L$ and $G >> \omega C$, we get the characteristic impedance

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{R}{G}} \text{ (real)} \tag{2.182}$$

and the propagation constant

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \approx \sqrt{RG} \text{ (real)} \tag{2.183}$$

Comparing this with that of a loss-less case we note that Z_0 is real (resistive) in both cases but γ is purely imaginary for a loss-less line but purely real for a very lossy line. The almost resistive nature of the characteristic impedance therefore does not guarantee the line to be loss less. In general, complex $Z_0 (\equiv R_0 + jX_0)$ and complex $\gamma (= \alpha + j\beta)$ represent moderately lossy lines.

2.14.1 Standing Waves on a Lossy Line

The voltage on a lossy transmission line can be written as (see Eqn (2.28))

$$V = V^+ e^{-(\alpha + j\beta)x} + V^- e^{+(\alpha + j\beta)x} \tag{2.184}$$

The voltage amplitude of the incident wave is $V^+ e^{-\alpha x}$ and that of the reflected wave is $V^- e^{+\alpha x}$. The forward (incident) wave, therefore, exponentially dies

down as it travels away from the generator (Remember $+x$ indicates away from generator). Similarly, the reflected wave grows away from the generator or decays towards the generator. Since the reflected wave originates from the load end, its exponentially decaying behavior with distance is identical to that of the incident wave (Fig. 2.53)

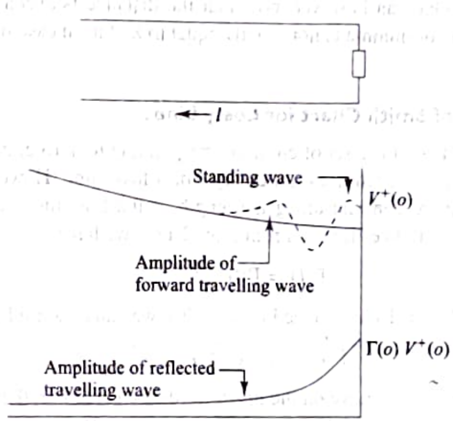


Fig. 2.53 *Amplitude variation of voltage waves on a lossy transmission line.*

The reflection coefficient at any point on the line is (see Eqns (2.32) and (2.37))

$$\Gamma(l) = \Gamma(0)e^{-2\gamma l} = \left(\frac{Z_L - Z_0}{Z_L + Z_0}\right) e^{-2\gamma l} \tag{2.185}$$

Substituting for $\gamma = \alpha + j\beta$ we get

$$|\Gamma(l)| = |\Gamma(0)| e^{-2\alpha l} \tag{2.186}$$

Since α is positive and l is positive towards the generator, $\exp(-2\alpha l)$ becomes smaller and smaller as we move towards the generator. In other words the reflection coefficient which the generator sees is always less than its value at the load. If $2\alpha l >> 1$ then $\Gamma(l) \to 0$ and the generator does not see any reflected wave. Consequently, from the generator end the system appears matched and the generator delivers maximum power to the line input. The power, however, does not get delivered to the load but is lost in the line itself. If the power is not a limitation, one can then use a lossy line for the matching purposes.

The standing wave pattern for a lossy line is as shown in Fig. 2.53. The standing wave slowly dies down to merge with the exponential function of the forward wave. In this case it is not quite correct to define the VSWR as the ratio of the V_{max} and V_{min} on the line as one would not get the same value for the VSWR at different locations on the line. If at all VSWR is to be defined it should be defined as a function of distance as

$$\rho(l) = \frac{1 + |\Gamma(l)|}{1 - |\Gamma(l)|} = \frac{1 + |\Gamma_L|e^{-2\alpha l}}{1 - |\Gamma_L|e^{-2\alpha l}} \quad (2.187)$$

In practice one may use adjacent voltage maximum and minimum to obtain the value of $\rho(l)$. One can however, note that the distance between two adjacent voltage maxima or minima is not exactly equal to $\lambda/2$ as in case of the loss-less line.

2.14.2 Use of Smith Chart for Lossy Lines

Since the Smith chart is a set of circles corresponding to normalized resistances and reactances, it as such does not change for a lossy line. However, since the VSWR is no more constant along the length of the line, the constant VSWR circles have to be reinvestigated. From Eqn (2.185) we have

$$\Gamma(l) = \Gamma(0)e^{-2(\alpha + j\beta)l} \quad (2.188)$$

Since $\Gamma(0) = \Gamma_L \equiv |\Gamma_L|e^{j\theta_L}$, (see Eqn (2.129)), we can re-write Eqn (2.188) as

$$\Gamma(l) = |\Gamma_L|e^{-2\alpha l}\, e^{j(\theta_L - 2\beta l)} \quad (2.189)$$

It is clear that as we move on the line towards the generator, the phase of $\Gamma(l)$ decreases similar to that in the case of a loss-less line. But in this case even the magnitude of $\Gamma(l)$ decreases. The plot of $\Gamma(l)$ as a function of l, therefore, is not a circle but a converging spiral as shown in Fig. 2.54.

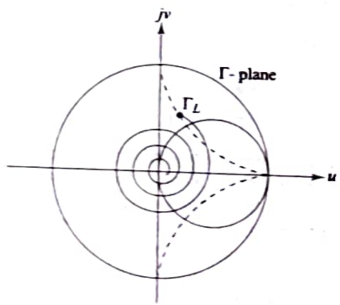


Fig. 2.54 *Variation of reflection coefficient on a lossy transmission line.*

In this case, therefore, there are no constant VSWR circles. For any impedance transformation, ideally one has to draw the Γ-spiral. For moderately low-loss lines however, one can make some approximations. One can assume that even if α is not negligibly small it is much smaller compared to β. Physically, it means that the reduction in the amplitude of a wave due to the loss is small over a distance

of $\lambda/2$. For one section of $\lambda/2$ of a line the VSWR is constant but it is not same for different $\lambda/2$ sections. Figure 2.55 shows the variation of VSWR as a function of distance and its stair-case approximation. In this case, the impedance transformation is carried out in steps of $\lambda/2$.

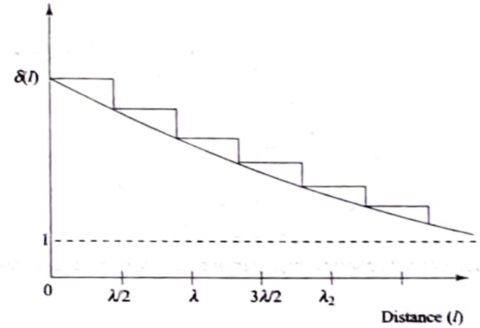


Fig. 2.55 *VSWR as a function of distance on a lossy transmission line.*

As an illustrative example, let us transform an impedance Z_L over a distance say l. In the case of loss-less line we would have marked the impedance Z_L on the Smith chart and would have moved clockwise by a distance l along the constant VSWR circle passing through Z_L to get the transformed impedance. In case of a lossy line the steps however will be as follows: (see Fig. 2.56).

1. Mark the point Z_L on the chart (point P).
2. Draw a constant - VSWR circle through P.
3. Move along the constant VSWR circle by a distance l to reach to point Q.
4. Move the point Q radially towards the center of the chart by a distance δ to get Q′, where δ is the loss correction and is given as

$$\delta = |\Gamma_L|(1 - e^{-2\alpha l}) \quad (2.190)$$

5. Read the value of impedance at Q′.

The VSWR at the load corresponds to point S whereas, its value at a distance of l approximately corresponds to point S′. Generally, a scale for ρ as a function of loss is marked at the bottom of the Smith chart. If not, one can use Eqn (2.189) to obtain the reduction in the reflection coefficient and subsequently the VSWR.

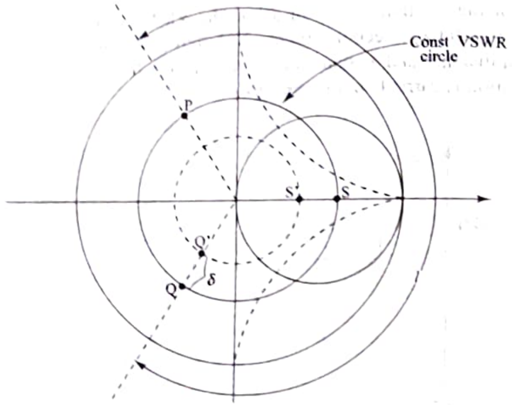


Fig. 2.56 *Use of Smith chart for impedance calculations on a lossy transmission line.*

2.15 MEASUREMENT OF LINE PARAMETERS

We have seen in the beginning of this chapter that, a transmission line can be either characterized by its primary constants R, L, G and C or by its secondary constants γ and Z_0. Usually, the knowledge of γ and Z_0 is sufficient for transmission line calculations and one rarely needs the values of the primary constants. The secondary constants γ and Z_0 of a line are evaluated by conducting short-circuit and open circuit tests on any arbitrary length of the line.

Take any arbitrary length of the line say, L, and measure the input impedance Z_{oc} and Z_{sc}(technique for measuring impedance has been described earlier) of the line with its other end open and short respectively. From Eqn (2.45) we have

$$Z_{oc} = Z_0 \coth \gamma L \tag{2.191}$$

$$Z_{sc} = Z_0 \tanh \gamma L \tag{2.192}$$

Multipying Eqns (2.191) and (2.192) we get

$$Z_0^2 = Z_{sc} Z_{oc} \tag{2.193}$$

$$\Rightarrow \qquad Z_0 = \sqrt{Z_{sc} Z_{oc}} \tag{2.194}$$

Dividing Eqn (2.192) by Eqn (2.191) we get

$$\tanh \gamma L = \sqrt{\frac{Z_{sc}}{Z_{oc}}} \equiv A \text{ (say)} \tag{2.195}$$

where A is a complex number.

Now expanding $\tanh \gamma l$ we get

$$\frac{e^{\gamma L} - e^{-\gamma L}}{e^{\gamma L} + e^{-\gamma L}} = A \tag{2.196}$$

$$\Rightarrow \qquad e^{2\gamma L} = \frac{1+A}{1-A} \equiv R\ e^{j\theta} \text{ (say)} \tag{2.197}$$

$$\Rightarrow \qquad e^{2\alpha L}\, e^{j2\beta L} = R\ e^{j(\theta \pm 2m\pi)} \tag{2.198}$$

where, m is an integer. Separating the real and imaginary parts we get

$$\alpha = \frac{1}{2L} \ln R = \frac{1}{2L} \ln \left|\frac{1+A}{1-A}\right| \tag{2.199}$$

$$\beta = \frac{1}{2L}(\theta \pm 2m\pi) = \frac{1}{2L}\left[\angle\left(\frac{1+A}{1-A}\right) \pm 2m\pi\right] \tag{2.200}$$

Using Eqn (2.199) one can uniquely evaluate the attenuation constant of the line, whereas calculation of β using Eqn (2.200) has an ambiguity due to unknown m. The ambiguity is expected as the phase characteristics of a transmission line repeat over every $\lambda/2$ distance. The information about how many $\lambda/2$ sections are there in the length of the line is not present in the measured input impedance.

The ambiguity in β can be resolved either by choosing a length L less than $\lambda/2$, or by multifrequency measurements. In the first choice, $L < \lambda/2$ and therefore, $m = 0$ giving

$$\beta = \frac{1}{2L}\angle\left(\frac{1+A}{1-A}\right) \tag{2.201}$$

It should be noted however, that due to short length the calculation of α becomes inaccurate in this case.

At this point it is worthwhile to ask the question, as to why one has to measure β at all? Since $\beta = 2\pi/\lambda$ and the wavelength $\lambda = c/f$, where c is the velocity of light, we can calculate β directly if we know the frequency of operation f using

$$\beta = \frac{2\pi f}{c} \tag{2.202}$$

Of course we intrinsically assumed here that the wave travels with the velocity of light, c. In fact this assumption is not always true (this fact will become clear in the subsequent sections of this book). If velocity of the wave is say v, the phase constant β can be written as

$$\beta = \frac{2\pi f}{v} \tag{2.203}$$

Evaluation of β indirectly evaluates the velocity of a wave on the transmission line.

One can use multi-frequency measurements to obtain v and once v is known, β at a frequency f can be computed using Eqn (2.203) (here, we are assuming that v is independent of frequency).

Let us say we conduct the short and open circuit test on a line of length L at frequency f_1 and get A. Then from Eqn (2.200) we have

$$\beta_1 = \frac{1}{2L}\angle\left(\frac{1+A}{1-A}\right) + 2m\pi \qquad (2.204)$$

Now, increase the frequency slowly till you get Z_{oc} and Z_{sc} same as that at f_1. Let this frequency be f_2. (Note: It has been assumed here that the loss of the cable is same at the two frequencies. This is more or less true if the two frequencies are not widely separated.) Now, since the input impedances are same at the two frequencies f_1 and f_2, the electrical lengths of the line at the two frequencies must be differing by $\lambda/2$, or m must have been increased by 1. The phase constant, would therefore, be

$$\beta_2 = \frac{1}{2L}\angle\left(\frac{1+A}{1-A}\right) + 2(m+1)\pi \qquad (2.205)$$

Subtracting Eqn (2.204) from Eqn (2.205) we get

$$\beta_2 - \beta_1 = \frac{\pi}{L} \qquad (2.206)$$

Using Eqn (2.203) for β_1 and β_2, we get

$$\frac{2\pi f_2}{v} - \frac{2\pi f_1}{v} = \frac{\pi}{L} \qquad (2.207)$$

$$\Rightarrow \qquad v = 2L(f_2 - f_1) \qquad (2.208)$$

The phase constant can be then obtained as

$$\beta = \frac{\pi f}{L(f_2 - f_1)} \qquad (2.209)$$

Although, the above measurement procedure appears very straight forward theoretically, in practice it encounters certain problems. The problems are mainly due to difficulties in realizing ideal open and short-circuit at the line end. An open circuit generally has some fringing capacitance and a short-circuit has a small looping inductance. Multiple measurements on lines of different lengths can help in removing errors due to nonideal short or open circuits at the line ends.

2.16 VARIOUS TYPES OF TRANSMISSION LINES

In practice we find a variety of transmission lines, like the coaxial lines, parallel wire lines, etc. Although, structures like waveguides and optical fibres (explained in later sections) also fall in the catagory of transmission line, we confine our discussion to only those transmission lines which have two conductors separated by a dielectric. In the following sections we describe some commonly used transmission lines and also give expression for their characteristic impedances.

2.16.1 Co-axial Cable

The co-axial cable transmission line is as shown in Fig. 2.57. The outer conductor is hollow circular cylinder and the inner conductor is generally a circular rod. The volume between the two conductors is filled with a dielectric material like teflon or perspex, etc.

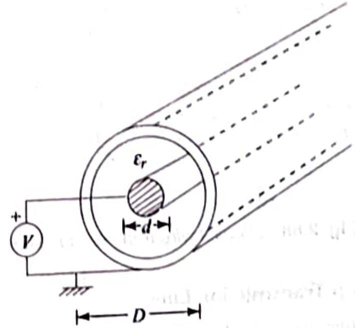


Fig. 2.57 *Co-axial cable.*

The characteristic impedance of a coaxial cable is given as

$$Z_o = \frac{138}{\sqrt{\epsilon_r}} \log_{10}\left(\frac{D}{d}\right) \qquad (2.210)$$

where D is the inner diameter of the outer conductor and d is the diameter of the inner conductor. ϵ_r is the dielectric constant of the material filling the cable (see Fig. 2.57).

2.16.2 Parallel Wire Transmission Line

A parallel wire transmission line consists of two similar conducting circular rods seperated by a dielectric (or air) as shown in Fig. 2.58. The whole structure can be encapsulated in a plastic shell or can be kept open. Due to symmetry, the ground is floating for this configuration and one can assume the potential on the two conductors to be equal and opposite with respect to some hypothetical ground point. The characteristic impedance for a parallel wire line is given by ($D >> d$)

$$Z_0 = \frac{276}{\sqrt{\epsilon_r}} \log \frac{2D}{d} \qquad (2.211)$$

Since D is large compared to d, the characteristic impedance of this line is typically few hundred ohms. This line, therefore, finds application in those cases where high impedances are required. Two standard parallel wire lines of 300Ω and 600Ω are commonly used in telephone networks and antenna feeder systems.

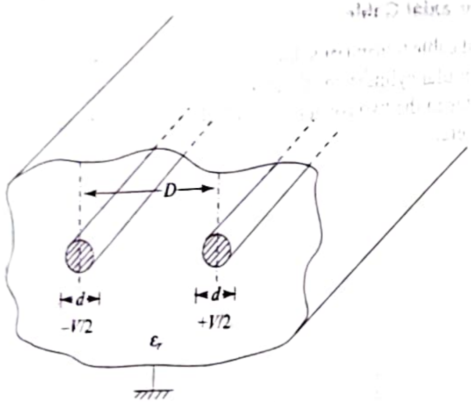


Fig. 2.58 *Parallel wire transmission line.*

2.16.3 Microstrip Transmission Line

This transmission line consists of an infinitely large conducting plane and a flat metal strip placed at a distance from it (Fig. 2.59). The region between the conducting plane and the strip may be filled with a dielectric.

The characteristic impedance for this line is approximately given as

$$Z_0 \approx \frac{377}{\sqrt{\epsilon_r}[(W/h)+2]} \tag{2.212}$$

W and h are shown in Fig. 2.59.

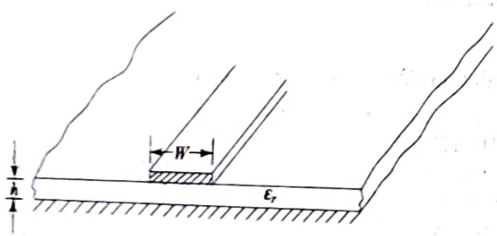


Fig. 2.59 *Micro-strip transmission line.*

This type of transmission lines are encountered in high frequency printed circuits. The characteristic impedance generally lies in a range which is compatible with the co-axial cables.

2.17 SUMMARY

In this chapter, we developed a special but important case of electromagnetic waves, the transmission line. The transmission lines form the foundation of the high frequency circuits. For frequencies beyond a few MHz electrical circuits can not be correctly analyzed without using the concepts of transmission line. Today, when computer and data network speeds are approaching GHz, the knowledge of transmission line is vital to the electrical and computer engineers. A thorough understanding of transmission line also helps in understanding more complex phenomenon of electromagnetic waves in the three dimensional space. In the following chapters, we discuss the general phenomenon of electromagnetic waves, however, we make frequent references to the results of the transmission lines derived and discussed in this chapter.

Review Questions

2.1 What is transit time effect?

2.2 When does transit time effect become appreciable?

2.3 What is a distributed element?

2.4 Why should high frequency circuits be analyzed with distributed element approach?

2.5 What is a transmission line?

2.6 What equations govern voltage and current on a transmission line?

2.7 What is a travelling wave ?

2.8 What is a propagation constant, attenuation constant and phase constant?

2.9 What are the units of attenuation and phase constants?

2.10 What is characteristic impedance? What does it signify?

2.11 If the ratio of voltage and current for a travelling wave is negative. What does it signify?

2.12 What is a standing wave?

2.13 Why do we get standing waves on a transmission line?

2.14 Are standing waves desirable on a transmission line? Why?

2.15 What are full and partial standing waves?

2.16 What is voltage reflection coefficient?

2.17 How is reflection coefficient related to the local impedance?

2.18 What is impedance transformation?

2.19 If impedance at any point on a transmission line is known, can we find impedance uniquely at any point on the line?

2.20 What are loss-less and low-loss transmission lines?

2.21 What is the characteristic impedance of a loss-less transmission line?

2.22 How does reflection coefficient vary along a lossy and loss-less transmission line?

2.23 Over what distance does the impedance repeat along a transmission line?

2.24 What is voltage standing wave ratio?

2.25 What is the range of VSWR?

2.26 How is VSWR related to the reflection coefficient?

2.27 What is normalized impedance and what is its importance?

2.28 What is the impedance at voltage maximum and voltage minimum?

2.29 What is the maximum and minimum impedance seen on a line which is terminated in a particular load?

2.30 What is matched load?

2.31 Why is matching of load impedance important?

2.32 If impedance at some point on a line is Z_0 what will be the load impedance?

2.33 On a loss-less transmission line if the reflection coefficient is Γ, what % of incident power is delivered to the load?

2.34 On a loss-less line the real power is same at every location on the line but the imaginary power varies? Why?

2.35 What is the Smith chart?

2.36 Why is the graphical representation of a transmission line like the Smith chart important?

2.37 In what curves the constant resistance and constant reactance lines in the impedance plane are mapped to in the reflection coefficient plane?

2.38 What is the speciality of constant resistance and constant reactance circle in the Γ-plane?

2.39 As we move along a transmission line, how does a point move on the Smith chart?

2.40 What is constant VSWR circle?

2.41 Identify the load impedances in the following cases:

(a) Partial standing wave and the voltage drops towards the generator at the load point.

(b) Fully standing wave and the voltage rises on the line as we move from load towards generator.

(c) No standing wave.

(d) Partial standing wave with VSWR = 3 and voltage is minimum at the load?

(e) Fully standing wave and voltage is maximum at the load point.

2.42 What is the difference between impedance and admittance Smith charts?

2.43 Name applications of transmission lines.

2.44 Why at high frequencies should the impedance be measured using transmission lines ?

2.45 Can a short circuited section of a transmission line be used for realizing an inductance?

2.46 Under what condition does a short circuited or open circuited section of a transmission lines behave like a series and parallel resonant circuit?

2.47 What is the Q of a section of a loss-less transmission line?

2.48 Can a section of transmission line be used as a voltage or current transformer?

2.49 What is a quarter wavelength transformer? What impedances can be matched using this transformer?

2.50 What is the single stub matching technique? What is its advantage over the quarter wavelength transformer?

2.51 What are the drawbacks of the single stub matching technique?

2.52 What is double stub matching technique ? What is its advantage over the single stub matching technique ?

2.53 What is the limitation of the double stub matching technique and how is it overcome?

2.54 What is a lossy transmission line?

2.55 How does VSWR vary along a lossy transmission line?

Problems

2.1 Find the transit time over a parallel wire transmission line of length 50 cm. A sinusoidal signal of 1.5 GHz is applied to one end of the line. What is the phase difference between the signals at the two ends of the line?

2.2 For a transmission line the primary constants are R = 0.2 Ω/m, G = 0, L = 0.3 μ H/m and C = 15 pF/m. Find the phase and attenuation constants of the line at 900 MHz. Also, find the characteristic impedance of the line.

2.3 In Question 2.2 what should be the value of G so that the line becomes distortionless? What will be the characteristic impedance of the distortionless line?

2.4 On a 50 Ω loss less line two waves travel in opposite directions. At some location on the line and at some instant the voltage and current are respectively 25 V and 2 A. Find the voltage and current at a distance of $\pm\lambda/3$ from the point at the same instant.

2.5 A transmission line has complex propagation constant $\gamma = 1 + j10$ per meter at 1 GHz. Find the distance over which the amplitude of a traveling wave will decrease by 20 %. What phase change will the wave undergo over that distance?

2.6 A transmission line has $Z_0 = 65 + j5\Omega$ and $\gamma = 1 + j20$ per meter. Is the line loss less? Find the primary constants of the line at 900 MHz. For these primary constants, over what frequency range can the line be treated as a low loss line ? (Note: for a low loss line $\alpha << \beta$, say $\alpha < 1\%$ of β).

2.7 A coaxial cable has 10 dB loss per 100 m length. A 10V-3A signal is connected to one end of the 50 m long cable and the other end of the cable is connected to a matched load. Find the power loss in the cable and the power delivered to the load.

2.8 A line has $Z_0 \approx 70\ \Omega$ and $\alpha = 5$ neper per meter. A 100 V/50 Ω generator is connected to one end of a long piece of the line. What is the power supplied by the generator? Over what length of the line about 99 % of the input power will be absorbed by the line?

2.9 A transmission line with $Z_0 = 50 - j5$ and $\gamma = 0.2 + j\,2.5$ per meter is connected to a load impedance of 100 + J 50 Ω. Find the reflection coefficient at the load end of the line. Also, find the reflection coefficient and the impedance at a distance of 3 m from the load.

2.10 On a 50 Ω line the reflection coefficient at the load is $0.7\angle 30^\circ$. If the propagation constant of the line is $\gamma = 20\angle 89^\circ$ per meter. Find the impedance at a distance of 4 m form the load.

2.11 A $\lambda/3$ long 50 Ω loss-less line is connected to a load at one end, and to $100\angle 0$V/50Ω generator at the other end. If the voltage at the input terminal of the line is $75\angle 30^\circ$, find the load impedance and the power delivered to the load.

2.12 A 300 Ω parallel wire line is connected to an antenna of 75 + j35 Ω input impedance. Find the VSWR on the line. What is the maximum and minimum impedance seen on the line?

2.13 On a loss-less transmission line connected to a load the maximum voltage and current are 100 V and 4 A respectively. If 25% of the incident power is reflected by the load find the minimum and maximum impedance seen on the line.

2.14 A 300 Ω loss-less transmission line connected to a load has maximum and minimum currents of 20 A and 12 A respectively. What is the power delivered to the load?

2.15 A 75 Ω loss-less coaxial cable is connected to an antenna of impedance 100 + j35 Ω at one end and to a generator at the other end. The internal impedance of the generator is 50 Ω. What should be the length of the cable so that maximum power is delivered to the antenna?

2.16 A 50 Ω line of length $3\lambda/5$ is connected to an admittance of 0.03 − j0.01 ℧ at one end, and a 50 V − 75 Ω generator at the other end. What are the amplitudes of the forward voltage and current travelling waves on the line? Find the complex powers at the input and load ends of the line.

2.17 A constant VSWR circle is mapped onto the impedance plane. Find the equation of the transformed curve. Identify the curve and plot it for VSWR = 1, 5, ∞.

2.18 Transform impedances which have same phase angle onto the complex reflection coefficeint plane. Draw the transformed curves for phase angles of 0°, 30°, 45°, 60°, 90°.

2.19 What will be the equation of the curve in the complex Γ-plane corresponding to $|Z|$ = constant ?

Smith chart based problems (2.20 to 2.26)

2.20 Mark following points on the smith chart.
(i) $\Gamma = 0.3\angle -45^\circ$
(ii) $\Gamma = 0.2 + j0.5$
(iii) $\bar{z} = 3 + j2$
(iv) $\bar{y} = 0.5 - j2$
(v) short and open circuit on the impedance chart
(vi) short and open circuit on the admittance chart
(vii) $g = 1, b = 0.6$ on the admittance chart
(viii) VSWR = 2, $\angle\Gamma = 120^\circ$
(ix) $Z = 3$ and $x = 2r$
(x) $Z = 2$ and maximum phase angle of Z

2.21 An impedance of 75 − j30 Ω is connected to a 100 Ω line. Find complex reflection coefficient at the load, the complex reflection coefficent and the impedance at a distance of 0.35λ from the load. What is the VSWR measured on the line?

2.22 On a line the maximum and minimum voltages measured are 25 V and 15 V respectively. The distance between adjacent maximum and minimum voltages is 20 cm. If minimum voltage occurs at a distance of 25 cm from the load, find the load admittance. The characteristic impedance of the line is 100 Ω.

2.23 Maximum and minimum impedances seen on line are 150 Ω and 50 Ω. What is the load impedance if maximum current occurs at a distance of 0.15λ from the load?

2.24 A transmission line of 50 Ω characteristic impedance is terminated in a load impedance of 80 − j 30 Ω. At a distance of 0.3λ a short circuited stub made of 75 Ω line is connected in parallel to the main line such that the impedance is real at the junction. What is the length of the stub? What will be the VSWR on the main line beyond the junction towards the generator?

2.25 An inductive reactance of 100 Ω is to be realized using an open circuited section of a 50 Ω line at 2 GHz. Find the length of section. If 10 % variation in the reactance is acceptable, over what bandwidth will the reactance realization be satisfactory?

2.26 A 300 Ω line is connected in a load impedance of 200 -j100 Ω. A short circuted section of the transmission line is connected in parallel with the load impedance. What should be the length of the section so as to get minimum reflection from the load?

2.27 A $\lambda/2$ section of line has 0.05 dB loss. If the line is open circuited at one end, find the impedance at the other end of the line. The characteristic impedance of the line is 50 Ω. What will be the quality factor of the section?

2.28 A 1.25λ long section of a 75 Ω line is short circuited at one end and open circuited at the other. The voltage measured at the mid point of the line is 40 V. If the loss in the line is 0.2 dB per meter and the wavelength of the signal is 5 m, find the energy stored and energy dissipated on the line. Hence, find the quality factor of the line. Assume that the line has a velocity factor 0.66. (velocity factor is the ratio of the velocity of a wave on the line to the velocity of the light in vacuum).

2.29 A short circuited section of a line is to be used as a parallel resonant circuit at 2.4 GHz. Find the length of the section. If the line has 3dB/m loss, what is the 3-dB bandwidth of the resonant circuit?

2.30 Two long cables with characteristic impedances 100 Ω and 200 Ω are to be joined through a quarter wavelength transform at 900 MHz. If the velocity factor for the transformer section is 0.8, find the length and the characteristic impedance of the transformer.

2.31 Two transmission lines with characteristic impedances Z_1 and Z_2 are connected to get a perfect match at frequency f_0 through a quarter wavelength transformer. Derive an expression for power transfer efficiency at the junction as a function of frequency deviation Δf around f_0 ($\Delta f \ll f_0$). For Problem 2.30 if the power transfer efficiency is to be greater than 90 %, find the frequency range over which the transformer can be used.

2.32 An impedance of 50 + j50 Ω is to be matched to 50 Ω using a quarter wavelength transformer. Find the location and the characteristic impedance of the quarter wavelength transformer.

2.33 If the single stub matching is employed for Problem 2.32, what is the location and length of the stub? The stub is a short circuited section of 150 Ω line connected in parallel with the main line.

2.34 An admittance of 0.03 - j0.05 ℧ is to be matched to 50 Ω impedance using single stub matching. Work out all series and shunt, and open and short circuited stub configurations.

2.35 If G = 0 for a transmission line, show that

(a) $X_0 = R_0(1 - LCv_p^2)$ Where $v_p = \omega/\beta$

(b) $\alpha = \beta[\frac{(W_C - W_L)}{(W_C + W_L)}]^{1/2}$

Where W_C is the average energy stored per unit length in the distributed capacitance of the line, and W_L is the average energy stored per unit length in the distributed inductance of the line when the line is terminated in its characteristic impedance.

2.36 What is the condition for distortionless transmission? For a general line find the condition on the primary constants of the distortionless line. What is wave velocity and characteristic impedance of a distortionless line?

2.37 A loss less transmission line has a characteristic impedance of 600 Ω and is quarter wavelength long. Find the voltage at the open circuited receiving end if the other end of the line is connected to a generator which has 50 Ω internal impedance and a generated voltage of 100 V.

2.38 A short circuited transmission line of Z_0 = 50 Ω is fed with a 250 V generator. Internal impedance of generator is 75Ω. If the length of the line is λ/6, find the current at the short circuited end of the line.

2.39 At 2 MHz the input impedance of a 5m long coaxial line under short and open circuit conditions are 17 + j20 Ω and 120 - j 140 Ω respectively. Is the line loss-less? Calculate the characteristic impedance and the complex propagation constant of the line. Velocity of wave on the transmission line is greater then 2×10^8 m/sec.

2.40 A coaxial transmission line of characteristic impedance 50 Ω is to deliver 10 kW of power at a frequency of 10 MHz over a distance of several hundred feet to a load. The rms value of current in the line conductor must not exceed 25 Amp at any point. What is the highest value of VSWR that can be tolerated on the line?

2.41 A transmission line 10 m long operating at 100 MHz has a characteristic impedance of 600 Ω, an attenuation factor of 0.3 dB/m. The line is to provide a signal voltage of 40 V rms across a terminal load of 100 Ω.
(a) What is required rms voltage at the input terminals of the line ?
(b) What are peak values of voltage and current on the line and where do they occur? (make use of computer)

2.42 Consider a lossless line terminated in a load impedance of 40 + j30 Ω. Determine Z_0 for the line to get minimum VSWR on the line. Find this minimum VSWR and the location of the voltage minimum from the load.

2.43 Show that a lossy line, in general, is dispersive i.e. the velocity of the wave is a function of frequency.

2.44 A lossless transmission line with Z_0 = 50Ω is terminated in a load which is a series combination of 60 Ω resistance and 50 μH inductance. The length of the line is 100 m. If a 1 MHz square wave of 2 V pp amplitude is launched on the line, find voltage waveform across the load. Use computer program if necessary. Plot the voltage waveform to the scale.

2.45 For lossy transmission line the characteristic impedance $Z_0 = R_0 + jX_0$. If R,L,G,C are resistance, inductance, conductance and capicitance/unit length of the line show that

$$\alpha = \frac{R}{2R_0} + \frac{G|Z_0|^2}{2R_0}$$

$$\beta = \frac{\omega L}{2R_0} + \frac{\omega C|Z_0|^2}{2R_0}$$

2.46 A transmission line of Z_0 = 50 Ω is terminated in a load to give a VSWR of 3. The first voltage maximum is at 5 cm from the load and the next is at 25 cm from the load. Find the minimum value of VSWR that can be obtained by placing a stub in parallel with the line at the load.

2.47 Find an expression for the quality factor for a resonant section of a transmission line using frequency response of the circuit.

2.48 A lossy line has a characteristic impedance of 50Ω. About 100 m length of this line is terminated with an impedance 200 + j 50 Ω. The magnitude of the measured input impedance of the line is within 10 percent of the characteristic impedance. Calculate the value of the attenuation factor in dB/m.

2.49 If d is the distance between two points on either side of any voltage minimum at which the magnitude of the voltage is $\sqrt{2}$ times the V_{min}, show that

$$VSWR = \frac{\sqrt{1 + \sin^2(\beta d/2)}}{\sin(\beta d/2)}$$

2.50 A 50 Ω line is terminated in a load impedance of 75 - j 69 ohms. The line is 3.5 m long and is excited by a source at 50 MHz. Velocity of propagation is 3×10^8 m/sec. Find the input impedance, the complex input reflection coefficeint, the VSWR and the position of the voltage minimum.

2.51 The VSWR of an ideal 70 ohm line is measured as 3.2 and a voltage minimum is observed 0.23 wavelength in front of the load. Find the load impedance. Determine the maximum phase angle for the impedance at any point on the line.

2.52 A 70 ohm line is terminated in an impedance of 50 + j 10 ohms. Find the position and length of a short circuited stub required for matching if the stub is to be added in (a) series (b) parallel.

2.53 A 50 ohms transmission line is terminated with a load of 200 + j 300 ohms. A double stub tuner consisting of a pair of short circuited 50 ohm lines connnected in shunt to the main line at points spaced by 0.25 λ is located with one stub at 0.2 λ from the load. Find the lengths of the stubs.

2.54 A transmission line is terminated in a load which can attain any value of admittance given by the lower half of the smith chart. Can these loads be matched by a double stub tuner with stub spacing of a 5λ/16. If yes, design the tuner. If no, give the range of loads which cannot be matched.

CHAPTER 3

Maxwell's Equations

In the previous chapter, we discussed the limitations of the lumped element analysis for high frequency circuits and introduced the concept of distributed elements. Apart from the normal independent parameter 'time', 'space' was also introduced in the circuit analysis. The electrical quantities however were still voltages and currents which were scalar in nature. One can quickly realize that the concept of voltage and current is rather suitable for circuits having wires, resistances, capacitances, etc. If one goes to non-circuit like configurations, use of voltage or current becomes rather inappropriate. For example if one has to investigate the unbound medium like free space, the use of voltage and current appears unattractive. In this situation, one then naturally has to fall back on the more fundamental quantities like electric and magnetic fields. Due to vector nature of these fields, the analysis becomes relatively more involved compared to that with voltages and currents. However, we will see that some of the concepts developed with voltage and current can be extended to the field analysis and it would be a good practice to validate the generalized field concepts with the special examples of transmission line.

In this chapter, we develop the basic vector equations which govern the electric and magnetic fields. These equations are called the Maxwell's equations. In general, we develop the equations for the time varying electric and magnetic fields, which can be reduced to their static form by appropriate substitution for the time varying function. Before we get into the derivation of the Maxwell's equations some basic definations regarding co-ordinate systems, a brief discussion on the vector operations, and a few vector theorems are in order. It is essential that a consistent set of conventions be followed throughout so as not to commit any mistake in defining the vector directions. The basics of vector algebra and vector calculus are given in Appendices B and C.

It is an interesting phenomenon that the most fundamental of the electromagnetic fields is the charge. A stationary charge produces an electric field. However, when the same charge is kept in uniform motion, it constitutes a

current and then we get magnetic field. So, at the origin of the electric or magnetic fields we have nothing but charges. The presence of stationary charge is felt by the electric field which the charge generates around it. Similarly, the presence of current is felt by the magnetic field which is generated by it.

In the following sections we develop basic definitions of electric and magnetic fields and then establish relation between them using laws of electromagnetics.

3.1 BASIC QUANTITIES OF ELECTROMAGNETICS

3.1.1 Electric Field E and Electric Displacement Density D

Let us consider a charge Q at some location in space. For simplicity we generally assume the charge to be located at the origin of the co-ordinate system as shown in Fig. 3.1.

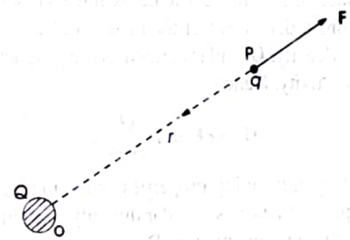


Fig. 3.1 *Force between two charges.*

Let us now place a test charge q at a distance r from the origin. Then according to the 'Coloumb's law' the test charge q will experience a force **F**. The force will be attractive if Q and q are of opposite sign, and the force will be repulsive if Q and q are of same sign. Consequently, for charges with same polarity, the force will have direction indicated by the thick arrow, and for charges with opposite polarity the force will be in the direction of the dotted arrow. In both the cases, the force vector is along the line joining the two charges. The force is given as

$$\mathbf{F} = \hat{\mathbf{r}}\frac{Qq}{4\pi\epsilon r^2} \tag{3.1}$$

Here, $\hat{\mathbf{r}}$ is the unit vector in the direction of the thick arrow.

The electric field **E** *at point P is defined by the force per unit test charge*, giving,

$$\mathbf{E} = \frac{\mathbf{F}}{q} = \hat{\mathbf{r}}\frac{Q}{4\pi\epsilon r^2} \tag{3.2}$$

The force **F** and **E** are vector quantities and they have same direction. It is interesting to note that as the test charge q reduces to zero, the force **F** also reduces to zero but the electric field **E** remains unchanged. The electric field therefore does not require a test charge for its existance although its presence can be felt only by placing a test charge on that location.

The quantity ϵ is a characteristic parameter of the medium surrounding the charge Q, and is called the permittivity of the medium. Even no medium like vacuum has a finite permittivity, and is generally denoted by ϵ_0. The permittivity of the free-space (vacuum) is

$$\epsilon_0 = \frac{1}{36\pi} \times 10^{-9}\ \text{F/m} \tag{3.3}$$

Generally, other media have permittivities ϵ higher than that of vaccuum. The ratio of ϵ and ϵ_0 is called the relative permittivity or the dielectric constant of the medium and is denoted by ϵ_r.

$$\epsilon_r = \frac{\epsilon}{\epsilon_0} \tag{3.4}$$

We observe here that, since, the permittivity is different for different media, the same charge Q produces different electric fields at the same distance r from it. A quantity which remains independent of the medium characteristics is called the electric displacement density, $\mathbf{D}$, and is defined as the product of the electric field and the medium permittivity. Hence,

$$\mathbf{D} = \epsilon\mathbf{E} = \hat{\mathbf{r}}\frac{Q}{4\pi r^2} \tag{3.5}$$

If the medium is isotropic, meaning properties of the medium are not direction dependent, then the permittivity ϵ is a scalar quantity. For this medium it can be seen from Eqn (3.5) that the direction of $\mathbf{D}$ is same as that of $\mathbf{E}$. On the other hand, for an anisotropic medium (a medium for which the properties are direction dependent) the permittivity in general is a tensor and then Eqn (3.5) has to be re-written as

$$\mathbf{D} = \bar{\bar{\epsilon}} : \mathbf{E} \tag{3.6}$$

where $\bar{\bar{\epsilon}}$ is a 3×3 matrix.

In the cartesian co-ordinate system this can be written as

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{3.7}$$

For an anisotripic medium, the direction of $\mathbf{D}$ is not same as that of $\mathbf{E}$. As can be seen from Eqn (3.7) D_x is not produced by only E_x but also by E_y and E_z. Same is true for D_y and D_z. Depending upon the elements of the permittivity tensor, $\mathbf{D}$ can have any arbitrary direction with respect to $\mathbf{E}$. The isotropic case is a special case of Eqn (3.7) where $\epsilon_{xx} = \epsilon_{yy} = \epsilon_{zz} = \epsilon$ and other off diagonal elements $\epsilon_{xy}, \epsilon_{xz}, \epsilon_{yx}, \epsilon_{yz}, \epsilon_{zx}, \epsilon_{zy} \equiv 0$. For isotropic medium Eqn (3.7) reduces to

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & \epsilon \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{3.8}$$

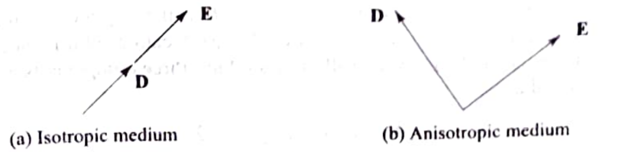


Fig. 3.2 *Electric field and electric displacement vectors for isotropic and anisotropic media.*

Equation (3.8) is same as Eqn (3.5). Figure 3.2 shows the $\mathbf{E}$ and $\mathbf{D}$ for isotropic and anisotropic media. Inside crystals and inside the magnetized plasma (ionosphere) the dielectric constant is a tensor.

3.1.2 Electric Scalar Potential

In the presence of an electric field if we move a charge from one point to another, we have to do a work. The work done is essenıially stored in the charge in the form of potential energy. This energy is characterized by a parameter called the electric potential. Since the work done is a scalar quantity, this potential is called electric scalar potential. The electric potential V of a point P with respect to some reference point O is defined as the work done in carrying a unit positive charge from point O to P. Then by defination the potential of the reference point is zero as the work done in carrying a unit charge from O to itself is identically zero.

Let us say that the point P_1 is at an infinitesimally small distance 'dl' from P (Fig. 3.3).

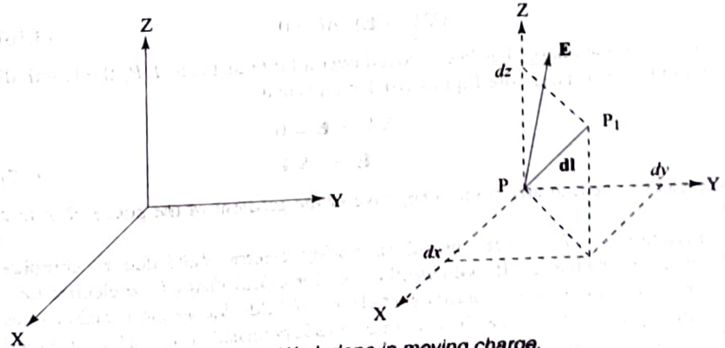


Fig. 3.3 *Work done in moving charge.*

Then if the electric field at P is $\mathbf{E}$ the work required to move a unit charge from P to P_1 is

$$dV = -\mathbf{E} \cdot d\mathbf{l} \tag{3.9}$$

Without losing generality, we can take the point P as the origin of a co-ordinate system. Since, the point P_1 can have any arbitrary location with respect to the co-ordinate axes, the length vector **dl** in general has three components along the three principal axes.

$$\mathbf{dl} = dx\hat{\mathbf{x}} + dy\hat{\mathbf{y}} + dz\hat{\mathbf{z}} \tag{3.10}$$

Similarly, the electric field can be written in its components as

$$\mathbf{E} = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} + E_z\hat{\mathbf{z}} \tag{3.11}$$

Substituting for **E** and **dl** in Eqn (3.9) we get

$$dV = -(E_x dx + E_y dy + E_z dz) \tag{3.12}$$

Now, if the potential of the point P is V, the change in potential over infinitesimal distance **dl** (using differential calculus) is

$$dV = \frac{\partial V}{\partial x}dx + \frac{\partial V}{\partial y}dy + \frac{\partial V}{\partial z}dz \tag{3.13}$$

Realizing that

$$\nabla V = \hat{\mathbf{x}}\frac{\partial V}{\partial x} + \hat{\mathbf{y}}\frac{\partial V}{\partial y} + \hat{\mathbf{z}}\frac{\partial V}{\partial z} \tag{3.14}$$

We can write Eqn (3.13) as

$$dV = \nabla V \cdot \mathbf{dl} \tag{3.15}$$

Substituting for dV in Eqn (3.9) we get

$$(\nabla V + \mathbf{E}) \cdot \mathbf{dl} = 0 \tag{3.16}$$

Now, since, the charge has been moved over a finite distance PP_1 the length dl can not be zero. Therefore Eqn (3.16) demands that

$$\nabla V + \mathbf{E} = 0$$

$$\Rightarrow \qquad \mathbf{E} = -\nabla V \tag{3.17}$$

The electric field at any point is negative of the gradient of the potential at that point.

Equation (3.17) is very useful in finding electric field due to complex distribution of charges. If one directly goes for calculation of the electric field one has to carry out vector addition of electrical fields due to individual charges which is quite cumbersome. Instead, one can add potentials due to all charges and then can find its gradient to give the electric field. Since the potentials are scalars their addition is much simpler compared to the addition of the electric fields.

Equation (3.17) is also used in defining the unit of the electric field. The potential V has a commonly used unit of volt. The operator ∇ is a spatial derivative and therefore has the unit of m^{-1}. The electric field hence should have a unit V/m or Vm^{-1}. Instead of Eqn. (3.17) if we had used Eqn (3.2) we would get the unit of

electric field as Newton/coulomb since the electric field is force per unit charge. So, Volt/meter and Newton/coulomb are equivalent units. The commonly used unit of the electric displacement density **D** will be derived later after we have covered more laws of electromagnetics.

3.1.3 Magnetic Field and Magnetic Flux Density

As we have mentioned earlier, the stationary charges give rise to electric field whereas the charges in motion produce magnetic field. Since the charges in motion constitute a current, we can take the current as the starting point for the generation of the magnetic field. As the presence of a charge is felt by the electric field produced by it, the presence of a current can be felt by the magnetic field generated by it. The definition of magnetic field could have been written on the lines similar to that of the electric field, i.e. *the magnetic field is the force experienced by a unit magnetic monopole*. However, there are no magnetic monopoles in reality like the electric charges. The magnetic field therefore has to be defined differently. Due to a current element the magnetic field is defined by the Biot-Savart law.

Consider an infinitesimally small piece of a wire carrying current I in it. The length of the piece is say dl. The current moment is then defined by the product Idl. Since, the piece of the wire can be oriented in any direction the length dl is a vector quantity and therefore should be denoted by **dl**. The current I is a scalar quantity making the current moment I**dl** a vector quantity. Without loss of generality, let us assume that the current element is located at the origin of the co-ordinate system as shown in Fig. 3.4.

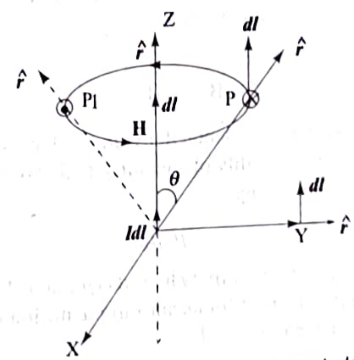


Fig. 3.4 *Magnetic field due to a current element.*

According to Biot-Savart law the magnetic field **dH** (field due to infinitesimally small current element) at a point P is given as

$$\mathbf{dH} = \frac{I\mathbf{dl} \times \hat{\mathbf{r}}}{4\pi r^2} \tag{3.18}$$

where $\hat{\mathbf{r}}$ is the unit vector in the direction of the line joining the current element and the observation point P. Since, the vectors **dl** and $\hat{\mathbf{r}}$ for this point (and also for point P_1) lie in the plane of the paper, the cross-product of the two has a direction perpendicular to the plane of the paper. Since, the direction of the current is assumed upwards, the direction of **dl** is upwards and the direction of the magnetic field **dH** will be going into the paper as indicated by $\otimes$ at point P. If we take some other observation point P_1 on the left of the current element, by right hand screw rule, the magnetic field will come out of the paper as indicated by $\odot$. If one streches his imagination a little, he can then see that the magnetic field forms a circular loop around the axis of the current element.

From Eqn (3.18) it is clear that the magnitude of the magnetic field is directly proportional to the current moment and is inversely proportional to the square of the distance between the source and the observation point. The cross-product $\mathbf{dl} \times \hat{\mathbf{r}}$ also suggests that the strength of the magnetic field is maximum when **dl** and $\hat{\mathbf{r}}$ are perpendicular to each other, that is, for θ equal to $\frac{\pi}{2}$. As θ reduces, the angle between **dl** and $\hat{\mathbf{r}}$ reduces and consequently the strength of the magnetic field reduces. For extreme case of $\theta = 0$ the vectors **dl** and $\hat{\mathbf{r}}$ are colinear making cross product and hence the magnetic field identically zero. The magnetic field, therefore, is identically zero at the axis of the current element.

It should also be noted that in Eqn (3.18) there is no quantity which depends upon the medium property. The magnetic field induced by a current element is, therefore, independent of the medium surrounding it.

Another quantity which characterizes the magnetic field is the magnetic flux density **B**. The magnetic flux density is defined as the total magnetic lines of forces passing through a unit area placed perpendicular to them. The magnetic field **H** is related to **B** through a medium characteristic parameter called 'permeability' of the medium as

$$\mathbf{B} = \mu \mathbf{H} \tag{3.19}$$

The permeability of vacuum is denoted by μ_0 and its value is $4\pi \times 10^{-7}$ Henry/meter. A ratio of permeability of a medium to that of the vacuum is called the relative permeability μ_r giving

$$\mu = \mu_r \mu_0 \tag{3.20}$$

As we have seen in case of permittivity here also μ is a scalar for a magnetically isotropic medium and it is a tensor for an anisotropic medium. For an anisotropic medium, therefore, we have in general

$$\mathbf{B} = \bar{\bar{\mu}} : \mathbf{H} \tag{3.21}$$

where $\bar{\bar{\mu}}$ is a 3×3 matrix.

In the cartesian co-ordinate system this can be written as

$$\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = \begin{bmatrix} \mu_{xx} & \mu_{xy} & \mu_{xz} \\ \mu_{yx} & \mu_{yy} & \mu_{yz} \\ \mu_{zx} & \mu_{zy} & \mu_{zz} \end{bmatrix} \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} \tag{3.22}$$

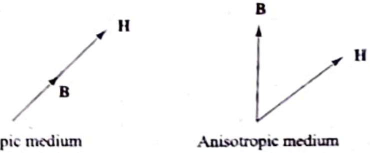


Fig. 3.5 *Magnetic field and magnetic flux density vectors in isotropic and anisotropic media.*

As shown in Fig. 3.5, in an isotropic medium the directions of **B** and **H** are same, whereas they are different in an anisotropic medium.

It should be emphasized here that an anisotropic medium does not mean that it has both ϵ and μ as tensors. A medium can be electrically anisotropic but magnetically isotropic and vice-versa.

The unit of magnetic field **H** can be derived from Eqn (3.18). On the right hand side of Eqn (3.18) the unit of current is Ampere(A), the unit of **dl** is meter(m), the unit of distance r is meter and the unit vector $\hat{\mathbf{r}}$ is unitless. The unit of the magnetic field therefore is Ampere/meter. The unit of magnetic flux density is either 'Tesla' or Webers/m^2.

3.1.4 Conduction Current Density J

In situations where the current has spatial variation it is rather inadequate to define just the total current flow in the system. One can then use the current density as the primary parameter which is defined as the current flow per unit area.

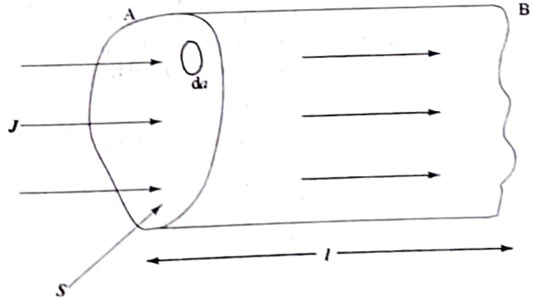


Fig. 3.6

Consider a cylinder made of a conducting material with an arbitrary infinitesimal cross-section S (see Fig. 3.6). Let us assume that the current density **J** exists in the cylinder over a length 'l'. If the conductivity of the cylinder is σ

its resistivity is $1/\sigma$. The resistance R of the cylinder between its two ends A and B is then given by

$$R = \frac{\text{Resistivity} \times \text{length}}{\text{area of cross section}} = \frac{1}{\sigma} \cdot \frac{l}{S} \tag{3.23}$$

For simplicity assuming that the current density is constant all over the cross-section, the total current in the conductor is

$$I = JS \tag{3.24}$$

Due to the current flow in the conductor, there is a voltage drop V between point A and B. If we assume the conducting cylinder to be of infinitesimal size we can assume the electric field associated with the voltage drop to be constant between A and B. Then one can write the voltage V to be

$$V = El \tag{3.25}$$

Now from Ohm's law we have

$$V = IR \tag{3.26}$$

Substituting for V, I and R from Eqns (3.23), (3.24) and (3.25) we get

$$El = JS\frac{1}{\sigma}\frac{l}{S} \quad \Rightarrow \quad J = \sigma E \tag{3.27}$$

The magnitude of the conduction current density J is proportional to the electric field strength E. Now, from the Ohm's law we also know that the current flows in the direction in which the potential drop is maximum. That is, the current flows in the direction in which the potential V has maximum change. This direction is opposite to the direction of the gradient of V. From Eqn (3.17) the direction opposite to gradient of V is same as that of the electric field **E**. Therefore, we can conclude that not only the magnitude of J is proportional to E but its direction is also same as that of E. The Eqn (3.27) for vector **J** and **E** can be written as

$$\mathbf{J} = \sigma \mathbf{E} \tag{3.28}$$

For anisotropic materials the conductivity also is a tensor and then the Eqn (3.28) has to be re-written as

$$\mathbf{J} = \overline{\overline{\sigma}} : \mathbf{E} \tag{3.29}$$

where the tensor conductivity is given as

$$\overline{\overline{\sigma}} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \tag{3.30}$$

For an anisotropic material the direction of the conduction current is not same as that of **E** as is evident from Eqns (3.29) and (3.30). For a medium like the ionosphere the conductivity is a tensor.

3.2 BASIC LAWS OF ELECTROMAGNETICS

In the previous section, we got introduced to some of the fundamental quantities of electromagnetics like electric field **E**, magnetic field **H**, etc. In the following sections, we state physical laws which establish relationship between various electromagnetic quantities. We will see that the laws can be mathematically written in integral or differential form. Depending upon the suitability one can choose the integral form or the differential form.

3.2.1 Gauss's Law

The Gauss's law states that, the *total outward electric displacement through any closed surface surrounding charges is equal to the total charge enclosed.*

Consider a charge Q at some location in space (Fig. 3.7). The electric displacement density **D** at some point P at a distance r from Q is given by Eqn (3.5) as

$$\mathbf{D} = \hat{\mathbf{r}}\frac{Q}{4\pi r^2} \tag{3.31}$$

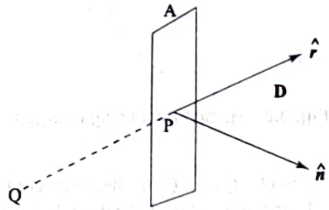


Fig. 3.7 *Electric displacement vector and area normal.*

Suppose, we place a frame enclosing area A at location P. Then the electric displacement passing through the frame indeed will depend upon the area of cross-section A of the frame but it will also depend upon the orientation of the frame with respect to **D**. If the direction of **D** is normal to the area A, the electric displacement through the frame will be DA. On the other hand, if the direction of **D** is tangential to the area A, the electric displacement through the frame will be zero. In general we can see that the electric displacement through the frame is proportional to the projection of the area A perpendicular to **D**. So if the angle between **D** and the normal to the area $\hat{\mathbf{n}}$ is θ, the projected area will be $A\cos\theta$. The electric displacement through the frame therefore will be $DA\cos\theta$. In vector algebra this is the dot product of **D** and the area **A**, where the direction of an area is defined by its normal. Therefore,

$$\text{Electric displacement} = \mathbf{D} \cdot \mathbf{A}$$

Let us now consider a closed surface S surrounding the charges Q_1, Q_2, Q_3 as shown in Fig. 3.8. Then, through an incremental area **da** on the surface, the outward electric displacement will be $\mathbf{D} \cdot d\mathbf{a}$. The total electric displacement can be obtained by integrating over the surface S. Then according to the Gauss's law

$$\oint_S \mathbf{D} \cdot d\mathbf{a} = Q_1 + Q_2 + \dots + Q_n \tag{3.32}$$

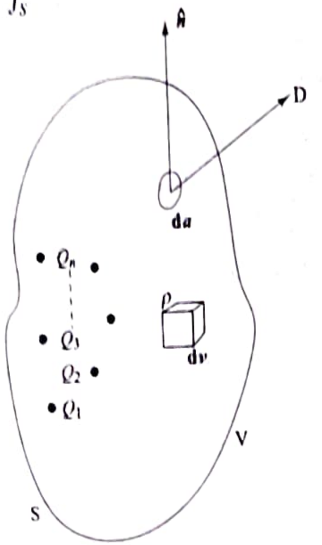


Fig. 3.8 *Region containing charges*

Instead of discrete charges $Q_1, Q_2, \dots Q_n$ if there is a continious distribution of charge inside the closed surface, we have to find total charge by integrating over the enclosed volume. The distributed charges can be correctly represented by a charge density ρ which in general is a function of space. The ρ is the charge per unit volume having units Coulomb/m^3. The charge in a small volume dv will be ρdv. The total charge enclosed by the volume can be obtained by integrating over the volume V. For distributed charges Eqn (3.32) then becomes

$$\oint_S \mathbf{D} \cdot d\mathbf{a} = \iiint_V \rho dv \tag{3.33}$$

This is the Gauss law in the integral form.

The equivalent differential form can be obtained by applying the Divergence theorem to Eqn (3.33). Using Eqn (C.28), the LHS of Eqn (3.33) can be written as

$$\oint_S \mathbf{D} \cdot d\mathbf{a} = \iiint_V (\nabla \cdot \mathbf{D})\, dv \tag{3.34}$$

Substituting in Eqn (3.33) and bringing all the terms on the left hand side of the equality sign we get

$$\iiint_V [(\nabla \cdot \mathbf{D}) - \rho]\, dv = 0 \tag{3.35}$$

Equation (3.35) has to be valid for any arbitrary volume. This can happen provided the integrand is identically zero making

$$\nabla \cdot \mathbf{D} = \rho \tag{3.36}$$

This is the differential form of the Gauss's law.

From Eqn (3.36) the physical meaning of divergence becomes more clear. The divergence of **D** represents the net outward flow of **D** per unit volume.

The Gauss's law can be effectively used in finding charge distribution in a region having electric field. One can find divergence of **D** to obtain charge density at that location. If there is no accumalation of charges in the given region the charge density ρ will be zero and the divergence of **D** will be identically zero. The total inward electric displacement for this region then will be equal to the total outward electric displacement.

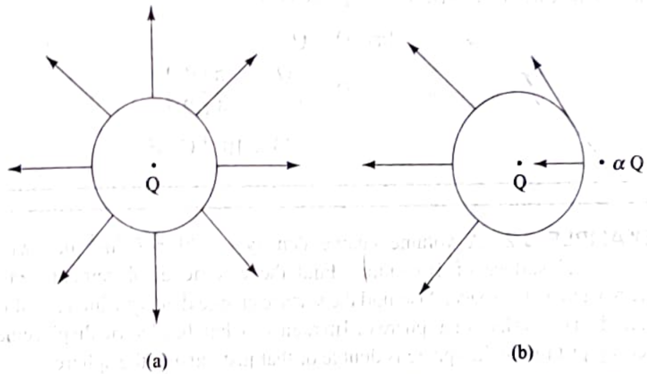


Fig. 3.9 *Electric displacement vectors for (a) a single charge Q (b) two charges Q and α Q*

One can also make an important observation here that, any charge which is outside the surface S, although affects the distribution of electric displacement on S, does not alter the net outward electric displacement. Consider a charge Q at the center of a spherical surface a shown in Fig. 3.9(a). The electric displacement is radially outward, and hence is normal to the spherical surface at every point. The total outward electric displacement is Q.

Now, suppose we place a charge α times Q (α is much greater than 1) outside the spherical surface. The direction and magnitude of the electric displacement on the surface of the sphere will be completely altered (Fig. 3.9(b)) but the net outward displacement will still be equal to Q because the charge enclosed by the surface is still Q only.

EXAMPLE 3.1 A volume charge density inside a hollow-sphere is $\rho = 10e^{-20r}$ C/m^3. Find the total charge enclosed within the sphere. Also find the electric flux density on the surface of the sphere.

Solution:

The total charge enclosed

$$Q = \int_V dv = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\int_{r=0}^{2} \rho\ r^2 \sin\theta\, dr\, d\theta\, d\phi$$

$$= \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta\, d\theta \int_0^{2} 10e^{-20r} r^2 dr$$

$$= 40\pi \int_0^{2} r^2 e^{-20r} dr = \frac{40\pi}{4000} = \frac{\pi}{100}$$

Since the distribution of charge is spherically symmetric, the electric flux density is also uniform on the surface of the sphere. By Gauss law the total electric flux from the surface of a sphere is Q.

$$\Rightarrow \quad 4\pi r^2 D = Q$$

$$\Rightarrow \quad D = \frac{Q}{4\pi r^2} = \frac{\pi/100}{4\pi(2)^2}$$

$$= 6.25 \times 10^{-4}\ \mathrm{C/m^2}$$

EXAMPLE 3.2 A volume charge density of 20 r $\mathrm{C/m^3}$ lies within a spherical surface of 1m radius. Find the electric displacement density everywhere in the space. Also find the surface charge density which should be placed on the surface of a sphere of 1m radius so that the electric displacement density just inside the sphere is double of that just outside the sphere.

Solution:

The charge enclosed within a sphere of radius $r < 1$ m is

$$Q(r) = \int_0^{2\pi}\int_0^{\pi}\int_0^{r} 20r \cdot r^2 \sin\theta\, d\theta\, d\phi\, dr$$

$$= 4\pi \int_0^{r} 20r^3 dr$$

$$= 20\pi r^4\ \mathrm{C}$$

The displacement density on the surface of sphere of radius r is

$$D = \frac{Q(r)}{4\pi r^2} = \frac{20\pi r^4}{4\pi r^2} = 5r^2\ \mathrm{C/m^2}$$

For $r > 1$ m, the charge is only within the radius of 1 m. Hence,

$$Q(r > 1) = Q(r = 1) = 20\pi\ \mathrm{C}$$

The D just within the surface of the sphere of 1m radius is

$$D = \frac{20\pi}{4\pi(1)^2} = 5\ \mathrm{C/m^2}$$

If the surface charge density of the sphere is ρ_s, the D just out side is $5 + \rho_s = 5/2$ (given.)

We, therefore, get $\rho_s = -2.5\ \mathrm{C/m^2}$.

EXAMPLE 3.3 The electric flux density is given as $\mathbf{D} = x^3\hat{\mathbf{x}} + x^2 y\hat{\mathbf{z}}$. Find the charge density inside a cube of side 2m placed centered at the origin with its sides along the coordinate axes.

Solution:

The volume charge density is (see appendix C for expression of divergence in the spherical coordinate system)

$$\rho = \nabla\cdot\mathbf{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}$$

$$= \frac{\partial}{\partial x}(x^3) + 0 + \frac{\partial}{\partial z}(x^2 y)$$

$$= 3x^2$$

The charge enclosed by the cube is

$$Q = \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} \rho\, dx\, dy\, dz = \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} 3x^2 dx\, dy\, dz$$

$$= 12\int_{-1}^{1} x^2 dx = 12.\frac{2}{3} = 8\ \mathrm{C}$$

EXAMPLE 3.4 The electric flux density is given by $\mathbf{D} = \frac{100\cos 2\theta}{r}\hat{\theta}$ $\mathrm{C/m^2}$. Find the charge enclosed within the region $1 < r < 2$, $0 < \theta < \pi/2$ rad.

Solution:

The volume charge density is

$$\rho = \nabla\cdot\mathbf{D} = \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(D_\theta \sin\theta)$$

since D_r and D_ϕ are $= 0$.

$$\rho = \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left\{\frac{100\cos 2\theta}{r}\sin\theta\right\}$$

$$= \frac{100}{r^2 \sin\theta} \frac{\partial}{\partial\theta} \left\{ \frac{\sin 3\theta - \sin\theta}{2} \right\}$$

$$= \frac{50}{r^2 \sin\theta} \{3\cos 3\theta - \cos\theta\}$$

The charge enclosed in the region is

$$Q = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi/2} \int_{r=1}^{2} \frac{50}{r^2 \sin\theta} \{3\cos 3\theta - \cos\theta\}\, r^2 \sin\theta\, d\theta\, d\phi\, dr$$

$$= 2\pi(50) \int_1^2 dr \int_0^{\pi/2} (3\cos 3\theta - \cos\theta)\, d\theta$$

$$= 100\pi [\sin 3\theta - \sin\theta]_0^{\pi/2}$$

$$= -200\pi \text{ C}$$

3.2.2 Gauss's Law for Magnetic Flux Density

As we write the Gauss's law for the electric displacement and the electric charges, we can identically write the law for the magnetic flux density and the magnetic charges. *The total magnetic flux coming out of a closed surface is equal to the total magnetic charge (poles) inside the surface.* However, there are no isolated magnetic monopoles. The magnetic poles are always found in pairs with opposite polarity. As a result, there are always equal number of north and south poles inside any closed surface making net magnetic charge identically zero inside a volume. The total outward magnetic flux from any closed surface must therefore be identically equal to zero. Writing mathematically,

$$\oint_S \mathbf{B} \cdot \mathbf{da} = 0 \tag{3.37}$$

where $\mathbf{B}$ is the magnetic flux density. Applying the Divergence theorem to Eqn (3.37) (as done in Eqn (3.34)) we get

$$\iiint_V (\nabla \cdot \mathbf{B})\, dv = 0 \tag{3.38}$$

or

$$\nabla \cdot \mathbf{B} = 0 \tag{3.39}$$

Equation (3.37) or Eqn (3.39) is essentially a mathematical statement of non-existence of the free magnetic monopoles. In future at some stage if one invents free magnetic charges the right hand side of Eqn (3.39) has to be modified by magnetic charge density.

3.2.3 Ampere's Circuital Law

The Ampere's circuital law states that *the total magnetomotive force along a closed loop is equal to the net current enclosed by the loop.* The magnetomotive force is the line integral of the tangential component of the magnetic field around the loop. Writing mathematically the Ampere's circuital law is

$$\oint \mathbf{H} \cdot \mathbf{dl} = I \tag{3.40}$$

Now consider a loop as shown in Fig. 3.10. The distribution of current passing through the loop may not be uniform. Also, the current may not be flowing perpendicular to the plane of the loop. It is, therefore, appropriate to define vector current density through the loop. The total current enclosed by the loop is

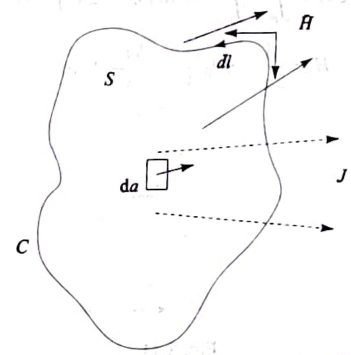


Fig. 3.10 *Schematic for Ampere's law. Thick arrows show the magnetic field and the dotted arrows show the current density.*

$$I = \iint_S \mathbf{J} \cdot \mathbf{da} \tag{3.41}$$

Now, applying Stokes theorem, Eqn (C.29) to Eqn (3.40), and substituting for I from Eqn (3.41), we get

$$\oint_C \mathbf{H} \cdot \mathbf{dl} = \iint_S (\nabla \times \mathbf{H}) \cdot \mathbf{da} = \iint_S \mathbf{J} \cdot \mathbf{da} \tag{3.42}$$

$$\Rightarrow \quad \iint_S \{\nabla \times \mathbf{H} - \mathbf{J}\} \cdot \mathbf{da} = 0 \tag{3.43}$$

Since Eqn (3.43) is valid for any arbitrary area, the integrand must be indentically zero. This gives

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{3.44}$$

This is the differential form of the Ampere's circuital law.

EXAMPLE 3.5 A long conducting cylinder of radius 'a', is placed along the z-axis. The cylinder carries a current density $J_0 r\hat{\mathbf{z}}$ A/m^2 for $r < a$. Find $\mathbf{H}$ inside and outside the conductor.

Solution:

By Ampere's circuital law

$$\oint \mathbf{H} \cdot \mathbf{dl} = \int_S \mathbf{J} \cdot \mathbf{da}$$

For current flowing in z-direction, the magnetic field is in the $\hat{\phi}$ direction. For $r < a$

$$H.2\pi r = \int_{\phi=0}^{2\pi} \int_{r=0}^{r} J_0 r dr d\phi = 2\pi \int_0^r J_0 r^2 dr$$

$$H = \frac{1}{2\pi r} 2\pi J_0 \frac{r^3}{3} = \frac{J_0 r^2}{3} \text{ A/m}$$

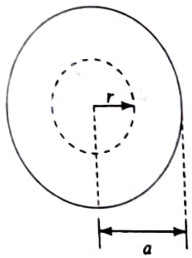


Fig. 3.11

For $r > a$, the enclosed current is only within $r < a$ Hence

$$2\pi Hr = 2\pi \int_0^a J_0 r^2 dr$$

$$\Rightarrow \quad H = \frac{J_0 a^3}{3r}$$

EXAMPLE 3.6 In a conducting medium the magnetic field is given as $\mathbf{H} = y^2 z\hat{\mathbf{x}} + 2(x+1)yz\hat{\mathbf{y}} - (x+1)z^2\hat{\mathbf{z}}$ A/m. Find the conduction current density at point $(2, 0, -1)$. Also find the current enclosed by the square loop $y = 1, 0 \le x \le 1, 0 \le z \le 1$.

Solution:

From Ampere's circuital law in differential form we get

$$\mathbf{J} = \nabla \times \mathbf{H} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y^2 z & 2(x+1)yz & -(x+1)z^2 \end{vmatrix}$$

$$\Rightarrow \quad \mathbf{J} = -2(x+1)y\hat{\mathbf{x}} + (y^2+z^2)\hat{\mathbf{y}} + (2yz - 2yz)\hat{\mathbf{z}}$$

$$= -2(x+1)y\hat{\mathbf{x}} + (y^2+z^2)\hat{\mathbf{y}}$$

$$\mathbf{J}\,|_{(2,0,-1)} = 1.0\hat{\mathbf{y}} \text{ A/m}^2$$

The square loop is in xz plane, i.e. perpendicular to the y axis. Current enclosed by the loop is

$$I = \int\int \mathbf{J} \cdot \mathbf{ds} = \int\int \mathbf{J} \cdot \hat{\mathbf{y}} dxdz$$

$$= \int_0^1 \int_0^1 J_y dxdz \text{ at } y = 1$$

$$= \int_0^1 \int_0^1 (1+z^2) dxdz = \frac{4}{3} \text{ A}$$

EXAMPLE 3.7 An imperfect conductor rod of circular cross section has radius of 0.01 m. The conductivity of the rod varies as $\sigma = 10^5(10^{-2} - r)$ ohm/m. If a 1 m length of the rod has 10 mV potential difference between its ends, find the magnetic field everywhere in the space.

Solution:

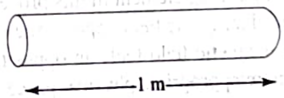


Fig. 3.12

The electric field inside the rod along its length $E = 10$ mV/1m $= 10^{-3}$ V/m.
The conduction current density is $J = \sigma E = 10^5(10^{-2} - r)$ A/m^2
Neglecting end effects, we get
Inside the conductor, i.e. for $r \le 1$ cm,

$$2\pi r H = 2\pi \int_0^r J r dr$$

$$H = \frac{1}{r} \int_0^r 10^5(10^{-2} - r) r dr$$

$$= \frac{1}{r}\left\{10^3 \frac{r^2}{2} - \frac{10^5 r^3}{3}\right\} = 500r - \frac{10^5}{3} r^2 \text{ A/m}$$

Outside the conductor i.e. for $r > 0.01$m, the total current enclosed is

$$I = 2\pi \left\{10^3 \frac{(10^{-2})^2}{2} - \frac{10^5 (10^{-2})^3}{3}\right\} = 0.105 \text{ A}$$

The magnetic field is

$$H = \frac{0.105}{2\pi r} = \frac{1.67 \times 10^{-2}}{r} \text{ A/m}$$

3.2.4 Faraday's Law of Electromagnetic Induction

The Biot Savart law tells us that the magnetic field is produced by a current. It therefore seems logical to find out whether the reverse is true, i.e. whether the magnetic fields would produce electricity. Faraday's experiments demonstrated that the static magnetic fields produce no electric current, but a time varying magnetic field produces an electromotive force in a closed loop which causes a current flow. According to Faraday's law, *the net electromotive force (EMF) in a closed loop is equal to the rate of change of magnetic flux (Φ) enclosed by the loop.*

Now, the current due to induced EMF will produce a magnetic field. Therefore there will be a magnetic field induced by the current which will be enclosed by the loop. There are two possibilities (i) the magnetic field due to the induced current is in same direction as the original field (ii) the magnetic field due to the induced current is in the opposite direction to the original field. In the first case, the two magnetic fields will add to enhance the net flux enclosed by the loop. This will increase the current and hence the magnetic flux enclosed. The process will be regenerative as there is no stabilizing element in this process. In the second case on the other hand, the induced magnetic field opposes the original magnetic field and therefore, the original magnetic field feels an opposition from the induced current. This process is a more appropriate physical process. We, therefore, find that the EMF in the loop is produced in such a way that the magnetic field due to the induced current is in opposite direction to that of the original field. This is called the Lenz's law.

Consider a loop in Fig. 3.13. The length segment **dl** and the area segment **da** are chosen by the right hand rule. Now, to satisfy the Lenz's law the current direction in the loop must be opposite to the direction of the length segment **dl**.

Mathematically the Faraday's law can be written as

$$V = \oint \mathbf{E} \cdot \mathbf{dl} = -\frac{\partial \Phi}{\partial t} \tag{3.45}$$

where V is the net EMF around the loop, and the negative sign is due to the Lenz's law.

Now if the loop has magnetic flux density **B**, the total flux enclosed by the loop is

$$\Phi = \iint_S \mathbf{B} \cdot \mathbf{da} \tag{3.46}$$

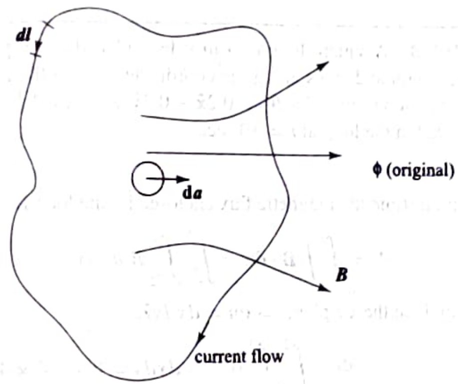


Fig. 3.13 *Schematic for Faraday's law of electromagnetic induction*

The Faraday's law Eqn (3.45) then becomes

$$\oint \mathbf{E} \cdot \mathbf{dl} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot \mathbf{da} \tag{3.47}$$

From Fig. 3.13 we can note that the magnetic flux enclosed by the loop can be varied in time in three ways

(i) keeping loop area stationary and varying the magnetic flux density with time,
(ii) keeping magnetic flux density static but changing the loop area,
(iii) changing the loop area in a time varying magnetic field.

Since, here we are interested in time varying fields we consider the first case only, that is, the area of the loop is stationary and the magnetic flux is time varying. In this case, we can take the time derivative in Eqn (3.48) inside the integral giving

$$\oint \mathbf{E} \cdot \mathbf{dl} = -\iint_S \frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{da} \tag{3.48}$$

Applying the Stoke's theorem (C.29) to Eqn (3.47) we get

$$\oint \mathbf{E} \cdot \mathbf{dl} = \iint_S (\nabla \times \mathbf{E}) \cdot \mathbf{da} = -\iint_S \frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{da} \tag{3.49}$$

$$\Rightarrow \qquad \iint_S \left\{ \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} \right\} \cdot \mathbf{da} = 0 \tag{3.50}$$

Again, using the argument that Eqn (3.50) has to be valid for any arbitrary area, we get

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{3.51}$$

This is the Faraday's law of electromagnetic induction in differential form.

EXAMPLE 3.8 A square loop of 4 m side is placed in xy-plane with its center at the origin and sides along the coordinate axes. If the magnetic flux density in the region is given by $B = (0.2\hat{\mathbf{x}} - 0.3\hat{\mathbf{y}} + 0.4\hat{\mathbf{z}})e^{-0.1t}$ Wb/m^2. Find the emf induced in the loop at $t = 10$ sec.

Solution:

At any instant of time the magnetic flux enclosed by the loop is

$$\Phi = \int\int \mathbf{B} \cdot \mathbf{da} = \int_{-2}^{2}\int_{-2}^{2} B_z dxdy$$

Since the loop is in the xy plane $\Rightarrow \mathbf{da} = dxdy\hat{\mathbf{z}}$.

$$\Phi = \int_{-2}^{2}\int_{-2}^{2} 0.4^{e^{-0.1t}} dxdy = 0.4 \times 4 \times 4 = 6.4e^{-0.1t}$$

$$\text{EMF} = -\frac{\partial \Phi}{\partial t} = 0.64\, e^{-0.1t}$$

At $t = 10$ sec, the EMF is

$$= -0.64\, e^{-0.1\times 10} = \frac{0.64}{e}\ \text{V}$$

EXAMPLE 3.9 Two charges Q_A and Q_B are separated by distance of 10 m. $Q_A = 10\cos\omega t$, and $Q_B = -10\cos\omega t$, where $\omega = 10^3$rad/sec. Find the magnetic flux density at a point which is at a distance of 10m from both the charges.

Solution:

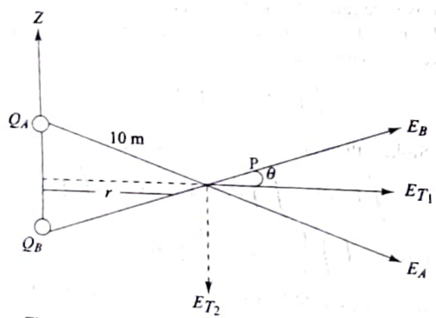


Fig. 3.14 *Electric field due to an electric dipole*

Let the observation point be P.

Total electric field at point P (refer Fig. 3.14) is

$$E_A = \frac{10\cos\omega t}{4\pi\epsilon_0 r}$$

$$E_B = -\frac{10\cos\omega t}{4\pi\epsilon_0 r}$$

$$E_{T1} = E_A\cos\theta + E_B\cos\theta = 0$$

$$E_{T2} = E_A\sin\theta - E_B\sin\theta = \frac{20\cos\omega t}{4\pi\epsilon_0 r}\sin\theta$$

Now, $\sin\theta = 5/10 = 1/2$.

$$\Rightarrow \quad E_{T2} = \frac{10\cos\omega t}{4\pi\epsilon_0 r}$$

Therefore, the vector electric field at point P is

$$\mathbf{E} = -\frac{10\cos\omega t}{4\pi\epsilon_0 r}\hat{\mathbf{z}}$$

Now, from Eqn (3.51) we have

$$\mathbf{H} = \frac{1}{j\omega\mu_0}(\nabla\times\mathbf{E}) = \frac{1}{j\omega\mu_0}\left\{\frac{\partial E_z}{\partial y}\hat{\mathbf{x}} - \frac{\partial E_z}{\partial x}\hat{\mathbf{y}}\right\}$$

Now, since $r = \sqrt{x^2+y^2}$, we get $\partial/\partial x(1/r) = -x/r^3$ and $\partial/\partial y(1/r) = -y/r^3$.

The magnetic field therefore is

$$\mathbf{H} = \frac{1}{j\omega\mu_0}\left(-\frac{10\cos\omega t}{4\pi\epsilon_0}\left\{-\frac{y}{r^3}\hat{\mathbf{x}} + \frac{x}{r^3}\hat{\mathbf{y}}\right\}\right)$$

At point P, $x = \sqrt{(10)^2 - (5)^2} = \sqrt{75}$ and $y = 0$, giving

$$\mathbf{H} = -\frac{1}{j\omega\mu_0}\frac{10\cos\omega t}{4\pi\epsilon_0}\left\{\frac{\sqrt{75}}{1000}\hat{\mathbf{y}}\right\}$$

$$= j\frac{\sqrt{3}}{80\pi}\frac{\cos\omega t}{\omega\epsilon_0\mu_0}\hat{\mathbf{y}}$$

3.3 MAXWELL'S EQUATIONS

Maxwell's equations are essentially a compilation of the mathematical equations derived from the basic laws of electromagnetics like the Gauss law, the Ampere's circuital law and the Faraday's law. While compiling the results however, Maxwell encountered a difficulty which was due to inconsistency in the Ampere's law.

Let us consider a closed surface (Fig. 3.15) having a volume charge density ρ. Let us assume that some charges are leaving the volume. As a result there is a rate of change of charge, that is, there is a current flow from the volume. If the current density on the surface of the volume is **J**, the net outward current from the volume is

$$\text{Net outward current} = \oint \mathbf{J} \cdot \mathbf{da} \tag{3.52}$$

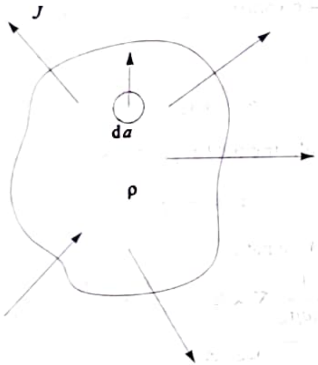


Fig. 3.15

At any instant if the volume change density is ρ, the total charge within the volume is $\oint_V \rho dv$ and the rate of decrease of charge in the volume is

$$-\frac{\partial}{\partial t}\oint_V \rho dv \tag{3.53}$$

Since the charges are to be conserved, the net outward current (rate of charge moving out of the volume) is equal to the rate of decrease of charge in the volume. We therefore get

$$\oint_S \mathbf{J} \cdot \mathbf{da} = -\frac{\partial}{\partial t}\oint_V \rho dv \tag{3.54}$$

Applying Divergene theorem (C.28) to Eqn (3.54) we get

$$\oint_S \mathbf{J} \cdot \mathbf{da} = \oint_V (\nabla \cdot \mathbf{J}) dv = -\frac{\partial}{\partial t}\oint_V \rho dv \tag{3.55}$$

$$\Rightarrow \qquad \oint_V \left\{\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t}\right\} dv = 0 \tag{3.56}$$

Since Eqn (3.56) again has to be valid for any arbitrary volume, the integrand should be identically zero, giving

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} \tag{3.57}$$

Equation (3.57) is called the 'continuity equation'.

The difficulty faced by Maxwell was precisely this that the Ampere's law is not consistent with the continuity equation. Let us see how?

The Ampere's law in differential form is given by Eqn (3.44). If we take divergence on both sides of Eqn (3.44), we get

$$\nabla \cdot (\nabla \times \mathbf{H}) = \nabla \cdot \mathbf{J} \tag{3.58}$$

The divergence of a curl of any vector ($\nabla \cdot (\nabla \times \mathbf{H})$) is identically zero. In other words, according to the Ampere's law, the divergence of **J** is identically zero, i.e.

$$\nabla \cdot \mathbf{J} \equiv 0 \tag{3.59}$$

This is inconsistent with the continuity Eqn (3.57) for a general time varying case. The correct modification to the Ampere's law was suggested by Maxwell.

Substituting for ρ from Eqn (3.36) in Eqn (3.57) we get

$$\nabla \cdot \mathbf{J} = -\frac{\partial}{\partial t}(\nabla \cdot \mathbf{D})$$

$$\Rightarrow \qquad \nabla \cdot \left(J + \frac{\partial \mathbf{D}}{\partial t}\right) = 0 \tag{3.60}$$

Equation (3.60) can be put in integral form by integrating over a volume and applying the divergence theorem, as

$$\oint_S \left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) \cdot \mathbf{da} = 0 \tag{3.61}$$

So, in Ampere's law if we regard $J + \frac{\partial D}{\partial t}$ as the total current enclosed by the loop, the law becomes consistent with the continuity equation. The term $\frac{\partial \mathbf{D}}{\partial t}$ obviously has dimension same as the current density **J** and is called the 'Displacement current density'. The displacement current is a notion of time varying electric field. For a static field, there is no displacemet current. This is the current which flows even in a medium, which does not have free charges.

At this point, one can distinguish between the current due to movement of free charges, **J**, and the curent due to time varying fields. The quantity **J** which is due to conduction of charges, is called the 'conduction current density', whereas, the rate of change of the electric displacement is called the 'displacement current density'.

In view of above discussion, the Ampere's law can be restated as:

The net magnetomotive force around a closed loop is equal to the total current, which is a combination of conduction and the dispalcement current.

Mathematically, therefore, Eqn (3.40) is written as

$$\oint \mathbf{H}\cdot d\mathbf{l} = -\int\int_S \left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right)\cdot d\mathbf{a} \tag{3.62}$$

and Eqn (3.44) becomes

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{3.63}$$

EXAMPLE 3.10 An air filled parallel plate capacitor is made of circular discs of area 2 m^2. The spacing between the discs is 0.1 m. If a voltage $20\cos 10^3 t$ volts is applied across the capacitor plates, find the displacement current density and the magnetic field between the capacitor plate. Also, show that the current flowing between the capacitor terminal is equal to the displacement current.

Solution:

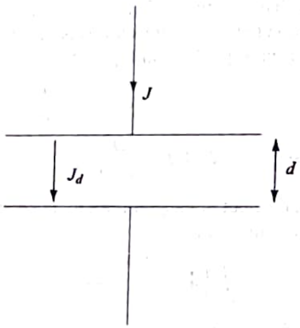


Fig. 3.16

The electric field inside the capacitor is

$$E = -\frac{V}{d} = -\frac{20\cos 10^3 t}{0.1} = -200\cos 10^3 t$$

The displacement current density is

$$J_d = \frac{\partial D}{\partial t} = \frac{\partial}{\partial t}(\epsilon_o E)$$
$$= 2\times 10^5 \epsilon_0 \sin 10^3 t \text{ A/m}^2$$

Since capacitor plate area, A, is 2 m^2, the total displacement current carried by the capacitor is

$$I_d = J_d A = 4\times 10^5 \epsilon_0 \sin 10^3 t \text{ A}$$

From Ampere's circuital law, (along a loop of radius r parallel to the capacitor plates)

$$2\pi r H = J_d \pi r^2$$

$$\Rightarrow \quad H = \frac{J_d r}{2} = 10^5 r\epsilon_0 \sin 10^3 t \text{ A/m}$$

Now the capacitor has a capacitance $C = \frac{\epsilon_0 A}{d} = \frac{\epsilon_0\times 2}{0.1} = 20\epsilon_0$.
Current in the capacitance terminal is

$$I = C\frac{dV}{dt} = 20\epsilon_0 \frac{d}{dt}(20\cos 10^3 t)$$
$$= -4\times 10^5 \epsilon_0 \sin 10^3 t$$

But, since the direction of the current in the external circuit is opposite, we get $I_d = I$.

EXAMPLE 3.11 In a medium having dielectric constant 9, the magnetic field is given as

$$H_\phi = \left(\frac{1}{r} + \frac{1}{r^2}\right)\sin\theta e^{-jr}\cos 10^8 t \text{ A/m}$$

Find the electric field and hence the displacement current density at $r = 10$ m and $\theta = \pi/4$.

Solution:

This is a problem in spherical coordinate system.
Since the conductivity $\sigma = 0$, we have $J = 0$, giving

$$\nabla \times \mathbf{H} = \mathbf{J_d} = \frac{\partial \mathbf{D}}{\partial t} = \frac{\partial}{\partial t}(\epsilon_0\epsilon_r \mathbf{E})$$

Since $H_r = 0$, $H_\theta = 0$

$$\mathbf{J_d} = \nabla \times \mathbf{H} = \frac{1}{r\sin\theta}\frac{\partial(H_\phi \sin\theta)}{\partial\theta}\hat{\mathbf{r}} - \frac{1}{r}\frac{\partial(rH_\phi)}{\partial r}\hat{\theta}$$
$$= \frac{1}{r\sin\theta}\left(\frac{1}{r} + \frac{1}{r^2}\right)2\sin\theta\cos\theta e^{-jr}\cos 10^8 t\hat{\mathbf{r}}$$
$$-\frac{1}{r}\left\{-j - \frac{j}{r} - \frac{1}{r^2}\right\}\sin\theta e^{-jr}\cos 10^8 t\hat{\theta}$$

The electric field is

$$\mathbf{E} = \int \frac{\mathbf{J_d}}{\epsilon_0\epsilon_r}dt$$

$$= \frac{1}{9\epsilon_0}\left[-\left(\frac{1}{r^2}+\frac{1}{r^3}\right)2\cos\theta e^{-jr}\frac{\sin 10^8 t}{10^8}\hat{r} - \left(\frac{j}{r}+\frac{j}{r^2}+\frac{1}{r^3}\right)\sin\theta e^{-jr}\frac{\sin 10^8 t}{10^8}\hat{\theta}\right]$$

At $r = 10$m and $\theta = \pi/2$, $1/r^2$ and $1/r^3$ terms are $<< 1/r$ term. Substituting $\epsilon_0 = \frac{1}{36\pi} \times 10^{-9}$ F/m, we get

$$\Rightarrow \quad \mathbf{E} \approx \frac{-j\sin\theta}{9\epsilon_0}\frac{e^{-jr}}{r}\frac{\sin 10^8 t}{10^8}\hat{\theta}$$

$$= -j\frac{\sin\theta}{9\epsilon_0}\frac{e^{-jr}}{r}\frac{\sin 10^8 t}{10^8}\hat{\theta}$$

$$= -j4\pi e^{-10j}\sin 10^8 t\hat{\theta} \text{ V/m}$$

After removing the inconsistency in the Ampere's law, the four equations which we get corresponding to the laws of electromagnetism, are called the Maxwell's equations. The Maxwell's equation can be written in integral as well as differential form as follows:

Differential form	Integral form	
$\nabla \cdot \mathbf{D} = \rho$	$\oint_S \mathbf{D}\cdot d\mathbf{a} = \iiint \rho dv$	(3.64)
$\nabla \cdot \mathbf{B} = 0$	$\oint_S \mathbf{B}\cdot d\mathbf{a} = 0$	(3.65)
$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$	$\oint_C \mathbf{E}\cdot d\mathbf{l} = -\iint \frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{a}$	(3.66)
$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$	$\oint_C \mathbf{H}\cdot d\mathbf{l} = \iint \left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right)\cdot d\mathbf{a}$	(3.67)

The differential form of Maxwell's equations are point relations i.e. they establish relationship between fields and sources at any point in space. The differential form involves derivatives of the fields with respect to space and time.

Since $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$, the derivative form is suitable only in those situations where medium and fields are continuous in space and the space derivatives exist. At media interfaces where the medium properties change abruptly, the differential form of Maxwell's equations can not be used. In this situation the integral form of Maxwell's equations is to be used. (Note that, the integral form can be used in any situation). When we apply the integral form of Maxwell's equations at the media interface, we get relationship between fields in the two media. These are generally referred to as the 'boundary conditions'. Before we discuss boundary conditions however, let us introduce the concept of surface charge and surface current.

3.4 SURFACE CHARGE AND SURFACE CURRENT

Surface charge and surface current truely exist on the surface of a conducting medium. In practice, we never find true surface charges or true surface currents. However, we will see later that in certain situations (like surface of a conductor) the charges and currents are confined to a very thin layer below the surface and therefore they can be approximated by surface charge and surface current respectively. Figure 3.17 shows the charges and current close to a surface within a distance d. Now, the charge inside the layer under a unit surface area is

$$\rho_s = \rho d$$

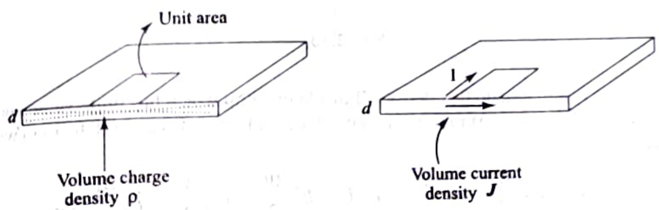


Fig. 3.17

In the limit when $d \to 0$ we get true surface charge density $\rho_s = \lim_{d\to 0} \rho d$. Of course for ρ_s to be finite, ρ should $\to \infty$. Since the dimensions of ρ are Coulomb/m^3, the dimensions of the surface charge are Coulomb/m^2.

Similarly, the current flowing under a strip of unit width of the surface is

$$J_s = J \cdot d$$

Again, in the limit when $d \to 0$ we get the surface current J_s. The direction of J_s is same as that of J. As argued in the case of the surface charge, the finite surface current density corresponds to infinite volume current density. Since the unit of current density J is A/m^2, the unit of J_s is A/m.

3.5 BOUNDARY CONDITIONS AT MEDIA INTERFACE

A media interface is a surface across which the media properties change abruptly. Since a medium is characterized by its ϵ, μ, σ, an abrupt change in any one or all of the parameters forms a media interface. Figure 3.18 shows media interface formed by two media having parameters $\epsilon_1, \mu_1, \sigma_1$ and $\epsilon_2, \mu_2, \sigma_2$, respectively.

Let us first consider the boundary condition due to the Faraday's law. Let us consider a small rectangular ($L \times W$) loop across the boundary as shown in the Fig. 3.18 and let the electric field along different sides of the loop be as shown in the Fig. 3.18. Suffix t represents components tangential to the boundary (interface), whereas sufix n represents components normal to the boundary. Let us now assume

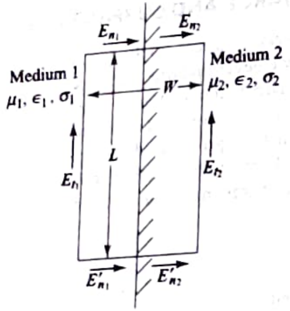


Fig. 3.18

that there is a magnetic flux density **B** inside the loop which has finite value and finite time derivative. Applying Faraday's law, i.e. Eqn (3.47) around the loop we get

$$E_{t_1}L + E_{n_1}\frac{W}{2} + E_{n_2}\frac{W}{2} - E_{t_2}L - E'_{n_2}\frac{W}{2} - E'_{n_1}\frac{W}{2} = \int \frac{\partial \mathbf{B}}{\partial t}\cdot \mathbf{da}$$

$$= \frac{\partial B_\perp}{\partial t}LW \quad (3.68)$$

where $B_\perp$ is the component of **B** perpendicular to the plane of the loop i.e. in the direction $\perp$ to the plane of the paper.

Now, let us reduce the area of the loop to zero making $W \to 0$. The loop then becomes a line along the boundary and Eqn (3.68) becomes

$$E_{t_1}L - E_{t_2}L = 0$$

$$\Rightarrow \quad E_{t_1} = E_{t_2} \quad (3.69)$$

Equation (3.69) gives the first boundary condition that 'the tangential component of electric field across a media interface is always continuous'.

Let us now consider the boundary condition corresponding to the Gauss's law. Now, consider a box of area $A \times h$ as shown in Fig. 3.19. Let us assume there is volume charge density ρ in the vicinity of the interface, and a surface charge density ρ_s at the interface.

In the limit when the height of the box goes to zero, the net electric displacement from the curved surface is zero since the area of the curved surface goes to zero. The net electric displacement from the box is then $(D_{n_1} - D_{n_2})A$. Then according to the Gauss law, this should be equal to total charge enclosed within the box, i.e.

$$(D_{n_1} - D_{n_2})A = \lim_{h\to 0} \rho Ah + \rho_s A \quad (3.70)$$

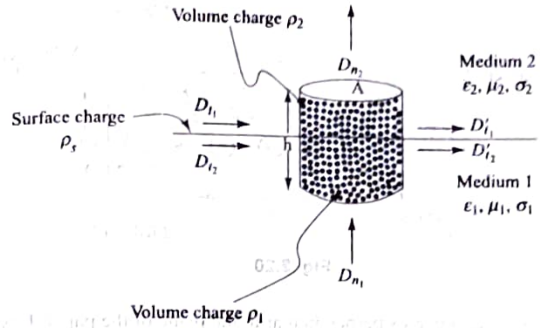


Fig. 3.19

$$\Rightarrow D_{n_1} - D_{n_2} = \rho_s \quad (3.71)$$

In absence of any surface charge, $\rho_s = 0$ and the boundary condition reduces to

$$D_{n_1} = D_{n_2} \quad (3.72)$$

The second boundary condition is then, 'the normal component of the electric displacement is continuous at the boundary in the absence of the surface charge'. In presence of the surface charge, the normal component of the electric dispalcement is discontinuous by an amount equal to the surface charge density.

Let us now apply the above treatment to the magnetic field and the magnetic flux density.

In Fig. 3.19 if we replace D by B we get relation identical to Eqn (3.71) with ρ_s replaced by magnetic surface charge density. Since there are no isolated magnetic charges, the magnetic surface charge density is identically zero, and consequently we always have

$$B_{n_1} = B_{n_2} \quad (3.73)$$

The third boundary condition, therefore is, 'the normal component of magnetic flux density is always continuous at the boundary'.

To obtain the boundary condition corresponding to the Ampere's law, let us consider a rectangular loop as shown in Fig. 3.20. Inside the loop there is volume current density J and surface current density J_s.

Applying Ampere's Law (Eqn (3.62)) to the loop we get in the limit when height of the loop $h \to 0$

$$H_{t_2}L - H_{t_1}L = J_sL + \lim_{h\to 0} J\cdot Lh + \lim_{h\to 0}\left(\frac{\partial D}{\partial t}\right)\cdot Lh \quad (3.74)$$

For finite $\dfrac{\partial \mathbf{D}}{\partial t}$ and **J** the limits tend to zero giving

$$H_{t_2} - H_{t_1} = J_s \quad (3.75)$$

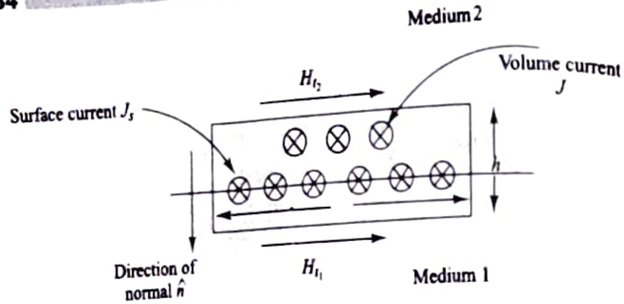


Fig. 3.20

Here, J_s is flowing inwards perpendicular to the plane of the paper. If we define the normal to the boundary $\hat{\mathbf{n}}$ pointing from medium 2 to medium 1, we can write Eqn (3.75) as

$$\hat{\mathbf{n}} \times (H_{t_2} - H_{t_1}) = \mathbf{J_s} \tag{3.76}$$

or in general

$$\hat{\mathbf{n}} \times (\mathbf{H_2} - \mathbf{H_1}) = \mathbf{J_s} \tag{3.77}$$

Since $\hat{\mathbf{n}} \times (H_{n_1} - H_{n_2}) \equiv 0$.

Note that the right hand convention has been used to write the cross product. If the direction of normal is changed from medium 1 to medium 2, the direction of the surface current will be opposite. The choice of the normal should be made correctly to get the correct direction of the surface current.

In the absence of any surface current, $J_s \equiv 0$ and the boundary condition Eqn (3.75) reduces to

$$H_{t_2} = H_{t_1} \tag{3.78}$$

The fourth boundary condition, therefore, is 'the tangential component of the magnetic field is continuous at the boundary if there is no surface current at the boundary'.

EXAMPLE 3.12 Just inside the surface of a dielectric slab the electric field is 10 V/m and makes an angle of 60° with the surface. If the dielectric constant of the slab is 4, find the electric field and its direction just above the surface.

Solution:

The tangential and normal components of the field in medium (1) (dielectric) are

$$E_{t_1} = 10 \cos 60^\circ = 5 \text{ V/m}$$

$$E_{n_1} = 10 \sin 60^\circ = \frac{5\sqrt{3}}{2} \text{ V/m}$$

From the boundary condition Eqns (3.69) and (3.72) we have

$$E_{t_2} = E_{t_1} = 5 \text{ V/m}$$

and

$$D_{n_1} = D_{n_2} \quad \Rightarrow \quad \epsilon_0 E_{n_2} = \epsilon_0 (4) E_{n_1}$$

$$\Rightarrow \quad E_{n_2} = 4 E_{n_1} = 10\sqrt{3} \text{ V/m}$$

Total electric field above the surface is

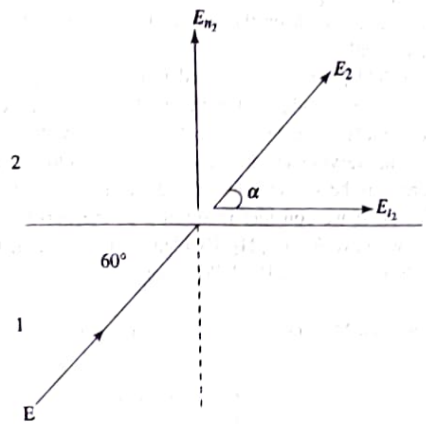


Fig. 3.21

$$\Rightarrow \quad E_2 = \sqrt{E_{t_2}^2 + E_{n_2}^2}$$

$$= \sqrt{325} = 18.0278 \text{ V/m}.$$

Angle which the field makes with the surface is

$$\alpha = \tan^{-1}\left(\frac{E_{n_2}}{E_{t_2}}\right) = \tan^{-1}\left(\frac{10\sqrt{3}}{5}\right) = 73.9^\circ$$

3.5.1 Boundary Conditions at the Surface of a Perfect Conductor

A perfect conductor is a medium which has infinite conductivity ($\sigma = \infty$). Of course, a perfect conductor is an abstraction as there is no media which has infinite conductivity. In practice however, properties of good conductors like silver, copper can be closely approximated by that of the perfect conductor.

The first thing we note about the perfect conductor is that the electric field inside it has to be zero. This is due to simple reason that any finite electric field will develop infinite conduction current density $\mathbf{J}$ $(J = \sigma E)$ since $\sigma = \infty$ for perfect conductor. So, for any finite current density J the electric field inside a conductor has to be identically zero. The infinite current density, however, is possible if it is confined to a slab of zero thickness (a current sheet), at the surface of a conductor, giving finite surface current. The true surface current therefore exists only at the surface of a perfect conductor. For a real life conductor, no matter how high its conductivity is, there is no surface current.

Since the electric field is zero inside a conductor, there can not be any charge inside a conductor and a charge has to move to the surface of the conductor. We, therefore, can have charges only on the surface of the conductor giving the so called the 'surface charge'.

For time varying fields, the magnetic field $\mathbf{H}$ is related to the time varying electric field. Since inside a conductor the electric field is zero, the time varying magnetic field also cannot exist inside a perfect conductor. We therefore conclude that there are no time varying electric and magnetic fields inside a perfect conductor and there may be surface charge and surface current at the surface of a perfect conductor. If we consider medium 2 as the perfect conductor, as shown in Fig 3.22, we have $\mathbf{E_2}, \mathbf{D_2}, \mathbf{H_2}, \mathbf{B_2}$ identically zero and the boundary conditions Eqns (3.69), (3.71), (3.73), (3.75) become

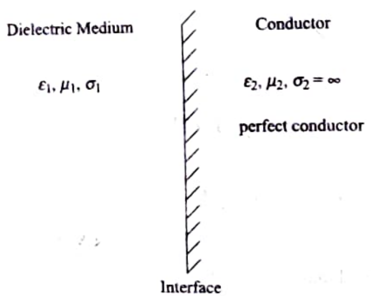


Fig. 3.22 *Dielectric-conductor Media interface*

$$E_{t_1} = 0 \tag{3.79}$$

$$D_{n_1} = \rho_s \tag{3.80}$$

$$B_{n_1} = 0 \tag{3.81}$$

$$-\hat{\mathbf{n}} \times \mathbf{H_1} = \mathbf{H_1} \times \hat{\mathbf{n}} = \mathbf{J_s} \tag{3.82}$$

We, therefore, make an important obervation and that is, at a perfect conducting surface, the electric field is always normal to the surface ($E_{t_1} = 0$), and the magnetic field is always tagential to the surface ($B_{n_1} = \mu H_{n_1} = 0$).

Table 3.1 summarizes the boundary conditions for source-free dielectric-dielectric interface and dielectric-conductor interface.

Table 3.1

Dielectric-dielectric interface	Dielectric-conductor interface
$E_{t_1} = E_{t_2}$	$E_{t_1} = 0$
$D_{n_1} = D_{n_2}$	$D_{n_1} = \rho$
$H_{t_1} = H_{t_2}$	$H_{t_1} = J_s$
$B_{n_1} = B_{n_2}$	$B_{n_1} = 0$

EXAMPLE 3.13 A medium has infinite conductivity for $z \le 0$, and $\epsilon_r = 5$, $\mu_r = 20$ and $\sigma = 0$ for $z > 0$. If the electric field for $z > 0$ is given by $\mathbf{E} = 10\cos(3 \times 10^8 t - 10x)\hat{\mathbf{z}}$. Find the surface charge density and surface current density at location (2, 3, 0) at $t = 0.5$ n sec.

Solution:

The electric field inside a conductor is 0 i.e. $\mathbf{E} = 0$ for $z < 0$. Also $\mathbf{E}$ for $z > 0$ is normal to the surface of the conductor. The surface charge density is therefore

$$\Rightarrow \qquad \rho_s = \epsilon_0 \epsilon_r E_n = \epsilon_0(5)10\cos(3 \times 10^8 t - 10x)$$

At $x = 2$ m and $t = 0.5 \times 10^{-9}$sec

$$\rho_s = 50\epsilon_0 \cos(3 \times 10^8 \times 0.5 \times 10^{-9} - 10 \times 2) = 2.387 \times 10^{-10} \text{ C/m}^2$$

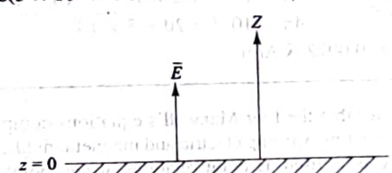


Fig. 3.23 *Electric field on a conducting surface*

From the Faraday's law

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\mu_0 \mu_r \frac{\partial \mathbf{H}}{\partial t}$$

$$\mathbf{H} = -\frac{1}{\mu_0 \mu_r} \int \nabla \times \mathbf{E} dt$$

Now,

$$\nabla \times \mathbf{E} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ 0 & 0 & E_z \end{vmatrix}$$

$$= \frac{\partial E_z}{\partial y}\hat{\mathbf{x}} - \frac{\partial E_z}{\partial x}\hat{\mathbf{y}}$$

Since E_z is not a function of y, $\partial E_z/\partial y \equiv 0$. We, therefore, get

$$\Rightarrow \quad \nabla \times \mathbf{E} = -\frac{\partial E_z}{\partial x}\hat{\mathbf{y}} = -100\sin(3 \times 10^8 t - 10x)\hat{\mathbf{y}}$$

$$\Rightarrow \quad \mathbf{H} = +\frac{100}{\mu_0\mu_r}\int \sin(3 \times 10^8 t - 10x)\hat{\mathbf{y}}\,dt$$

$$= -\frac{100}{\mu_0\mu_r}\frac{\cos(3 \times 10^8 t - 10x)}{3 \times 10^8}\hat{\mathbf{y}}$$

The surface current

$$\mathbf{J}_s = \hat{\mathbf{n}} \times \bar{H} = \hat{\mathbf{z}} \times \frac{-100}{\mu_0\mu_r}\frac{\cos(3 \times 10^8 t - 10x)}{3 \times 10^8}\hat{\mathbf{y}}$$

$$= \frac{100\cos(3 \times 10^8 t - 10x)}{3 \times 10^8 \mu_0\mu_r}\hat{\mathbf{x}}$$

At $x = 2$m and $t = 0.5 \times 10^{-9}$sec

$$\mathbf{J}_s = \frac{100\cos(3 \times 10^8 \times 0.5 \times 10^{-9} - 10 \times 2)}{4\pi \times 10^{-7} \times 20 \times 3 \times 10^8}$$

$$= 0.0072\ \hat{\mathbf{x}}\ \text{A/m}$$

We have seen earlier that the four Maxwell's equations completely describe the relationship between time varying electric and magnetic fields. The equations in differential form can be used as point relations at any point in space. The same equations in integral form provide the so called boundary conditions. The differential form of Maxwell's equation is suitable only for continuous media whereas the integral form can be applied even in case of media discontinuities. It should be remembered however that the Maxwell's equations are for time varying fields under the assumption that the space is time invariant, i.e. the shape, size, and other properties of a medium do not vary with time.

Substituting time derivatives identically to zero we get the Maxwell's equations for the static fields as

$$\nabla \cdot \mathbf{D} = \rho \qquad (3.83)$$

$$\nabla \cdot \mathbf{B} = 0 \qquad (3.84)$$

$$\nabla \times \mathbf{E} = 0 \qquad (3.85)$$

$$\nabla \times \mathbf{H} = \mathbf{J} \qquad (3.86)$$

We can make an important observation at this point and that is, the static electric fields are always conservative fields ($\nabla \times \mathbf{E} = 0$). The time varying fields on the other hand are always non conservative because $\mathbf{E}$ and $\mathbf{H}$ both coexist in a medium and their time derivatives are non zero.

3.6 SUMMARY

In this chapter, we stated the laws of electromagnetics and obtained their mathematical representation using vector algebra. Assuming that the space is time invariant, we derived the relationship between the general time-varying electric and magnetic fields. These relations are called the Maxwell's equations. Maxwell's equation govern in totality the phenomenon of electromagnetics. In differential form, the Maxwell's equations provide point relations beween the fields. Using the integral form of the Maxwell's equations, we obtained the boundary conditions at the media interface.

In the next and the subsequent chapters, we obtain solutions of the Maxwell's equations for various media configurations like the unbound medium, media interfaces, planar and rectangular waveguides, etc. First, we investigate the solution to the source-free Maxwell's equations (Propagation of electromagnetic waves), and finally we obtain the solution to the Maxwell's equations in the presence of electrical sources (Radiation of electromagnetic waves).

Review Questions

3.1 What is the Colomb's law?

3.2 How does the force between two electric charges vary as a function of distance?

3.3 What is electric field?

3.4 What is the unit of electric field?

3.5 What is the permittivity of the medium and what is its unit? What is the value of permittivity for vacuum?

3.6 How does the force between two charges vary as a function of permittivity? For which medium is the force maximum?

3.7 What is electric displacement density and what is its unit?

3.8 What is the relation between $\mathbf{D}$ and $\mathbf{E}$? Are $\mathbf{D}$ and $\mathbf{E}$ always oriented in the same direction?

3.9 What is an anisotropic medium?

3.10 What is electric scalar potential?

3.11 How are electric potential and electric field related?

3.12 What would be the direction of electric field between two points of the same potential?

3.13 In a region of uniform electric field, how does the electric potential vary as a function of distance?

3.14 What is Biot-Savart law?

3.15 Inside a circular loop carrying current what will be the direction of the magnetic field?

3.16 What is magnetic flux density and the magnetic filed? What is the relation between them?

3.17 What are units of **B**, **H**, and permeability?

3.18 What is Ohm's law?

3.19 In a conducting medium is it essential to have the conduction current in the direction of the electric field? Justify your answer with an example.

3.20 What is Gauss's law?

3.21 Write Gauss's law in integral and differential form.

3.22 Derive unit of **D** from Gauss's law.

3.23 What is the physical meaning of the equation $\nabla \cdot \mathbf{B} = 0$?

3.24 What is Ampere's circuital law?

3.25 Can you visualize the curl action in the Ampere's law from the Biot-Savart law?

3.26 Find the unit of **H** from the Ampere's law.

3.27 What is Faraday's law of electromagnetic induction?

3.28 Work out correct directions of the magnetic filed and the induced current in a loop.

3.29 What was the inconsistancy in the Ampere's law, and how was it resolved by Maxwell?

3.30 What is continuity equation? What is the origin of the continu .y equation?

3.31 What is displacement current and what is its unit?

3.32 Write th- Maxwell's equations in integral and differential forms.

3.33 What is dimention of the operator ∇?

3.34 What is the limitation of the differential form of the Maxwell's equations?

3.35 What is surface charge and surface current? What are their units?

3.36 Whai is boundary condition?

3.37 What are the boundary conditions on the electric field?

3.38 What are the boundary conditions on the magnetic field?

3.39 What are the boundary conditions on a conducting surface?

3.40 On a conducting surface since the tangential electric field is zero, explain flow of surface current.

Problems

3.1 A point charge Q is placed at a height, d above an infinitely large conducting sheet. What is the electric field and the surface charge density on the sheet?

3.2 The region between two concentric spheres of radii 1m and 2m respectively, is filled with a material with $\epsilon_r = 9$, and a volume charge density 10 C/m^3. Find the electric displacement in regions $r < 1$m, $1 \le r \le 2$m, and $r \ge 2$m.

3.3 The potential function in the free-space is given as

$$V = 5x^3y + 10x^2y^2z$$

Find the vector electric field and the volume charge density at point (4, 5, 2).

3.4 A spherical region of 5m radius has volume charge density $\rho = 10r^2$ and a surface charge density 5 C/m^2 on the surface of the sphere. Find the electric field everywhere in the space.

3.5 A volume charge density is placed inside a medium having conductivity σ and permittivity ϵ. Find the equation which governs the decay of charge density as a function of time. If 100 C/m^3 charge is placed inside copper, find the time over which the charge density will reduce to 1 C/m^3. For copper $\sigma = 5.6 \times 10^7$ ℧/m and $\epsilon = \epsilon_0$.

3.6 An infinitely long wire along the z-axis carries 10A current. Find the magnetic flux density at a distance of 5m from the wire.

3.7 A 10 m long wire is aligned with the z-axis and is symmetrically placed at the origin. Find the magnetic field at (i) point $(x, y, z) \equiv (1, 2, 5)$ (ii) point $(\rho, \phi, z) \equiv (2, \pi/3, 10)$ (iii) point $(r, \theta, \phi) \equiv (10, \pi/3, \pi/2)$.

3.8 A vector field is given as

$$\mathbf{F} = x^2y\hat{\mathbf{x}} + yz^2\hat{\mathbf{y}} + f(x, y, z)\hat{\mathbf{z}}$$

What should be the function $f(x, y, z)$ so that the vector **F** represents the magnetic flux density.

3.9 A circular loop of 10m diameter carries 2A current. Find the magnetic field strength at a distance of 20m along the axis of the loop. Also find the magnetic flux density in the plane of the loop as a function of distance from the center of the loop.

3.10 The magnetic field inside a cylindrical rod is given as

$$\mathbf{H} = 5(\rho - 100\rho^3)\hat{\boldsymbol{\phi}} \text{ A/m}$$

If the diameter of the rod is 10cm, find the current carried by the rod.

3.11 The magnetic flux density in a region is given as

$$\mathbf{B} = (x^4y^2\hat{\mathbf{x}} + 10x^2\hat{\mathbf{y}} - 4x^3y^2z\hat{\mathbf{z}})\cos(100\pi t) \quad \text{Wb/m}^2$$

A circular loop of radius 2m is placed parallel to the xy-plane at $z = 5$ m. Center of the loop lies on the z-axis. Find the emf induced in the loop.

3.12 Inside a material having $\epsilon_r = 5$ and $\sigma = 10^4$ ℧/m, the displacement current density is

$$J_d = 20\cos(10^7 t) \text{ A/m}^2$$

Find the conduction current density inside the material.

3.13 Two circular plates of 1cm diameter separated by a distance of 0.5mm from a capacitor. The region between the plates is filled with a dielectric having $\epsilon_r = 4$

and $\sigma = 0.1$ ℧/m. A 2GHz sinusoidal signal of amplitude 10V is applied between the capacitor plates. Find the rms value of the total current flowing between the capacitor terminal.

3.14 The electric field in a material with $\epsilon_r = 9\epsilon_0$ and $\mu = 25\mu_0$ is given by

$$\mathbf{E} = 10\cos(at - bz)\hat{\mathbf{x}} + 5\sin(at - bz)\hat{\mathbf{y}}$$

What should be the relation between a and b so that the fields are consistent with the source-free Maxwell's equations?

3.15 At air dielectric interface the field in the air has a magnitude 50V/m and makes an angle 30° with the normal to the interface. If the direction of the electric field vector is deflected by 20° towards the normal in the dielectric medium, find the dielectric constant of the medium.

3.16 A long wire carrying $10\cos(100t)$ A current is placed parallel to a conducting boundary at a distance of 5m. Find the surface charge and the surface current density on the conducting boundary.

3.17 A circular loop of 2m diameter carries a current $10\cos(10^6 t)$ A. A small loop of 0.1 m^2 area is placed parallel to the first loop at a distance of 50m along the axis of the loop. If the small loop has a resistance 5Ω, find the current induced in the loop.

3.18 The magnetic field in some region is given as

$$\mathbf{H} = \frac{10\sin\theta}{r^2} e^{j10^6 t}\hat{\boldsymbol{\phi}} \text{ A/m}$$

If the region has dielectric constant 3 and conductivity 0.1 ℧/m, find the electric field in the region. What are the conduction and displacement current densities in the medium.

3.19 A circular loop of 10cm radius is placed in the xy-plane. A magnetic flux density in the region is $\mathbf{B} = 5\cos(10^6 t)\hat{\mathbf{z}}$ Wb/m^2. If the resistance of the loop is 0.1Ω, find the magnitude and direction of the current in the loop at $t = 1\mu$sec.

3.20 Two circular discs of 20cm diameter form a parallel plane capacitor. The separation between the plates is 1cm. If a sinusoidal voltage $100\cos(10^4 t)$ V is applied between the capacitor plates, find the magnetic field between the capacitor plates. Neglect the fringing fields.

3.21 The $z = 0$ plane contains a current sheet. For $z < 0$, $\mu_{r1} = 5$ and $\epsilon_{r1} = 2$, and for $z > 0$, $\mu_{r2} = 20$ and $\epsilon_{r2} = 1$. If the magnetic fields in the two regions are

$$\mathbf{H}_1 = 6\hat{\mathbf{x}} + 5\hat{\mathbf{y}} + 8\hat{\mathbf{z}} \quad \text{A/m}$$

$$\mathbf{H}_2 = 2\hat{\mathbf{x}} - 3\hat{\mathbf{y}} + 2\hat{\mathbf{z}} \quad \text{A/m}$$

Find the linear current density in the current sheet.

CHAPTER 4

Uniform Plane Wave

In the previous chapter, we studied the development of Maxwell's equations in differential as well as integral forms. Maxwell's equations represent the electromagnetic phenomenon in its totality. Depending upon the medium properties and the distribution of the electric and magnetic sources, we get different solutions to Maxwell's equations. In this chapter, we will investigate the solution of Maxwell's equations in an unbound, source-free medium for time varying fields. For the time being, we defer the question as to how the electric and magnetic fields are generated? Assuming that the time varying electric and magnetic fields exist in an infinitely large medium, we ask a simple question, 'In what form do these fields exist?' We believe that nature always looks for the simplest possible solution to a problem. It means that, for a given problem, we do not unnecessarily look for a complicated solution. We choose the simplest possible solution which is consistent with the governing mathematical equations and the constraints. For example, if there are no constraints on the solution, we choose the uniform solution as it is the simplest form of a function. If the solution has to pass through two points, the simplest form is a linear function, and so on. The simplest solution to the Maxwell's equation in an unbound, homogenous medium for the time varying field is the 'uniform plane wave solution'. The uniform plane wave is, therefore, an important fundamental phenomenon of electromagnetics. Although we do not have either unbound media or ideal homogeneous media in practice, the concept of uniform plane wave is very useful as it provides the basic understanding of the electromagnetic wave propagation in a medium. Moreover, for media which have physical dimensions much larger than the wavelength, the solution closely approximates to the uniform plane wave solution. In practice, therefore, there are many problems where uniform plane wave solution is adequate and is quite useful.

In this chapter, we develop the understanding of uniform plane wave in cartesian geometry and investigate some interesting properties of a uniform plane wave.

4.1 HOMOGENEOUS UNBOUND MEDIUM

Let us consider a homogeneous, isotropic, unbound medium without any electric or magnetic source. For this case the medium permittivity ϵ and permeability μ are scalar quantities which are constants over the entire medium. The source free nature of the medium means that there are no free charges and currents in the medium making, $\rho = 0$ and $\mathbf{J} = 0$. The Maxwell's equations then reduce to

$$\nabla \cdot \mathbf{B} = 0 \tag{4.1}$$

$$\nabla \cdot \mathbf{D} = 0 \tag{4.2}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{4.3}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \tag{4.4}$$

From constitutive relations we have

$$\mathbf{B} = \mu \mathbf{H} \tag{4.5}$$

$$\mathbf{D} = \epsilon \mathbf{E} \tag{4.6}$$

Since the medium is homogeneous and non-time varying, μ and ϵ are constants as a function of space and time. The Maxwell's equations, therefore, reduce to

$$\nabla \cdot \mathbf{B} = \nabla \cdot (\mu \mathbf{H}) = \mu \nabla \cdot \mathbf{H} = 0 \tag{4.7}$$

$$\Rightarrow \quad \nabla \cdot \mathbf{H} = 0 \tag{4.8}$$

and

$$\nabla \cdot \mathbf{D} = \nabla \cdot (\epsilon \mathbf{E}) = \epsilon \nabla \cdot \mathbf{E} = 0 \tag{4.9}$$

$$\Rightarrow \quad \nabla \cdot \mathbf{E} = 0 \tag{4.10}$$

$$\nabla \times \mathbf{E} = -\frac{\partial (\mu \mathbf{H})}{\partial t} = -\mu \frac{\partial \mathbf{H}}{\partial t} \tag{4.11}$$

$$\nabla \times \mathbf{H} = \frac{\partial (\epsilon \mathbf{E})}{\partial t} = \epsilon \frac{\partial \mathbf{E}}{\partial t} \tag{4.12}$$

It can be seen from Eqns (4.11) and (4.12) that the time derivative of the magnetic field is related to the space derivative of the electric field and the time derivative of the electric field is related to the space derivative of the magnetic field. These two equations suggest that a time varying magnetic field can not exist without corresponding electric field and vice a versa. It is, therefore, clear that if the fields are time varying, both electric and magnetic fields have to co-exist and we can not get only electric field or only magnetic field which varies with time. Then this is conceptually different than the electrostatic or magnetostatic fields which are non-time varying and therefore can exist without each other.

Equations (4.11) and (4.12) are similar to the transmission line Eqns (2.5) and (2.6) for voltage and current respectively. The two equations are coupled in nature. These equations, however, are more general compared to that for the transmission lines as they represent the fields in three dimensional space whereas the transmission line is a one dimensional geometry.

On the lines similar to that used for the transmission line analysis, the two coupled equations for $\mathbf{E}$ and $\mathbf{H}$ are decoupled by differentiating the equations once more with respect to the space coordinate. Since, now, we are dealing with the three dimensional space, the space derivatives are the vector space derivative operators. Taking curl of Eqn (4.11) we get

$$\nabla \times \nabla \times \mathbf{E} = -\mu \nabla \times \frac{\partial \mathbf{H}}{\partial t} \tag{4.13}$$

$$\nabla \times \nabla \times \mathbf{H} = \epsilon \nabla \times \frac{\partial \mathbf{E}}{\partial t} \tag{4.14}$$

Interchanging space and time derivatives on the RHS, we get

$$\nabla \times \nabla \times \mathbf{E} = -\mu \frac{\partial (\nabla \times \mathbf{H})}{\partial t} \tag{4.15}$$

$$\nabla \times \nabla \times \mathbf{H} = \epsilon \frac{\partial (\nabla \times \mathbf{E})}{\partial t} \tag{4.16}$$

Substituting for $(\nabla \times \mathbf{H})$ in Eqn (4.15) from Eqn (4.12) and for $(\nabla \times \mathbf{E})$ in Eqn (4.16) from Eqn (4.11) we get

$$\nabla \times \nabla \times \mathbf{E} = -\mu \frac{\partial}{\partial t}\left(\epsilon \frac{\partial \mathbf{E}}{\partial t}\right) = -\mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{4.17}$$

$$\nabla \times \nabla \times \mathbf{H} = -\epsilon \frac{\partial}{\partial t}\left(\mu \frac{\partial \mathbf{H}}{\partial t}\right) = -\mu\epsilon \frac{\partial^2 \mathbf{H}}{\partial t^2} \tag{4.18}$$

Using the vector identity $\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$, where $\mathbf{A}$ is any arbitrary vector, the LHS of Eqns (4.17) and (4.18) can be written as

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{4.19}$$

$$\nabla(\nabla \cdot \mathbf{H}) - \nabla^2 \mathbf{H} = -\mu\epsilon \frac{\partial^2 \mathbf{H}}{\partial t^2} \tag{4.20}$$

Since from Eqns (4.8) and (4.10), $\nabla \cdot \mathbf{H} = 0$ and $\nabla \cdot \mathbf{E} = 0$, Eqns (4.19) and (4.20) reduce to

$$\nabla^2 \mathbf{E} = \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{4.21}$$

$$\nabla^2 \mathbf{H} = \mu\epsilon \frac{\partial^2 \mathbf{H}}{\partial t^2} \tag{4.22}$$

Equations (4.21) and (4.22) are the wave equations and their solutions represent the wave phenomenon in three dimensional space.

We, therefore, observe that if at all the time varying fields have to exist in a homogeneous, unbound medium, they have to exist in the form of a wave. Moreover, both electric and magnetic fields have to exist together. Hence this phenomenon is called 'Electromagnetic Wave'.

4.2 WAVE EQUATION FOR TIME HARMONIC FIELDS

For time varying fields the **E** and **H** vectors are functions of the space (x, y, z) and also functions of time t. If we assume that the fields are sinusoidal function of time (time harmonic fields) and if we take the time variation explicitly out, then, we can write a field vector **F** (where **F** could be either **E** or **H**) as

$$\mathbf{F}(x, y, z, t) = \mathbf{F}(x, y, z)e^{j\omega t} \tag{4.23}$$

Here ω is the angular frequency of the time varying field. Now onwards we will assume that a vector **F** is only a function of space (x, y, z) and its time variation $e^{j\omega t}$ is implicit.

For time harmonic fields we observe that

$$\frac{\partial}{\partial t}\mathbf{F}(x, y, z, t) = \frac{\partial}{\partial t}\{\mathbf{F}(x, y, z)e^{j\omega t}\} \tag{4.24}$$

$$= j\omega\mathbf{F}(x, y, z)e^{j\omega t} = j\omega\mathbf{F}(x, y, z, t) \tag{4.25}$$

Similarly,

$$\int \mathbf{F}(x, y, z, t)dt = \int \mathbf{F}(x, y, z)e^{j\omega t}dt \tag{4.26}$$

$$= \frac{1}{j\omega}\mathbf{F}(x, y, z)e^{j\omega t} + C \tag{4.27}$$

$$= \frac{\mathbf{F}(x, y, z, t)}{j\omega} + C \tag{4.28}$$

where C is a constant function of time. Since we are interested here in only time varying quantities, we take $C \equiv 0$. From Eqns (4.25) and (4.28) we see that a time derivative is equivalent to multiplication by $j\omega$ and a time integration is equivalent to division by $j\omega$. We, therefore, have

$$\frac{\partial}{\partial t} \equiv j\omega \tag{4.29}$$

$$\frac{\partial^2}{\partial t^2} \equiv j\omega \cdot j\omega = -\omega^2 \tag{4.30}$$

and so on.

Replacing $\partial^2/\partial t^2$ by $-\omega^2$ in Eqns (4.21) and (4.22) we get the wave equations for the time harmonic fields as

$$\nabla^2\mathbf{E} = -\omega^2\mu\epsilon\mathbf{E} \tag{4.31}$$

$$\nabla^2\mathbf{H} = -\omega^2\mu\epsilon\mathbf{H} \tag{4.32}$$

Each of the Eqns (4.31) and (4.32), in fact, is a set of three equations, one for each vector component of **E** and **H**. In other words, each component of **E** and **H** satisfies the wave equation. Since the ∇^2 operator is a scalar operator,

$$\left(\nabla^2 \equiv \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2}\right) \tag{4.33}$$

the wave equation for each component is a scalar equation as

$$\nabla^2\psi = -\omega^2\mu\epsilon\psi \tag{4.34}$$

where, ψ could be E_x, E_y, E_z or H_x, H_y, H_z.

4.3 SOLUTION OF THE WAVE EQUATION

As mentioned at the beginning of the chapter, our philosophy is to find the simplest possible solution which satisfies the given constraints, in the present case, the wave equation. Let us, therefore, ask a question 'Can we have an electric field which is time varying but is constant through out the space?' That is, can an electric field

$$\mathbf{E}(x, y, z, t) = \mathbf{K}e^{j\omega t} \tag{4.35}$$

be a solution of the wave equation, where **K** is a vector which has same orientation and magnitude throughout the space.

Substituting from Eqn (4.35) into Eqn (4.31) we get

$$\nabla^2\mathbf{K} = -\omega^2\mu\epsilon\mathbf{K} \tag{4.36}$$

Since **K** is constant over the space its space, derivatives are zero, therefore $\nabla^2\mathbf{K} = 0$. It implies that

$$\omega^2\mu\epsilon\mathbf{K} = 0 \tag{4.37}$$

Since μ and ϵ for a medium cannot be zero, either ω should be zero or **K** should be zero. If $\omega = 0$, the field **E** represents a DC field or an electrostatic field. Here since we are analysing time varying fields, $\omega \neq 0$ and the only option we have is that $\mathbf{K} \equiv 0$ making $\mathbf{E} \equiv 0$.

We, therefore, conclude that a uniform electric field in the three-dimensional space can not exist if it is time varying, and since the magnetic field is related to the electric field, even the magnetic field cannot exist. We also understand that no time varying field can exist without varying as a function of space.

Having seen that the simplest uniform solution for the wave equation does not exist, let us see whether the next level of complexity that is the uniformity of the field in a plane leads us anywhere! Without losing generality let us orient our coordinate system such that the electric field **E** aligns along the x-axis. Since now we are assuming the field to be constant in a plane, it could be constant in xy or xz or yz plane.

Let us first consider a case where the electric field is constant in a plane perpendicular to the direction of the field. Let us take an x-directed **E** field which is constant in the yz-plane (it means the field varies only as a function of x). The field can then be written as

$$\mathbf{E}(x, y, z, t) = E_0(x)e^{j\omega t}\,\hat{\mathbf{x}} \tag{4.38}$$

Let us substitute **E** in one of the Maxwell's equations, Eqn (4.11),

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} = -j\omega\mu \mathbf{H} \tag{4.39}$$

$$\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix} = -j\omega\mu \mathbf{H} \tag{4.40}$$

Since **E** is not a function of y and z, we have $\partial/\partial y \equiv 0$ and $\partial/\partial z \equiv 0$. Further **E** has only x-component so $E_y = E_z = 0$. Eqn (4.40) hence becomes

$$\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & 0 & 0 \\ E_0(x) & 0 & 0 \end{vmatrix} = -j\omega\mu \mathbf{H} \tag{4.41}$$

The determinant in Eqn (4.41) is identically zero giving

$$-j\omega\mu \mathbf{H} = 0 \tag{4.42}$$

Again, except when $\omega = 0$, the magnetic field **H** has to be zero. Since no time varying electric field can exist without corresponding magnetic field, $E_0(x)$ has to be identically zero. We hence have second conclusion that an electric or magnetic field which is time varying and is constant in a plane perpendicular to its direction, also cannot exist.

Let us now consider a field which is constant in a plane containing the field vector and see whether it has a non-zero solution. Without losing generality let us assume that the x-oriented electric field is constant in the xy-plane, i.e. it is a function of z only as given by

$$\mathbf{E}(x, y, z, t) = E_0(z)e^{j\omega t}\hat{\mathbf{x}} \tag{4.43}$$

For this case, $\partial/\partial x \equiv 0$ and $\partial/\partial y \equiv 0$. Substituting for **E** from Eqn (4.43) to Eqn (4.40) we get

$$\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ 0 & 0 & \frac{\partial}{\partial z} \\ E_0(z) & 0 & 0 \end{vmatrix} = -j\omega\mu \mathbf{H} \tag{4.44}$$

Simplifying Eqn (4.44) we get

$$\Rightarrow \qquad -\frac{\partial E_0(z)}{\partial z}\hat{\mathbf{y}} = -j\omega\mu\{H_x\hat{\mathbf{x}} + H_y\hat{\mathbf{y}} + H_z\hat{\mathbf{z}}\} \tag{4.45}$$

Equating different components on the two sides of Eqn (4.45) we get

$$H_x = 0 \tag{4.46}$$

$$H_y = \frac{1}{j\omega\mu}\frac{\partial E_0(z)}{\partial z} \tag{4.47}$$

$$H_z = 0 \tag{4.48}$$

Since $E_0(z)$ is not a constant function of z, $\partial E_0(z)/\partial z \neq 0$ and consequently H_y exists, E also exists.

One can note that in the above discussion, although we have taken x-oriented field constant in xy-plane, the arguments are valid for any plane which contains the field vector.

We can summarize the above discussion as follows:

1. A time varying electric or magnetic field which is uniform in the three-dimensional space, cannot exist.
2. A time varying field which is constant in a plane perpendicular to the field direction, also cannot exist.
3. The simplest form of field which can exist is the field which is constant in a plane containing the field vector, and consequently has spatial variation along the direction perpendicular to the constant field plane.

4.4 UNIFORM PLANE WAVE

As discussed above, the time varying fields which can exist in an unbound, homogeneous medium, are constant in a plane containing the field vector and have wave motion perpendicular to the plane. This phenomenon is then called the 'Uniform plane wave'.

Let us take an x-directed E-field which is constant in the xy-plane. The, field, therefore is given as

$$\mathbf{E}(x, y, z, t) = E_x(z)e^{j\omega t}\hat{\mathbf{x}} \tag{4.49}$$

Substitutinig Eqn (4.49) in the wave Eqn (4.21) and noting that

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial y} \equiv 0 \tag{4.50}$$

we get

$$\frac{d^2E_x(z)}{dz^2} = -\omega^2\mu\epsilon E_x(z) \equiv \gamma^2 E_x(z) \tag{4.51}$$

Note that since E_x is a function of z only, the partial derivative has been changed to full derivative.

Equation (4.51) is identical to the transmission line voltage equation with $\gamma \equiv \alpha + j\beta$ (is the propagation constant which has been defined in Chapter 2)

$$\gamma^2 = -\omega^2\mu\epsilon \tag{4.52}$$

$$\Rightarrow \qquad \gamma = j\omega\sqrt{\mu\epsilon} \equiv j\beta \tag{4.53}$$

The phase constant β for the medium therefore is $\omega\sqrt{\mu\epsilon}$ and the attenuation constant is zero.

On the lines similar to that for the transmission line, the solution to the Eqn (4.51) can be written as

$$E_x = E_x^+ e^{-j\beta z} + E_x^- e^{j\beta z} \tag{4.54}$$

The two terms on RHS represent the travelling waves moving in $+z$ and $-z$ directions respectively. Since in this case, γ is purely imaginary, (the attenuation constant for the medium is zero), the two waves travel with constant amplitude and their amplitudes are E_x^+ and E_x^- respectively anywhere in the space.

Substituting for E_x from Eqn (4.54) in the Maxwell's Eqn (4.11) we get

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{4.55}$$

$$\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ 0 & 0 & \frac{\partial}{\partial z} \\ E_x & 0 & 0 \end{vmatrix} = -j\omega\mu\mathbf{H} \tag{4.56}$$

$$\Rightarrow \quad \frac{\partial E_x}{\partial z}\hat{\mathbf{y}} = -j\omega\mu\mathbf{H} = -j\omega\mu\{H_x\hat{\mathbf{x}} + H_y\hat{\mathbf{y}} + H_z\hat{\mathbf{z}}\} \tag{4.57}$$

$$\Rightarrow \quad -j\beta E_x^+ e^{-j\beta z} + j\beta E_x^- e^{j\beta z} = -j\omega\mu H_y \tag{4.58}$$

$$\Rightarrow \quad H_y = \frac{\beta}{\omega\mu}E_x^+ e^{-j\beta z} - \frac{\beta}{\omega\mu}E_x^- e^{j\beta z} \tag{4.59}$$

From Eqn (4.59) we observe that the magnetic field has only y-component and it also has two travelling waves moving in $+z$ and $-z$ directions with respective amplitudes of $\beta E^+/\omega\mu$ and $-\beta E^-/\omega\mu$. It is then interesting to note that the ratio of the electric field and the magnetic field for a wave travelling in $+z$ direction is

$$\frac{E_x}{H_y} = \frac{E^+}{\frac{\beta}{\omega\mu}E^+} = \frac{\omega\mu}{\beta} \tag{4.60}$$

and that travelling in $-z$ direction is

$$\frac{E_x}{H_y} = \frac{E^-}{-\frac{\beta}{\omega\mu}E^-} = -\frac{\omega\mu}{\beta} \tag{4.61}$$

Since E_x has unit of Volts/m and H_y has unit of Amp/m, E_x/H_y has unit of V/A = impedance. The quantity $\omega\mu/\beta$ therefore has the unit of impedance, Ω. This impedance is called the *intrinsic impedance* of the medium, and is normally denoted by η.

Substituting for β from Eqn (4.53), we have

$$\eta = \frac{\omega\mu}{\beta} = \frac{\omega\mu}{\omega\sqrt{\mu\epsilon}} = \sqrt{\frac{\mu}{\epsilon}} \tag{4.62}$$

The intrinsic impedance is a property of the medium. The intrinsic impedance η for a medium serves the same purpose as the characteristic impedance Z_0 serves for a transmission line.

For the free-space (vacuum),

$$\mu = \mu_0 = 4\pi \times 10^{-7}\,\text{H/m}, \qquad \epsilon = \epsilon_0 \approx \frac{1}{36\pi} \times 10^{-9}\,\text{F/m}$$

The intrinsic impedance of the free-space, η_0 is

$$\eta_0 = \sqrt{\frac{4\pi \times 10^{-7}}{\frac{1}{36\pi} \times 10^{-9}}} \approx 120\pi \quad \Omega \tag{4.63}$$

For any other medium

$$\mu = \mu_0\mu_r \qquad \epsilon = \epsilon_0\epsilon_r$$

where μ_r and ϵ_r are the relative permeability and relative permittivity of the medium respectively, and the intrinsic impedance is

$$\eta = \sqrt{\frac{\mu}{\epsilon}} = \sqrt{\frac{\mu_0\mu_r}{\epsilon_0\epsilon_r}} = \eta_0\sqrt{\frac{\mu_r}{\epsilon_r}} \tag{4.64}$$

The above discussion suggests that a x-directed electric field E_x which is constant in xy-plane has a wave motion in z-direction (i.e. perpendicular to the plane in which it is constant), has an associated magnetic field which is oriented along y-direction and has a magnitude of E_x/η. If the wave is travelling in $+z$ direction, $H_y = E_x/\eta$ and if the wave is travelling in $-z$ direction, $H_y = -E_x/\eta$. We therefore reach a very important conclusion that the vectors **E**, **H** and the direction of the wave propagation are perpendicular to each other in space. The vectors **E**, **H** and the direction of wave propagation form the three coordinate axes in that order, and follow the right hand screw rule. That is, if the rotation from **E** to **H** is along the fingers of your right hand, the wave motion is along the direction of the thumb as shown in Fig. 4.1. Since both electric and magnetic field vectors are perpendicular to the direction of the wave propagation, the wave is called the 'Transverse Electromagnetic (TEM) Wave'.

In an unbound medium the electromagnetic wave is always transverse. Since light is an electromagnetic wave and in majority of the situations its wavelength is much smaller compared to the size of the medium, light behaves like a TEM wave in most cases.

In general, the electric field may not be oriented either along the x-axis or along the y-axis, as shown in Fig. 4.2.

The field **E** can be decomposed into two components E_x and E_y which for a wave travelling in $+z$ direction are related to H_x and H_y respectively as

$$H_y = \frac{E_x}{\eta} \tag{4.65}$$

$$H_x = -\frac{E_y}{\eta} \tag{4.66}$$

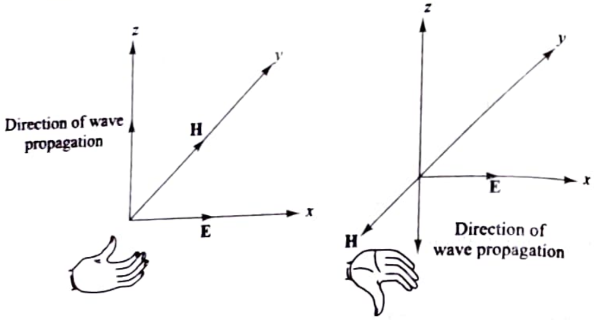


Fig. 4.1 *Directions of wave propagation and electric and magnetic field.*

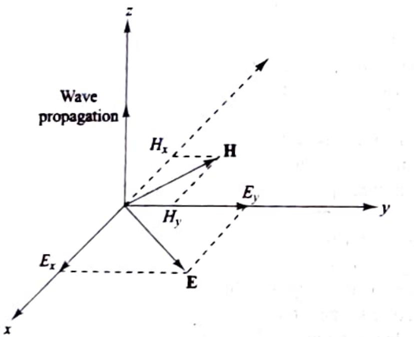


Fig. 4.2 *Arbitrarily oriented electric and magnetic fields with respect to the coordinate system.*

The total magnetic field **H** and the total electrical field **E** are related as

$$|\mathbf{H}| = \sqrt{H_x^2 + H_y^2} = \sqrt{\frac{E_x^2}{\eta^2} + \frac{E_y^2}{\eta^2}} = \frac{|\mathbf{E}|}{\eta} \tag{4.67}$$

$$\Rightarrow \qquad \frac{|\mathbf{E}|}{|\mathbf{H}|} = \eta \tag{4.68}$$

From Eqn (4.68) we see that the ratio of the electric and the magnetic field magnitudes is equal to the intrinsic impedance of the medium and therefore is purely decided by the medium parameters μ and ϵ ($\eta = \sqrt{\mu/\epsilon}$). This means that for a uniform plane wave, the magnitude of **H** is known if the magnitude of **E** and

the medium parameters are known. Moreover, the direction of **H** is known if the direction of **E** and the direction of wave propagation are known. Note that this relationship between **E** and **H** is valid at every instant of time and at every point in the space. In other words, if we know the vector electric field, the direction of wave propagation and the medium parameters, the magnetic field is uniquely defined. There is no need to keep a seperate account of the magnetic field. That is why, most of the wave propagation discussion is centered around **E** field and the behavior of the magnetic field is implicit.

EXAMPLE 4.1 A uniform plane wave at frquency of 300 MHz travels in vacuum along $+y$ direction. The electric field of the wave at some instant is given as $\mathbf{E} = 3\hat{\mathbf{x}} + 5\hat{\mathbf{z}}$. Find the phase constant of the wave and also the vector magnetic field.

Solution:

The phase constant

$$\beta = \omega\sqrt{\mu\epsilon} = \frac{2\pi \times 300 \times 10^6}{3 \times 10^8} = 2\pi \text{ rad/m}$$

The wave is travelling along $+y$ direction. Therefore, the **H** vector should be lie in xz- plane. Let the vector magnetic field be given by $\mathbf{H} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\hat{\mathbf{z}}$. Since for a uniform plane wave **E** and **H** are perpendicular to each other,

$$\mathbf{E} \cdot \mathbf{H} = 0$$

$$\Rightarrow \qquad 3A + 5B = 0 \qquad \ldots(1)$$

Also we have $\frac{|E|}{|H|} = \eta = 120\pi\ \Omega$

$$\Rightarrow |H| = \sqrt{A^2 + B^2} = \frac{|E|}{\eta} = \frac{\sqrt{9+25}}{120\pi} = 15.46 \times 10^{-3} \text{ A/m} \qquad \ldots(2)$$

Solving (1) and (2), we get

$$A = \pm\frac{5}{\sqrt{120\pi}} \quad \text{and} \quad B = \mp\frac{3}{\sqrt{120\pi}}$$

The vector magnetic field is

$$\mathbf{H} = \frac{5\hat{\mathbf{x}} - 3\hat{\mathbf{z}}}{\sqrt{120\pi}} \text{ A/m}$$

EXAMPLE 4.2 A uniform plane wave travelling in a medium having dielectric constant 9, has peak electric field of 10 V/m. The frequency of the wave is 1 GHz. Find the wavelength and peak magnetic field of the wave. If at some location ($z = 0$) and some instant ($t = 0$), the eletric field is 5 V/m, find the magnitudes of the electric and magnetic fields at $z = 2$ m and $t = 50$ msec. Assume the wave to be moving in $+z$ direction.

Solution:

The electric field of a wave travelling in $+z$ direction can be written as

$$E = Re\{10e^{j\omega t - j\beta z + j\phi}\} = 10\cos(\omega t - \beta z + \phi)$$

where $\omega = 2\pi f = 2\pi \times 10^9$ rad/s.

$$\text{and } \beta = \omega\sqrt{\mu\epsilon} = \omega\sqrt{\mu_0\epsilon_0\epsilon_r} = \omega\sqrt{\mu_0\epsilon_0} \times \sqrt{\epsilon_r}$$

The dielectric constant ϵ_r of the medium is 9.

$$\Rightarrow \qquad \beta = 2\pi \times 10^9\sqrt{4\pi \times 10^{-7} \times \frac{1}{36\pi} \times 10^{-9}}\sqrt{9} = 20\pi \text{ rad/m.}$$

Since, the electric field is not maximum at $z = 0, t = 0$, it should have a phase, say ϕ. We, therefore, have

$$5 = 10\cos(\phi)$$

$$\Rightarrow \qquad \phi = \cos^{-1}(\frac{1}{2}) = \frac{\pi}{3}$$

The expression for electric field is then

$$E = 10\cos(2\pi \times 10^9 t - 20\pi z + \frac{\pi}{3})$$

The intrinsic impedance of the medium is

$$\eta = \frac{\eta_0}{\sqrt{\epsilon_r}} = \frac{120\pi}{\sqrt{9}} = 40\pi\ \Omega$$

The magnetic field is given by

$$|H| = \frac{|E|}{\eta} = \frac{10}{40\pi}\cos(2\pi \times 10^9 t - 20\pi z + \frac{\pi}{3})$$

At $z = 2$ m and $t = 50$ msec $= 50 \times 10^{-3}$ sec, we have

$$E = 10\cos(2\pi \times 10^9 \times 50 \times 10^{-3} - 20\pi \times 2 + \frac{\pi}{3}) = -9.87 \text{ V/m}$$

$$H = \frac{E}{40\pi} = -0.0785 \text{ A/m}$$

4.5 WAVE POLARIZATION

It has been shown in the earlier sections that, a uniform plane wave is transverse in nature. The **E** and **H** field vectors are perpendicular to each other and they lie in a plane transverse to the direction of the wave propagation at all times and at all locations in space. If we look at this transverse nature of the wave carefully, we

note that although **E** and **H** vectors have to lie in a plane transverse to the wave motion, nowhere is it suggested that the direction of **E** or **H** should remain fixed as a function of time or space. The analysis shows that, at every point in space and at every instant of time, **E** and **H** should be perpendicular to each other and the ratio of their magnitudes should be equal to η. Therefore, if both **E** and **H** vectors rotate in the transverse plane by same angle and scale in same proportion, the transverse nature of the wave is not affected. We may then have various possibilities for **E** and **H** fields as shown in Fig. 4.3.

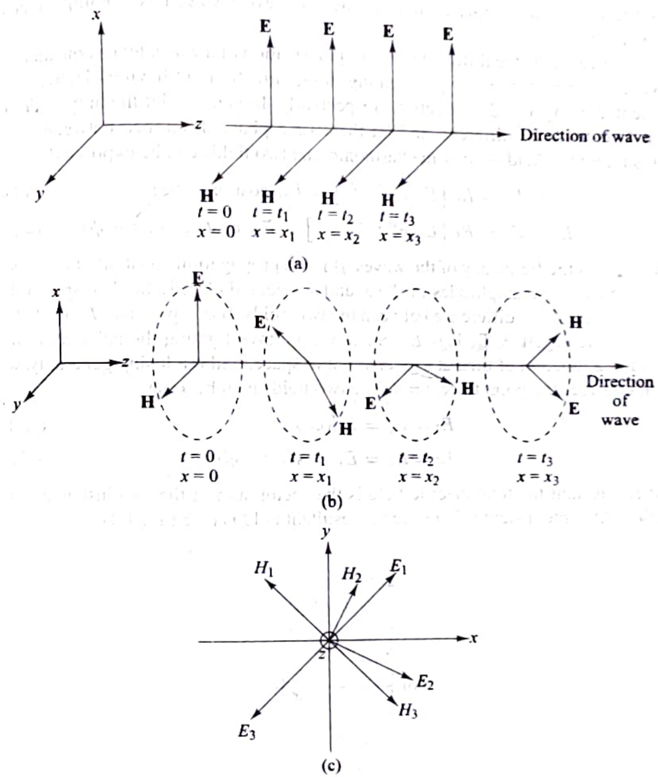


Fig. 4.3 *E and H at different time at some location in space (a) Direction of E remains same with time (b) Direction of E systematically changes with time (c) Direction of E randomly changes with time.*

Again, since the **H** field is unambiguosly specified with the knowledge of **E** and the direction of the wave propagation, we will discuss the behavior of the **E** field only. The temporal behavior of the electric field is described as the 'wave polarization'.

Wave polarization is defined by the temporal behavior of the electric field of a TEM wave at a given point in space. In other words, the state of polarization of a wave is described by the geometrical shape which the tip of the electric field vector draws as a function of time at a given point in space. Polarization is a fundamental characteristic of a wave, and every wave has a definite state of polarization.

To investigate the different states of polarization of a wave, let us consider two waves of same frequency propagating in $+z$ direction but having electric fields oriented along x and y directions respectively. In general, let the amplitudes of the two fields be different, and let there be a phase difference between them. Assuming the fields to be time harmonic, the two fields can be expressed as

$$E_x = E_1 = Re\left[E_{x0}e^{j\omega t - j\beta z}\right] = E_{x0}\cos(\omega t - \beta z) \tag{4.69}$$

$$E_y = E_2 = Re\left[E_{y0}e^{j\omega t + j\phi - j\beta z}\right] = E_{y0}\cos(\omega t - \beta z + \phi) \tag{4.70}$$

where ω is the frequency of the waves, β is the propagation constant of the wave, E_{x0} and E_{y0} are amplitudes of the x and y oriented electric fields respectively, and ϕ is the phase difference between the two fields. If ϕ is positive, E_y leads E_x, and if ϕ is negative, E_y lags E_x. Since we are investigating the behavior of the field as a function of time at a given point in space, without losing generality, we can consider the point to be $z = 0$. The two fields then become

$$E_1 = E_x = E_{x0}\cos\omega t \tag{4.71}$$

$$E_2 = E_y = E_{y0}\cos(\omega t + \phi) \tag{4.72}$$

At any instant the total electric field is the vector sum of the two instantaneous fields. At some instant t therefore the resultant field is (see Fig. 4.4)

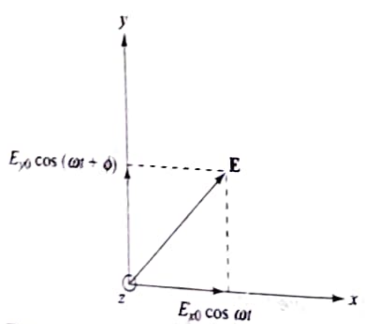


Fig. 4.4 *Total electrical field as vector sum of two components.*

$$\mathbf{E} = E_{x0}\cos\omega t\,\hat{\mathbf{x}} + E_{y0}\cos(\omega t + \phi)\hat{\mathbf{y}} \tag{4.73}$$

$$= \sqrt{E_{x0}^2\cos^2\omega t + E_{y0}^2\cos^2(\omega t + \phi)}\angle\tan^{-1}\left\{\frac{E_{y0}\cos(\omega t + \phi)}{E_{x0}\cos\omega t}\right\} \tag{4.74}$$

It is clear from Eqn (4.74) that the magnitude as well as the direction of the resultant electric field changes as a function of time. The locus of the tip of the E vector can be obtained by eliminating the time parameter from Eqns (4.71) and (4.72).

From Eqn (4.71) we have

$$\cos\omega t = \frac{E_x}{E_{x0}} \tag{4.75}$$

$$\Rightarrow \quad \sin\omega t = \sqrt{1 - \frac{E_x^2}{E_{x0}^2}} \tag{4.76}$$

Expanding Eqn (4.72) and substituting from Eqns (4.75) and (4.76) we get

$$\frac{E_y}{E_{y0}} = \cos\omega t\cos\phi - \sin\omega t\sin\phi \tag{4.77}$$

$$= \frac{E_x}{E_{x0}}\cos\phi - \sqrt{1 - \frac{E_x^2}{E_{x0}^2}}\sin\phi \tag{4.78}$$

Bringing first term of RHS to LHS and squaring both sides yields,

$$\left[\frac{E_x}{E_{x0}}\cos\phi - \frac{E_y}{E_{y0}}\right]^2 = \left[1 - \frac{E_x^2}{E_{x0}^2}\right]\sin^2\phi \tag{4.79}$$

Rearranging the equation and noting that $\sin^2\phi + \cos^2\phi = 1$, the equation of the locus of the tip of the field **E** is

$$\frac{E_x^2}{E_{x0}^2} - \frac{2E_xE_y\cos\phi}{E_{x0}E_{y0}} + \frac{E_y^2}{E_{y0}^2} = \sin^2\phi \tag{4.80}$$

Equation (4.80) is the equation of an ellipse. In general, therefore, the tip of the **E** vector, for a time harmonic plane wave, draws an ellipse as a function of time (see Fig. 4.5). Since, over one time period of the field, ωt changes by 2π, the tip of the field draws the ellipse every one time period. It means that if the frequency of the wave is f, the ellipse is drawn f times a second. This wave is then called the 'elliptically polarized wave' or we say that the polarization of the wave is 'elliptical'. Of course depending upon E_{x0}, E_{y0} and ϕ the equation of ellipse and consequently the shape and orientation of the ellipse change. Let us therefore investigate the behavior of the polarization ellipse as a function of E_{x0}, E_{y0} and ϕ.

It is clear from the above discussion that we are interested only in the shape of the ellipse and not its absolute size. The shape of the ellipse can be characterized by two parameters, namely the axial ratio (AR) i.e. the ratio of the major to the

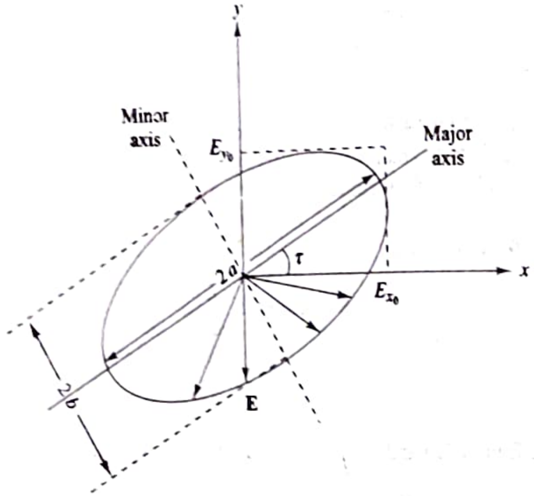


Fig. 4.5 *Polarization ellipse.*

minor axes of the ellipse, and the tilt angle (τ), i.e. the orientation of the major axis with respect to the x-direction. The same ellipse can also be characterized by the ratio of the two amplitudes, E_{y0}/E_{x0}, and the phase difference ϕ between them. Let us study how the ellipse changes with the ratio E_{y0}/E_{x0} and ϕ for few special cases.

Case I: Linear Polarization ($\phi = 0$) The two components E_x and E_y may or may not have same amplitude but let us assume that the phase difference between them is zero. For this case, Eqn (4.80) reduces to

$$\left[\frac{E_x}{E_{x0}} - \frac{E_y}{E_{y0}}\right]^2 = 0 \tag{4.81}$$

$$\Rightarrow \quad \frac{E_y}{E_{y0}} = \frac{E_x}{E_{x0}} \tag{4.82}$$

$$\Rightarrow \quad E_y = \left(\frac{E_{y0}}{E_{x0}}\right) E_x \tag{4.83}$$

This is the equation of a straight line with slope (E_{y0}/E_{x0}). The tip of electric field vector, therefore, draws a straight line if $\phi = 0$, irrespective of the amplitudes of the two field components. This polarization is hence called the 'Linear Polarization' or the wave is said to be linearly polarized.

Let us assume that at some instant of time t_o both components are zero (see Fig. 4.6). Then at a later time t_1, the two components will be $E_x = E_{x1}$ and $E_y = E_{y1}$ and the resultant field will be E_1 as shown in Fig. 4.6. Similarly, at time t_2 the field will be E_2 and so on. Exactly after quarter period, $E_x = E_{x0}$

and $E_y = E_{y0}$, and the electric field will reach to its maximum value E_{max}. In the next quarter period, the field will reduce to zero and in the following half period the field direction would reverse and would go through the same maximum to zero variation.

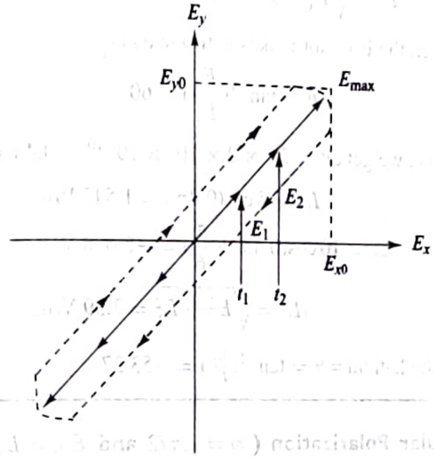


Fig. 4.6 *Linear polarization.*

Depending upon the ratio of E_{y0} and E_{x0} the slope of the line changes.

1. If $E_{x0} = 0$, the line becomes vertical and the wave is called 'vertically polarized wave', or the wave is said to have vertical polarization.
2. If $E_{y0} = 0$, the line becomes horizontal giving rise to a horizontally polarized wave, i.e. horizontal polarization.
3. If $E_{x0} = E_{y0}$, the wave is said to be linearly polarized with 45° polarization angle.

EXAMPLE 4.3 A uniform plane wave travelling in $+z$ direction has two components of the electric field $E_x = 5$ V/m and $E_y = 10\angle 30°$ V/m. At some instant say $t = 0$, the E_x component is maximum. Find the magnitude and direction of the field at $t = 0$ and at $t = 0.1$ nsec. Frequency of the wave is 2 GHz.

Solution:

The E_y component leads the E_x component by $30° = \frac{\pi}{6}$ rad.
The two components of the field can be written as

$$E_x = 5\cos(\omega t)$$

$$E_y = 10\cos(\omega t + \frac{\pi}{6})$$

(Note: Cosine function for E_x is chosen since at $t = 0$ the E_x field is maximum.)
At $t = 0$, we have $E_x = 5$ and $E_y = 10\cos\frac{\pi}{6} = \frac{10\sqrt{3}}{2}$
The field magnitude is

$$|E| = \sqrt{E_x^2 + E_y^2} = \sqrt{25 + 75} = 10 \text{ V/m}$$

Direction which the E vector makes with x-axis is

$$\theta = \tan^{-1}(\frac{E_y}{E_x}) = 60^\circ$$

At $t = 0.1$ nsec, we get $\omega t = 2\pi \times 2 \times 10^9 \times 10^{-10} = 0.4\pi$ rad

$$E_x = 5\cos(0.4\pi) = 1.545 \text{ V/m}$$

$$E_y = 10\cos(0.4\pi + \frac{\pi}{6}) = -2.078 \text{ V/m}$$

$$\Rightarrow \qquad |E| = \sqrt{E_x^2 + E_y^2} = 2.59 \text{ V/m}$$

Direction of the E-field = $\theta = \tan^{-1}(\frac{E_y}{E_x}) = -53.37^\circ$

Case II: Circular Polarization ($\phi = \pm\pi/2$ and $E_{x0} = E_{y0} = E_0$). The Eqn (4.80) now becomes

$$E_x^2 + E_y^2 = E_0^2 \tag{4.84}$$

Equation (4.84) is the equation of a circle. The tip of the electric field vector therefore draws a circle and the wave is said to be circularly polarized. We would note here that we get the same equation of circle (4.84) for $\phi = +\pi/2$ as well as for $\phi = -\pi/2$. It is worthwhile then to discover the effect the sign of the phase has on the polarization. Does it matter whether E_y leads or lags E_x? As far as the shape of curve is concerned, it does not matter whether ϕ is positive or negative, but it affects the way the curve is traced as a function of time. To understand this, let us consider Fig. 4.7. Let E_y be leading E_x by $\pi/2$. This means, at some instant t_0 when $E_x = 0$, $E_y = +E_{y0}$ and the resultant vector $E = E_{y0}$ is given by point A. At some small time later, E_y is reduced from E_{y0} and E_x is increased from zero in +ve direction. The resultant vector is then moved to point B making clockwise rotation. As the time increases, the vector further rotates clockwise and the tip of the vector draws a circle with clockwise movement.

Now consider the other case, i.e. E_y lags E_x by $\pi/2$. In this case, when $E_x = 0$, $E_y = -E_{y0}$ and the resultant vector is given by $E = -E_{y0}$, point A in Fig. 4.8. At a later time, E_y becomes less negative and E_x increases to some +ve value. The resultant vector then moves from A to B, making rotation of the vector anti-clockwise. In this case the circle is drawn by anti-clockwise movement of the tip of the vector.

The two cases of circular polarization corresponding to $\phi = \pm\pi/2$ are distinguished by their sense of rotation. In both cases the shape drawn by the tip of

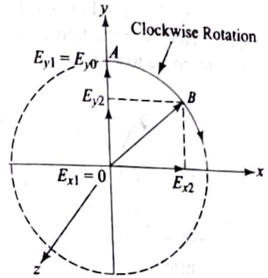


Fig. 4.7 *Clockwise rotation of the electric field vector as function of time.*

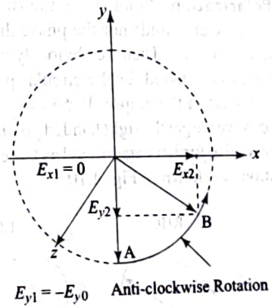


Fig. 4.8 *Anti-clockwise rotation of the electric field vector as a function of time.*

the E vector is a circle but for $\phi = +\pi/2$ the tip rotation is clockwise whereas it is anti-clockwise for $\phi = -\pi/2$. Of course, clockwise or anti-clockwise is arbitary and depends upon from which side of the paper (front or back) we look at the rotation. Now, if we assume that the wave is moving in +ve z direction, the wave motion is outward of the plane of the paper. We can then adopt a convention for the sense of rotation which will be uniformly followed by all. First, we turn ourselves in the direction of the wave motion (that is, we turn ourselves so that we see the wave receding from us). Now if the rotation of the vector is anti-clockwise (as in Fig. 4.8) or towards our right hand, we call the sense of polarization "Right Handed (RH)". If on the other hand, the rotation is clockwise (as in Fig. 4.7) i.e. towards our left hand we call the sense of polarization "Left handed (LH)". For a wave moving in $+z$ direction therefore $\phi = +\pi/2$ corresponds to LH and $\phi = -\pi/2$ corresponds to RH sense of polarization. Taking into account the sense of rotation, the circular polarization is further classified as Left circular(LC) and right circular (RC) polarization as shown in Fig. 4.9. In this figure, the wave is moving inward and therefore a clockwise rotation corresponds to RC wave and an anti-clockwise rotation corresponds to a LC wave.

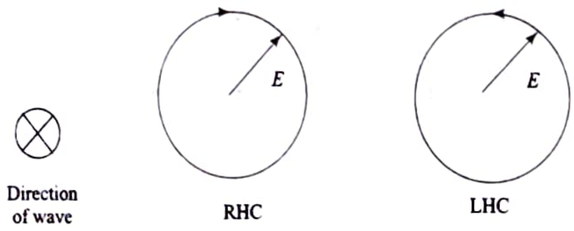


Fig. 4.9 *Two orthogonal circular polarizations.*

Case III: Elliptical Polarization In this case, the two field components E_x and E_y neither have the same amplitude nor the phase difference of 0 or $\pm\pi/2$ (i.e. $\phi \neq 0$ or $\pm\pi/2$ and $E_{x0} \neq E_{y0}$). Then we obviously get a general elliptically polarized wave. However, as discussed for the circular polarization, the sign of ϕ decides the sense of rotation of the ellipse. If ϕ is +ve we get the left handed polarization whereas if ϕ is -ve we get the right handed polarization. Consequently, the elliptical polarization is also further catagorized as left elliptical (LE) and right elliptical (RE) polarizations as shown in Fig. 4.10.

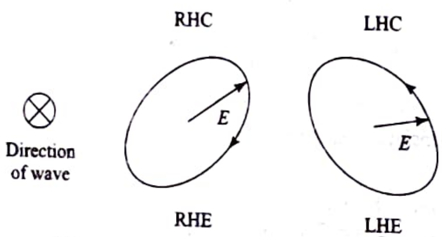


Fig. 4.10 *Two orthogonal elliptical polarizations.*

It is worth noting that the angle ϕ controls the bulging of the ellipse. For $\phi = 0$, the ellipse is compressed into a line and for $\phi = \pm\pi/2$ the ellipse is fully bulged to become a circle. The linear and circular polarizations are the special cases of elliptical polarization.

EXAMPLE 4.4 A uniform plane wave travels in $+x$ direction. The phasor electric field for the wave is $(3\hat{\mathbf{y}} + j5\hat{\mathbf{z}})$ V/m. Find the equation of the ellipse of polarization. Find the maximum magnitude of the field. Also find the sense of rotation.

Solution:

The wave travels in $+x$ direction and therefore the E-vector should lie in the yz plane. The two field components in y and z directions respectively can be writtten as

$$E_y = 3\cos\omega t$$

$$E_z = 5\cos(\omega t + \frac{\pi}{2})$$

(Note: j in z component indicates a phase difference of $+\frac{\pi}{2}$ between E_z and E_y). The equation of the ellipse is

$$\frac{E_y^2}{9} + \frac{2 \times E_y \times E_z \cos(\frac{\pi}{2})}{15} + \frac{E_z^2}{25} = \sin^2(\frac{\pi}{2})$$

$$\Rightarrow \frac{E_y^2}{9} + \frac{E_z^2}{25} = 1$$

The ellipse of polarization is shown in Fig. 4.11.
The maximum field magnitude = 5 V/m.
To find the sense of rotation, let us find the resultant field direction at say $t = 0$ and $t = \Delta t$ (Positive).
At $t = 0$, $E_y = 3$ and $Ez = 5\cos\frac{\pi}{2} = 0$. So, the field vector is vertically upwards.
At $t = \Delta t$, $E_y = 3\cos\omega\Delta t$ = +ve
and $E_z = 5\cos(\omega\Delta t + \frac{\pi}{2})$ =-ve
The resultant vector, therefore, moves leftward. This wave is travelling in +x-direction (Inside the paper) and hence the sense of rotation is Left Handed.

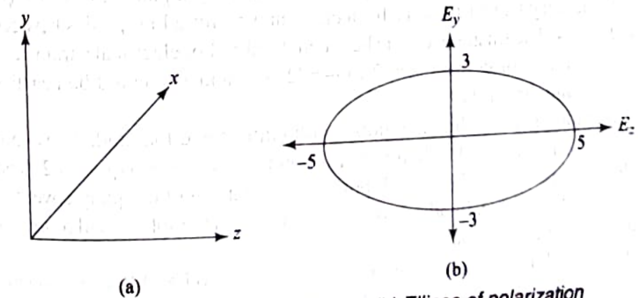


Fig. 4.11 *(a) Coordinate axes (b) Ellipse of polarization*

4.5.1 Representation of Polarization States

The state of polarization of a wave is represented by the shape, orientation and the sense of rotation of the polarization ellipse. The shape of the ellipse is characterized by the axial ratio (AR) and the orientation of the ellipse is measured by the tilt angle (τ), the angle which the major axis makes with the x-direction. By definition, AR is a positive quantity. However, if we assign a sign to AR, and adopt a convention that +ve AR means left handed rotation and $-$ve AR means right handed rotation, a pair (AR, τ) can unambigiously define the state of polarization including the sense of rotation. This representation is called wave parameter representation of the polarization state. The polarization state can also be defined in terms of the electrical quantities i.e. ratio (E_{y0}/E_{x0}) and ϕ.

Since there is a one-to-one correspondance between the ellipse parameters (AR, τ) and the electrical parameters (E_{y0}/E_{x0}) and ϕ, either of the representations is acceptable. However, one would need relations between the parameters of the two representations. For example, we may know the amplitudes and phases of two electric fields E_x and E_y and would like to know what polarization ellipse would be generated by these fields, or alternatively we may wish to generate some polarization ellipse and would like to know what E_x and E_y would be required to produce that ellipse.

The correspondence between two representations and a compact graphical representation of a state of polarization is due to Poincare'. It is shown that all states of polarization are compressed over the surface of a sphere called the Poincare' sphere. Let us define the following angles

$$\epsilon = \cot^{-1}(\pm AR) \tag{4.85}$$

$$\gamma = \tan^{-1}(E_{y0}/E_{x0}) \tag{4.86}$$

Then a state of polarization can be represented either by a pair (ϵ, τ) or (γ, ϕ). Since, AR can vary from 1 to ∞ (1 for a circle and ∞ for a line), ϵ lies between $-\pi/4$ and $+\pi/4$. The tilt angle can lie between 0 and π. For electrical parameters, if we take the range of ϕ from $-\pi/2$ to $+\pi/2$, E_{x0} and E_{y0} could be negative and γ would range from 0 to π.

Now, consider a sphere like the globe, with latitude and longitude lines on it (see Fig. 4.12). If we take the latitude of a point on the globe equal to 2ϵ and longitude of the point equal to 2τ, all polarization states will uniquely cover the total surface on the sphere. In other words, every point on the sphere will represent a unique state of polarization.

For the points lying on the equator of the sphere, $\epsilon = 0$ i.e. $AR = \infty$, and the state of polarization is therefore linear. For zero longitude point, $\tau = 0$ and hence the linear polarization is horizontal. As the longitude increases the polarization angle increases and at 180° longitude the polarization becomes vertical as τ becomes 90°.

The north and the south poles correspond to $\epsilon = \pm\pi/4$ i.e. $AR = \pm 1$ respectively. With the sign convention defined for AR, the north pole represents

a LC polarization, whereas the south pole represents a RC polarization. All other points which lie between the equator and the poles represent elliptical polarizations in general. However, since ϵ and hence AR is +ve in the northern hemisphere, the points in the northern hemisphere represent the 'Left handed' polarization. Similarly, the points in the southern hemisphere have AR $-$ve and hence represent the 'Right handed' polarization.

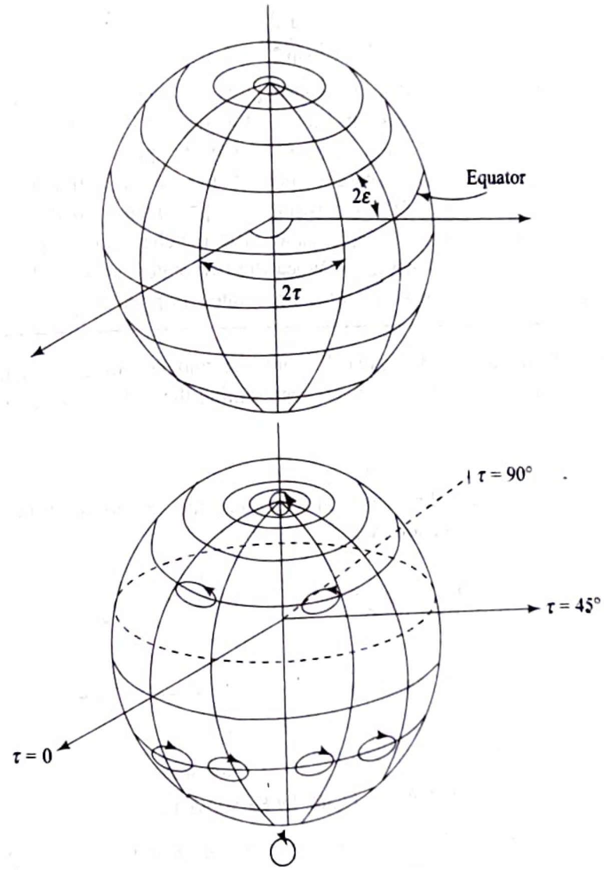


Fig. 4.12 *Poincare' shpere.*

Figure 4.12 shows all four angles ϵ, τ, γ and ϕ marked on the Poincare' sphere. Using spherical trigonometry we can establish the relation between the two pairs (ϵ, τ) and (γ, ϕ). The relations can be used to convert the electrical parameters to the ellipse parameters and vice-versa.

Conversion from electrical parameter to ellipse parameter:

$$\tan 2\tau = \tan 2\gamma \cos\phi \tag{4.87}$$

$$\sin 2\epsilon = \sin 2\gamma \sin\phi \tag{4.88}$$

Conversion from ellipse parameters to electrical parameters:

$$\cos 2\gamma = \cos 2\epsilon \cos 2\tau \tag{4.89}$$

$$\tan\phi = \frac{\tan 2\epsilon}{\sin 2\tau} \tag{4.90}$$

Apart from the fact that the Poincare' sphere provides a very compact and elegant representation of the states of polarization, its practical use is to find how close the two polarization states are, or how much interaction takes place between two states of polarization. The closeliness of the two polarization states is an important aspect of a communication system. When the two states of polarization are the same, they are called the 'matched states' of polarization. Two matched states of polarization have full interaction between them. When the two states are completely non-interacting they are known as the 'orthogonal states' of polarization.

EXAMPLE 4.5 For the wave in Example 4.4, find the latitude and longitude of the point on the Poincare' sphere representing the state of polarization of the wave.

Solution:

Reorienting the coordinate system such that the direction of the wave propagation is $+z$ direction we get

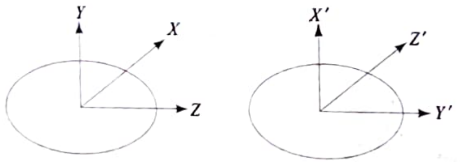


Fig. 4.13 *Figure for Example 4.5.*

The x' component corresponds to the y component and the y' component corresponds to the z component, as shown in Fig. 4.13.
The equation of ellipse is therefore

$$\frac{E_{x'}^2}{9} + \frac{E_{y'}^2}{25} = 1$$

The sense of rotation is LH. The axial ratio, is AR = +5/3
The major axis of the ellipse is along the y' axis giving tilt angle $\tau = \frac{\pi}{2}$

Also we get, $\epsilon = \cot^{-1}(+\frac{5}{3}) = 0.54$ rad
The point on the Poincare' sphere has longitude and latitude $(2\tau, 2\epsilon) = (\pi, 1.08)$.

EXAMPLE 4.6 A LHE polarized wave is to be generated using x and y polarized waves. The tilt angle and axial ratio of the ellipse of polarization is 30° and 3 respectively. Find the amplitude and phase of the x and y polarized waves. Assume the wave to be propagating in the $+z$-direction.

Solution:

For the polarization ellipse we have $\tau = 30^\circ = \frac{\pi}{6}$ and $\epsilon = \cot^{-1}(+3) = 18.4^\circ$
Using Eqns (4.89) and (4.90) we obtain

$$\cos 2\gamma = \cos 2\epsilon \cos 2\tau = \cos(36.8^\circ)\cos(60^\circ)$$

$$\Rightarrow \gamma = 33.2^\circ$$

and
$$\tan\phi = \frac{\tan 2\epsilon}{\sin 2\tau} = \frac{\tan(36.8^\circ)}{\sin(60^\circ)} = 0.8658$$

$$\Rightarrow \quad \phi = 40.88^\circ$$

If we assume that the wave is propagating in the +ve z direction, and the amplitudes of the x and y polarized waves are E_1 and E_2 respectively, we get

$$\frac{E_2}{E_1} = \tan(\gamma) = 0.65$$

and the phase difference between E_y and E_x is 40.88°

As will be seen in the later sections, every antenna system has a state of polarization, i.e. it maximally responds to a particular state of polarization. If now an electromagnetic wave with different state of polarization falls on the antenna, the power transfer efficiency from the wave to the antenna depends upon how close their states of polarization are. Closer the states of polarization higher the power transfer efficiency. If the two states are completely non-interacting (orthogonal), no power transfer takes place between the wave and the antenna. The antenna and the wave become transparent to each other.

Two orthogonal states of polarization can be used to double the signal transmission capacity of a communication channel. Since the orthogonal states of polarization have no transfer of power between them, one may transmit two waves of same frequency but orthogonal polarizations simultaneously and can increase the information transmission capacity by a factor of two.

The closeliness of the two states of polarization is directly related to the distance between the corresponding points on the Poincare' sphere. If the two states are denoted by points M and M' on the Poincare' sphere respectively (see Fig. 4.14),

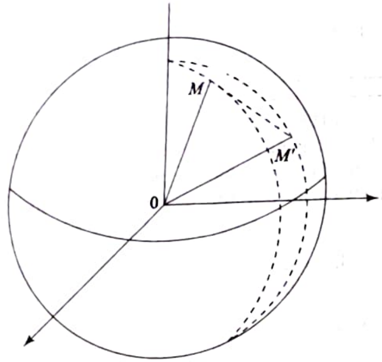


Fig. 4.14 *Two different states of polarization marked on the Poincare' sphere.*

the power transfer efficiency between the two states is

$$\eta_{pol} = \cos^2\left\{\frac{\angle MOM'}{2}\right\} \tag{4.91}$$

where $\angle MOM'$ is the angle subtended by the arc MM' on the center, 'O' of the Poincare' sphere. Naturally when M and M' are same, $\angle MOM' = 0$ and the efficiency is 1. When $\angle MOM' = 180°$, i.e., when M and M' are radially opposite points on the sphere, the efficiency is zero. In other words, any two polarization states represented by radially opposite points on the Poincare' sphere, are orthogonal states. This is a very useful feature of the Poincare sphere. By this, we can see that the RC and the LC polarizations, which correspond to south and north poles of the sphere respectively, are orthogonal states. Similarly, horizontal and vertical polarizations are orthogonal since both correspond to equitorial points but with longitudes 0° and 180°. It is also interesting to note that the power transfer efficiency between a circular and a linear polarization is always 1/2 as the circular polarizations correspond to the poles and the linear polarizations lie on the equator of the Poincare' sphere.

It is clear from the above discussion that, there are infinite pairs of orthogonal states of polarization. In general, we can draw the polarization ellipses of the two orthogonal states using the Poincare' sphere. Consider an arbitrary state of polarization denoted by $A = (\epsilon, \tau)$. Then its orthogonal state $\bar{A}$ will correspond to $(-\epsilon, \tau + \pi/2)$. Since the magnitude of ϵ corresponds to the axial ratio (AR) of the ellipse and its sign gives the sense of rotation, the two ellipses have same axial ratio but opposite sense of rotation. If a polarization state A lies in the northern hemisphere, its sense of rotation is LH and hence the sense of rotation for the orthogonal state $\bar{A}$ is RH, and vice-versa. The major axes of the two ellipses are perpendicular to each other since one has a tilt angle of τ and other of $(\tau + \pi/2)$.

The two polarization ellipses corresponding to the two orthogonal states can be as shown in Fig. 4.15(b)

As special cases of elliptical polarization, when AR becomes 1 or ∞ we get circular or linear polarizations respectively. The orthogonal states for the two special cases are as shown in Fig. 4.15(a) and (c).

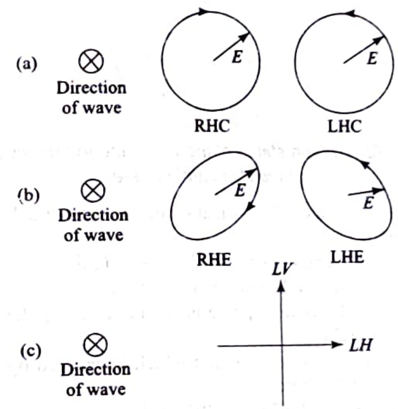


Fig. 4.15 *(a) Orthogonal states of circular polarization (b) Orthogonal states of elliptical polarization (c) Orthogonal states of linear polarization.*

Before we close the discussion on wave polarization, it is worth mentioning that an arbitrary state of polarization can be decomposed into any arbitrary pair of orthogonal states. In other words, any arbitrary polarization can be generated by superposition of any two orthogonal states of polarization. We may recall that, at the beginning of the section, we had started with two electric field components E_x and E_y and had derived the general elliptical state of polarization. Now, having developed the understanding of different states of polarization, we may notice that we in fact had taken two orthogonal linear polarizations and their combinations led to a variety of other polarizations like the circular and the elliptical polarizations. Instead of linear polarizations, we could have as well taken two circular polarizations (LC and RC) or two orthogonal elliptical polarizations to produce any required state of polarization.

EXAMPLE 4.7 A RHE polarized wave with tilt angles 45° and $AR = 2$ is received by a LHC antenna system. Find the power transfer efficiency from the wave to the antenna.

Solution:

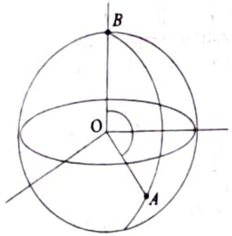


Fig. 4.16 *Polarization states of the wave and the receiving antenna marked on the Poincare' sphere.*

Let us first locate the states of polarization of the wave and the antenna on the Poincare' sphere.
For the wave: $\tau = 45^\circ$ and $\epsilon = \cot^{-1}(-2) = -0.4636$
(Note for RH polarization the axial ratio is negative.)
The location on the Poincare' sphere is (Point A in Fig. 4.16), $(long, lat)$= $(2\tau, 2\epsilon) = (\frac{\pi}{2}, -0.927)$
The antenna has a state of polarization which is denoted by point B (North Pole) on the Poincare' sphere.
The angle $\angle AOB = \frac{\pi}{2} + 0.927 = 2.4978$ rad
The polarization efficiency $\eta_{pol} = \cos^2(\frac{\angle AOB}{2}) = 0.10$

EXAMPLE 4.8 Show that a linear polarization can be generated by superposition of two circularly polarized waves. Explain how a linearly polarized wave with tilt angle of $\frac{\pi}{3}$ can be generated by two circularly polarized waves.

Solution:

Any arbitrary state of polarization can be obtained by superposition of any two orthogonal states of polarization. We can, therefore, obtain any state of polarization by combining LHC and RHC since they are orthogonal states.

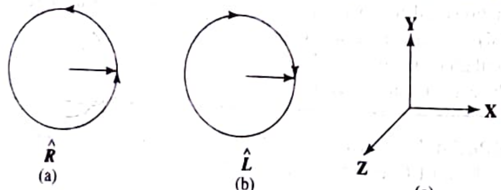


Fig. 4.17 *Orthogonal circular polarization for Example 4.8.*

Let us now define unity amplitude LHC and RHC fields. They obviously represent circular motion around a circle of unity radius. Let the unity LHC and RHC fields be denoted by vectors $\hat{\mathbf{L}}$ and $\hat{\mathbf{R}}$ respectively. (see Fig. 4.17) Let us now represent $\hat{\mathbf{L}}$ and $\hat{\mathbf{R}}$ in terms of unit vectors along x and y directions. As we have seen above, combination of $E_x = 1$ and $E_y = 1\angle\frac{\pi}{2}$ gives the LHC polarization, and $E_x = 1$ and $E_y = 1\angle\frac{-\pi}{2}$ gives the RHC polarization. We can, therefore, write

$$\hat{\mathbf{L}} = \hat{\mathbf{x}} + j\hat{\mathbf{y}}$$

$$\hat{\mathbf{R}} = \hat{\mathbf{x}} - j\hat{\mathbf{y}}$$

Solving the above two equations we get

$$\hat{\mathbf{x}} = \frac{\hat{\mathbf{L}} + \hat{\mathbf{R}}}{2}$$

and

$$\hat{\mathbf{y}} = \frac{\hat{\mathbf{L}} - \hat{\mathbf{R}}}{2j}$$

Any arbitary electric field E can be represented in terms of $(\hat{\mathbf{x}},\hat{\mathbf{y}})$ unit vectors or $(\hat{\mathbf{L}},\hat{\mathbf{R}})$ unit vectors. That is,

$$\Rightarrow \quad \mathbf{E} = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} = E_L\hat{\mathbf{L}} + E_R\hat{\mathbf{R}}$$

$$= E_x(\frac{\hat{\mathbf{L}} + \hat{\mathbf{R}}}{2}) + E_y(\frac{\hat{\mathbf{L}} - \hat{\mathbf{R}}}{2j})$$

where E_x, E_y, E_L, E_R are complex quantities. Seperating vector components, we get

$$E_L = \frac{E_x - jE_y}{2}$$

$$E_R = \frac{E_x + jE_y}{2}$$

Now we know that for a linear polarization, E_x and E_y are in phase. Without losing generality let us assume E_x and E_y to be real. We then get

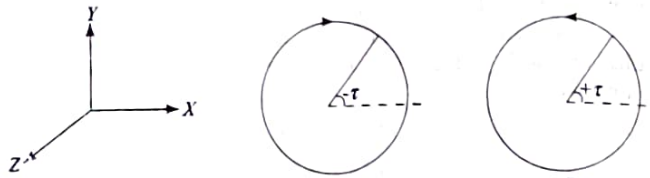
$$E_L = \frac{1}{2}\sqrt{E_x^2 + E_y^2} \quad \angle - \tan^{-1}(\frac{E_y}{E_x})$$

and

$$E_R = \frac{1}{2}\sqrt{E_x^2 + E_y^2} \quad \angle \tan^{-1}(\frac{E_y}{E_x})$$

This suggests that we can get linear polarization if $|E_L| = |E_R|$, and
The tilt angle $\tau = \tan^{-1}(\frac{E_y}{E_x}) = -\angle E_L = +\angle E_R$.
Now, the phase of a rotating vector means excess angle of rotation. If the phase is +ve, the excess angle is in the direction of rotation and if the phase is negative, the excess angle is in the direction opposite to the rotation.

Now, E_L has a phase of $-\tau$ and E_R a phase of $+\tau$. At some instant of time say $t = 0$, with respect to x-axis therefore we have E_L and E_R vectors as shown in Fig. 4.18.



Wave motion (a) E_L (b) E_R (c)

Fig. 4.18

The radii of the two circles are $\frac{1}{2}\sqrt{E_x^2 + E_y^2}$. We can verify that the superposition of E_L and E_R always gives a resultant vector which is at an angle τ with respect to the x-axis.

We can, hence, conclude that two circularly polarized waves of equal amplitude but with a phase difference of 2τ generate a linearly polarized wave of tilt angle τ.

For a linear polarization with $\tau = \frac{\pi}{3}$, the phase difference between E_L and E_R should be $\frac{2\pi}{3}$ with E_R leading E_L.

4.6 WAVE PROPAGATION IN CONDUCTING MEDIUM

In the previous section, we investigated the plane wave propagation in a source-free medium. We also assumed that the medium is an ideal dielectric, i.e. the conductivity of the medium was zero. In this section we study the wave propagation in a medium which does not have free charges but which has finite conductivity and consequently has conduction current **J**. In the presence of the finite conductivity, the Maxwell's equation for time harmonic fields can be written as

$$\nabla \cdot \mathbf{D} = 0 \quad \Rightarrow \quad \nabla \cdot \mathbf{E} = 0 \tag{4.92}$$

$$\nabla \cdot \mathbf{B} = 0 \quad \Rightarrow \quad \nabla \cdot \mathbf{H} = 0 \tag{4.93}$$

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} = -j\omega\mu_o\mu_r \mathbf{H} \tag{4.94}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + j\omega\epsilon \mathbf{E} = \mathbf{J} + j\omega\epsilon_0\epsilon_r \mathbf{E} \tag{4.95}$$

where, μ_r and ϵ_r are relative permeability and relative permittivity of the medium respectively. If the medium has conductivity σ, the conduction current density **J** is given by the Ohm's law as

$$\mathbf{J} = \sigma \mathbf{E} \tag{4.96}$$

Substituting for **J** in Eqn (4.95), we get

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} + j\omega\epsilon_0\epsilon_r \mathbf{E} = (\sigma + j\omega\epsilon_0\epsilon_r)\mathbf{E} \tag{4.97}$$

Equation (4.97) can be re-written as

$$\nabla \times \mathbf{H} = j\omega\epsilon_0 \left\{ \epsilon_r - j\frac{\sigma}{\omega\epsilon_0} \right\} \mathbf{E} \equiv j\omega\epsilon_0\epsilon_{rc}\mathbf{E} \tag{4.98}$$

where we define the relative permittivity of the conducting medium as

$$\epsilon_{rc} = \epsilon_r - j\frac{\sigma}{\omega\epsilon_0} \tag{4.99}$$

It is then clear that the relative permittivity (also called the dielectric constant) of a conducting medium is always complex and it is a function of frequency. The behavior of the medium now becomes frequency dependent.

Let us look at Eqn (4.97) once again. We note that, the two terms $\sigma\mathbf{E}$ and $j\omega\epsilon_0\epsilon_r\mathbf{E}$ correspond to the conduction and displacement current densities respectively. The conduction current is a characteristic of a conductor whereas, the displacement current is a characteristic of a dielectric. For a medium which has conduction as well displacement current one would then wonder whether to call the medium a conductor or a dielectric! The answer to this question lies in the relative contributions of the two currents. For a given electric field if the conduction current is larger compared to the displacement current, we can treat the medium like a conductor, whereas if the conduction current is negligible compared to the displacement current, we can treat the medium like a dielectric. That is,

$$\text{If} \quad \frac{\text{Cond. current density}}{\text{Disp. current density}} = \frac{\sigma}{\omega\epsilon_0\epsilon_r} \gg 1 \text{ — good conductor} \tag{4.100}$$

$$\text{If} \quad \frac{\text{Cond. current density}}{\text{Disp. current density}} = \frac{\sigma}{\omega\epsilon_0\epsilon_r} \ll 1 \text{ — good dielectric} \tag{4.101}$$

If the ratio $\sigma/\omega\epsilon_0\epsilon_r$ is ~ 1 then the medium can neither be called a good conductor nor a good dielectric.

Now, since the ratio $\sigma/\omega\epsilon_0\epsilon_r$ is a function of frequency, a medium (with conductivity σ and dielectric constant ϵ_r) can behave like a dielectric at one frequency and like a conductor at another frequency. Noting that the ratio is inversely proportional to ω, we can say in general, that towards the lower end of the electromagnetic spectrum a medium behaves more like a conductor (except when $\sigma = 0$) and as the frequency increases the behavior changes towards that of a dielectric.

This also suggests that, absolute value of conductivity is not very meaningful in calling a medium a good conductor. Copper, which has a conductivity of 5.6×10^7 ℧/m, is a good conductor but so is sea water which has a conductivity of 10^{-3} ℧/m. The only difference is, the sea water is a good conductor only at frequencies below few hundred kHz, whereas copper is a good conductor at all frequencies upto THz.

One can find the change over frequency f_T at which the medium behavior changes from conductor to dielectric or vice-versa by making the conduction and the displacement currents equal. At $f = f_T$ we have

$$\omega\epsilon_0\epsilon_r = \sigma \tag{4.102}$$

$$\Rightarrow \qquad f_T = \frac{\sigma}{2\pi\epsilon_0\epsilon_r} \tag{4.103}$$

For copper, taking $\epsilon_r \approx 1$ and $\sigma = 5.6 \times 10^7$ ℧/m, f_T is $\approx 10^{18}$ Hz. For sea water on the other hand, taking $\epsilon_r \approx 80$, and $\sigma = 10^{-3}$ ℧/m, f_T is ≈ 225 kHz.

The concept of complex dielectric constant is quite useful in analysing electromagnetics problems. For a non-ideal dielectric medium, one can first analyse the problem assuming the dielecric to be ideal with dielectric constant ϵ_r. The results for non-ideal dielectric medium can then be obtained by replacing the dielectric constant ϵ_r by ϵ_{rc}.

To analyse the plane wave propagation in a dielectric medium with finite conductivity, we use the concept of the complex dielectric constant. For a conducting medium, the wave Eqns (4.31) and (4.32) respectively, become

$$\nabla^2\mathbf{E} = -\omega^2\mu_0\mu_r\epsilon_0\epsilon_{rc}\mathbf{E} = \sqrt{j\omega\mu_0\mu_r(\sigma + j\omega\epsilon_0\epsilon_r)}\ \mathbf{E} \tag{4.104}$$

$$\nabla^2\mathbf{H} = -\omega^2\mu_0\mu_r\epsilon_0\epsilon_{rc}\mathbf{H} = \sqrt{j\omega\mu_0\mu_r(\sigma + j\omega\epsilon_0\epsilon_r)}\ \mathbf{H} \tag{4.105}$$

Here, we have explicitly written $\mu = \mu_0\mu_r$ and $\epsilon = \epsilon_0\epsilon_{rc}$. μ_0 and ϵ_0 are free space permeability and permittivity respectively. Since, here, we are primarily interested in non-magnetic media, we can take $\mu_r = 1$ in our further discussion of the wave propagation.

The Eqn (4.51) of a plane wave travelling in z direction now becomes

$$\frac{\partial^2 E_x}{\partial z^2} = -\omega^2\mu_0\epsilon_0\epsilon_{rc}E_x \equiv \gamma^2 E_x \tag{4.106}$$

The propagation constant of the wave therefore is

$$\gamma = \sqrt{-\omega^2\mu_0\epsilon_0\epsilon_{rc}} = +j\omega\sqrt{\mu_0\epsilon_0}\sqrt{\epsilon_{rc}} \tag{4.107}$$

Substituting for ϵ_{rc} from Eqn (4.99), the propagation constant of the wave is

$$\gamma = j\omega\sqrt{\mu_0\epsilon_0}\left\{\epsilon_r - j\frac{\sigma}{\omega\epsilon_0}\right\}^{1/2} \tag{4.108}$$

The propagation constant for a conducting medium, therefore, is complex and can be written as

$$\gamma = \alpha + j\beta \tag{4.109}$$

where

$$\alpha = Re(\gamma) = \omega\sqrt{\frac{\mu_0\epsilon_0\epsilon_r}{2}}\left[\sqrt{1+\frac{\sigma^2}{\omega^2\epsilon_0^2\epsilon_r^2}} - 1\right]^{1/2} \tag{4.110}$$

$$\beta = Im(\gamma) = \omega\sqrt{\frac{\mu_0\epsilon_0\epsilon_r}{2}}\left[\sqrt{1+\frac{\sigma^2}{\omega^2\epsilon_0^2\epsilon_r^2}} + 1\right]^{1/2} \tag{4.111}$$

Now, following the discussion on transmission lines, we note that, α is the attenuation constant and β is the phase constant. α decides the change in amplitude of a wave as it propagates in the medium and β decides the wavelength of the wave in the medium (wavelength $\lambda = 2\pi/\beta$). It is then interesting to note from Eqns (4.110) and (4.111) that in a dielectric medium with non-zero conductivity the attenuation constant and the wavelength of a wave are functions of frequency. It is also interesting that with increasing σ the attenuation of the wave increases and the wavelength of the wave decreases (β increases). From our basic understanding of electrical devices, we expect the attenuation (or loss) to decrease with increase in conductivity. The dielectric medium with finite conductivity however has exactly opposite behavior, i.e. increase in conductivity increases the attenuation. One would then wonder how there are opposite trends for the circuit behavior and the wave behavior! The surprise can be resolved by noting that in electrical circuits we primarily concentrate on the conductive current and treat the medium as a conductor whereas in the above analysis we have treated the medium as a dielectric with finite conductivity. A dielectric medium has attenuation if the conductivity is not zero and a conductor has attenuation if the conductivity is not infinite (this will become clear in later sections).

Since, here, we are primarily investigating the wave propagation in a dielectric medium, there is a wave attenuation if the conductivity of the medium is non-zero and for $\sigma = 0$ the attenuation is zero. A quantity called 'loss tangent' (denoted by $\tan\delta$) which is the tangent of the phase of the complex dielectric constant is normally used as a measure of the medium attenuation.

The loss tangent is

$$\tan\delta = \frac{\sigma}{\omega\epsilon_0\epsilon_r} \tag{4.112}$$

Smaller the loss tangent lesser is the attenuation and better is the dielectric.

4.6.1 Low-loss Dielectrics

A low-loss dielectric is a medium for which the loss tangent is very small, i.e. $\sigma/\omega\epsilon_0\epsilon_r << 1$. For this medium the expressions for the attenuation and the phase constants in Eqns (4.110) and (4.111) can be approximated. Making binomial expansion for the term inside the brackets of (4.110) we get

$$\left[\sqrt{1+\frac{\sigma^2}{\omega^2\epsilon_0^2\epsilon_r^2}} - 1\right]^{\frac{1}{2}} = \left[1 + \frac{1}{2}\frac{\sigma^2}{\omega^2\epsilon^2} - \frac{1}{8}\left(\frac{\sigma^2}{\omega^2\epsilon^2}\right)^2 + \cdots - 1\right]^{\frac{1}{2}} \tag{4.113}$$

$$= \left[\frac{1}{2}\frac{\sigma^2}{\omega^2\epsilon^2} - \frac{1}{8}\left(\frac{\sigma^2}{\omega^2\epsilon^2}\right)^2 + \cdots \right]^{\frac{1}{2}} \tag{4.114}$$

$$= \sqrt{\frac{\sigma^2}{2\omega^2\epsilon^2}} \left[1 - \frac{1}{2}\left(\frac{\sigma^2}{4\omega^2\epsilon^2}\right) + \cdots \right] \tag{4.115}$$

and the attenuation constant α can be written as

$$\alpha = \omega\sqrt{\frac{\mu_0\epsilon}{2}\frac{\sigma^2}{2\omega^2\epsilon^2}} \left[1 - \frac{\sigma^2}{8\omega^2\epsilon^2} + \cdots \right] \tag{4.116}$$

$$\approx \frac{\sigma}{2}\sqrt{\frac{\mu_0}{\epsilon}} \left[1 - \frac{\sigma^2}{8\omega^2\epsilon^2} \right] \tag{4.117}$$

$$\approx \frac{\sigma}{2}\sqrt{\frac{\mu_0}{\epsilon}} \tag{4.118}$$

Similarly the phase constant β can be approximated as

$$\beta \approx \omega\sqrt{\mu_0\epsilon}\left[1 + \frac{1}{8}\frac{\sigma^2}{\omega^2\epsilon^2} \right] \tag{4.119}$$

$$\approx \omega\sqrt{\mu_0\epsilon} \tag{4.120}$$

It can be seen from Eqn (4.118) that the attenuation constant α is directly proportional to the conductivity of the medium and is practically independent of frequency. The phase constant however is almost same as that of the loss-less medium (see Eqn (4.120)).

The intrinsic impedance of a low-loss medium is

$$\eta_d = \sqrt{\frac{j\omega\mu_0}{\sigma + j\omega\epsilon}} \tag{4.121}$$

$$\approx \sqrt{\frac{\mu_0}{\epsilon_o\epsilon_r}} = \eta_0\sqrt{\frac{1}{\epsilon_r}} \tag{4.122}$$

The phase of η_d is almost zero and hence the electric and magnetic fields are almost in phase with each other.

EXAMPLE 4.9 A dielectric material has relative permittivity 18 and loss tangent 10^{-3} at 100 MHz. Find the conductivity of the medium. Also find the distance over which the wave amplitude reduces to $1/e$ of its original amplitude?

Solution:

The loss tangent is

$$\tan\delta = \frac{\sigma}{\omega\epsilon_o\epsilon_r}$$

$$\Rightarrow \qquad \sigma = \omega\epsilon_o\epsilon_r \tan\delta$$

$$= 2\pi \times 10^8 \times \frac{1}{36\pi} \times 10^{-9} \times 18 \times 10^{-3}$$

$$= 10^{-4}\ \mho/\text{m}.$$

Since the loss-tangent is very small, the dielectric is a low-loss dielectric, and we can approximately get the attenuation constant of the wave as (see Eqn (4.118))

$$\alpha = \frac{\sigma}{2}\sqrt{\frac{\mu_o}{\epsilon_o\epsilon_r}} = \frac{\sigma}{2}\sqrt{\frac{\mu_o}{\epsilon_o}}.\frac{1}{\sqrt{\epsilon_r}} = \frac{10^{-4}}{2} \times 120\pi \times \frac{1}{\sqrt{18}} = 4.44 \times 10^{-3}\text{nepers/m}$$

The distance over which the wave amplitude reduces to 1/e of its original value is $1/\alpha = 225.08$ m

4.6.2 Good Conductor

As discussed in the previous sections, for a good conductor we have $\sigma/\omega\epsilon >> 1$ and we can apply some approximations to the amplitude and phase constants. The propagation constant in general can be written as

$$\gamma = \alpha + j\beta = \sqrt{j\omega\mu_0(\sigma + j\omega\epsilon_0\epsilon_r)} \tag{4.123}$$

$$\approx \sqrt{j\omega\mu_0\sigma} \qquad \text{since } \sigma >> \omega\epsilon_0\epsilon_r \tag{4.124}$$

since, $\sqrt{j} = \sqrt{e^{j\pi/2}} = e^{j\pi/4} = \cos\frac{\pi}{4} + j\sin\frac{\pi}{4} = \frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}} = \frac{1+j}{\sqrt{2}}$, we get

$$\gamma = \alpha + j\beta = \sqrt{\omega\mu_0\sigma}\left(\frac{1+j}{\sqrt{2}}\right) \tag{4.125}$$

giving

$$\alpha = \beta = \sqrt{\frac{\omega\mu_0\sigma}{2}} \tag{4.126}$$

It is interesting to note that for a good conductor, α and β are approximately equal. As the amplitude of a wave varies as $e^{-\alpha z}$, the wave amplitude reduces to $1/e$ of its value over a distance of $1/\alpha$. That is, over a distance of $1/\alpha = 1/\beta = \lambda/2\pi \approx \lambda/6$, the wave amplitude reduces to $1/e$ of its initial value. A conductor, therefore, acts like a very lossy medium. The field decays very rapidly from the point of its origin along the direction of the wave propagation.

At this point it is worthwhile to ask a question, "How is a field excited inside a conductor?" For transportation of electromagnetic energy, again we have to depend upon the wave propagation but the wave propagation is very lossy in a conductor. It is, therefore, apparent that we can not generate energy inside a conductor. We can generate energy inside a dielectric, and transport the energy to a conductor. The question then reduces to what happens to an electromagnetic wave when it reaches a conducting medium? Without worrying about the reflections, etc at the conductor surface, let us assume that after all transient adjustments have taken place, the field just inside the conductor surface has some value E_0. Due to high attenuation constant α, the field decreases rapidly as the wave propagates deeper in the conductor. The wave amplitude reduces to $1/e$ of its value at the surface, over a distance of $1/\alpha$ and within a distance of few times $1/\alpha$ the field reduces practically to zero. Figure 4.19 shows the amplitude of a wave as a function of depth inside a conductor. The field is, therefore, effectively confined to a layer which is $\sim 1/\alpha$ deep below the surface of the conductor. The thickness of the layer ($\sim 1/\alpha$) decreases as ω and σ increase. At a frequency of tens of MHz, and for conductivity of $\sim 10^7$ ℧/m (good conductor), the thickness lies in the range of μm. The field confinement then is just in the skin of the conductor. This effect is therefore called the 'skin effect', and the thickness of the layer is called the 'skin depth' or depth of penetration. The skin depth is given as

$$\delta = \frac{1}{\alpha} = \sqrt{\frac{2}{\omega\mu_0\sigma}} = \sqrt{\frac{1}{\pi f \mu_0 \sigma}} \tag{4.127}$$

The skin effect can be wisely exploited for shielding an electromagnetic wave. If a region with electromagnetic radiation is covered with a metal sheet of thickness much larger than the skin depth the fields outside the metal cover would be of negligibly small value.

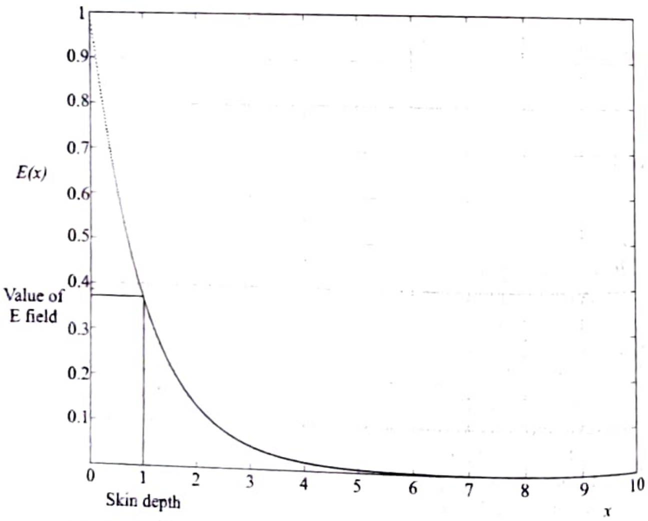


Fig. 4.19 *A decaying electrical field inside a good conductor.*

The skin effect which helps in isolating electromagnetic environments acts as a hinderance for the radio wave communication through the conducting walls. We all have noticed the poor reception of radio signals inside a train. The skin effect is the primary cause of that.

Table 4.1 shows the skin depth values for different conductors at a frequency of 1 MHz. The skin-depth is inversely proportional to the square-root of the frequency, and the square-root of the conductivity.

Table 4.1 Skin depth at 1 MHz

Material	Conductivity	Skin depth
Silver	6.17×10^7 ℧/m	63.87 μm
Copper	5.88×10^7 ℧/m	76.33 μm
Aluminium	3.65×10^7 ℧/m	83.17 μm
Doped Silicon	10^3 ℧/m	6.39 mm
Earth	5×10^{-3} ℧/m	7.12 m

The intrinsic impedance of a good conductor is

$$\eta_c \approx \sqrt{\frac{j\omega\mu_0}{\sigma}} = \sqrt{\frac{\omega\mu_0}{\sigma}}\angle 45^\circ \tag{4.128}$$

The magnetic field, therefore, lags the electric field by $\approx 45^o$ inside a good conductor.

EXAMPLE 4.10 A material has dielectric constant 25 and conductivity 2×10^6 ℧/m. What is the frequency above which the material cannot behave like a good conductor? If a plane wave of 10 MHz is incident on the material, effectively upto what depth can the wave penitrate the material, and what will be the wavelength of the wave inside the material?

Solution:

For any material to behave like a good conductor, we must have $\sigma >> \omega\epsilon_0\epsilon_r$. As a rule of thumb, we may say that when $\sigma > 10\omega\epsilon_0\epsilon_r$, the material is a good conductor. Therefore, the frequency above which the material does not behave like a good conductor, is

$$\omega = \frac{\sigma}{10\epsilon_0\epsilon_r} = \frac{2 \times 10^6}{10 \times \frac{1}{36\pi} \times 10^{-9} \times 25}$$

$$\Rightarrow \quad f = 1.44 \times 10^{14} \text{ Hz.}$$

Effective depth of wave penitration is the skin depth,

$$\delta = \frac{1}{\sqrt{\pi f \mu_0 \sigma}} = \frac{1}{\sqrt{\pi \times 10^7 \times 4\pi \times 10^{-7} \times 2 \times 10^6}} = 112.54\ \mu\text{m}$$

To find the wavelength we first calculate the phase constant $\beta = \sqrt{\pi f \mu_0 \sigma} =$ 8885.5 rad/m.
The wavelength $\lambda = \frac{2\pi}{\beta} = 0.707$ mm.

4.7 PHASE VELOCITY OF A WAVE

As we have seen above, the electric field of a plane wave travelling in $+z$ direction is written as

$$\mathbf{E}(z, t) = \mathbf{E}_0 e^{-\gamma z} \cdot e^{j\omega t} \tag{4.129}$$

For a general medium, γ is complex ($\gamma \equiv \alpha + j\beta$), and the electric field can be written as

$$\mathbf{E}(z, t) = E_0 e^{-\alpha z} \cdot e^{-j\beta z} \cdot e^{j\omega t} \tag{4.130}$$

$$= E_0 e^{-\alpha z} \cdot e^{j(\omega t - \beta z)} \tag{4.131}$$

As has been discussed in detail for the transmission lines, the phase ϕ of $\mathbf{E}$ is a composite function of space (z) and time (t). The constant phase point moves in $+z$ direction as a function of time. One can then ask a question, "with what speed does the constant phase point move in the space?"

Let us consider an observer standing at some location in space and seeing the wave passing by him. Since $z = constant$ for the observer, he observes the phase of the wave to be ωt, i.e. the phase increases linearly as a function of time. Now, consider an observer who is holding on to a particular phase point. Obviously, he has to move with the speed with which the phase point is moving so that he never leaves the point. Then for this observer, the phase appears stationary (constant) as a function of time. The velocity of the wave is same as the velocity of the observer for whom the phase is constant. This velocity is called the 'phase velocity' of the wave. For a wave travelling in $+z$ direction, the phase of the wave is

$$\phi = \omega t - \beta z = constant \tag{4.132}$$

Differentiating Eqn (4.132) with respect to time we get

$$\omega - \beta \frac{\partial z}{\partial t} = 0 \tag{4.133}$$

$$\Rightarrow \quad \text{phase velocity} : v_p = \frac{\partial z}{\partial t} = \frac{\omega}{\beta} \tag{4.134}$$

Since β in general, is a function of the dielectric constant, ϵ_r of the medium, and the conductivity σ, the phase velocity also changes from medium to medium. In fact, the phase velocity is one of the parameters which characterizes the medium. Let us find the phase velocities of a wave in different media.

1. Free Space: For free space (or vacuum) we have $\epsilon = \epsilon_0 = \frac{1}{36\pi} \times 10^{-9}$ F/m $\mu = \mu_0 = 4\pi \times 10^{-7}$ H/m. We therefore have

$$\beta = \omega\sqrt{\mu_0 \epsilon_0} = \frac{\omega}{3} \times 10^{-8} \text{ rad/m} \tag{4.135}$$

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = 3 \times 10^8 \text{ m/s} \tag{4.136}$$

This is the velocity of light in the free-space and is normally denoted by letter 'c'. One would appreciate the strength of the Maxwell's equations now which correctly predicted the wave phenomenon and provided correct estimate of the velocity of light in vacuum.

2. Pure Dielectric: For a pure dielectric we have $\sigma = 0$, $\epsilon = \epsilon_0 \epsilon_r$, $\mu = \mu_0$

$$\Rightarrow \quad \beta = \omega\sqrt{\mu_0 \epsilon_0 \epsilon_r} = \omega\sqrt{\mu_0 \epsilon_0}\sqrt{\epsilon_r} \tag{4.137}$$

$$\Rightarrow \quad v_p = \frac{\omega}{\omega\sqrt{\mu_0 \epsilon_0}\sqrt{\epsilon_r}} = \frac{c}{\sqrt{\epsilon_r}} \tag{4.138}$$

Equation (4.138) gives the velocity of light in a medium with dielectric constant ϵ_r. From Eqn (4.138) we note that the velocity of an electromagnetic wave in a dielectric medium is always less than that in the vacuum and is independent of the frequency of the wave.

We may recall from our high school physics that the refractive index n of a medium is defined as

$$n = \frac{\text{velocity of light in vaccum}}{\text{velocity of light in the medium}} \tag{4.139}$$

Therefore, from Eqns (4.138) and (4.139) we get

$$n = \sqrt{\epsilon_r} \tag{4.140}$$

The refractive index of a dielectric material is the square root of its dielectric constant.

3. Dielectric Medium with Loss: For a lossy dielectric, $\sigma \neq 0$, and the phase constant is given as (from Eqn (4.119))

$$\beta = \omega\sqrt{\mu_0 \epsilon_0 \epsilon_r}\left\{1 + \frac{1}{8}\frac{\sigma^2}{\omega^2 \epsilon^2}\right\} \tag{4.141}$$

hence the phase velocity is

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu_0 \epsilon_0 \epsilon_r}}\left\{1 + \frac{1}{8}\frac{\sigma^2}{\omega^2 \epsilon^2}\right\}^{-1} = \frac{c}{\sqrt{\epsilon_r}}\left\{1 + \frac{1}{8}\frac{\sigma^2}{\omega^2 \epsilon_0^2 \epsilon_r^2}\right\}^{-1} \tag{4.142}$$

From Eqn (4.142) we note that for a lossy dielectric medium, the phase velocity is a function of ω, i.e. the waves of different frequencies travel with different velocities. This phenomenon is called the 'dispersion', and the medium is said to be a 'dispersive medium'. One can then say that a loss-less dielectric is non-dispersive, whereas a lossy medium is dispersive and the dispersion increases with the loss in the medium. For a low-loss medium however, the dispersion is generally small.

4. Good Conductor: For a good conductor $\sigma >> \omega\epsilon$ and the phase constant is (Eqn (4.126))

$$\beta = \sqrt{\frac{\omega\mu_0\sigma}{2}} \tag{4.143}$$

$$\Rightarrow \quad v_p = \frac{\omega}{\sqrt{\frac{\omega\mu_0\sigma}{2}}} = \sqrt{\frac{2\omega}{\mu_0\sigma}} \tag{4.144}$$

Multiplying numerator and denominator within the square root sign of Eqn (4.144) by $\omega\epsilon_0$ and re-arranging we get

$$v_p = \frac{1}{\sqrt{\mu_0\epsilon_0}}\sqrt{\frac{2\omega\epsilon_0}{\sigma}} = c\sqrt{\frac{2\omega\epsilon_0}{\sigma}} \tag{4.145}$$

For a good conductor $\omega\epsilon_0/\sigma << 1$ and therefore $v_p << c$. The electromagnetic wave therefore slows down considerably in a conductor. As can be seen from the example, the phase velocity of an electromagnetic wave inside copper is few hundred meter/s which is of the order of the velocity of sound in copper.

Compared to a lossy dielectric, the conductor is much dispersive since v_p varies as $\sqrt{\omega}$. It should be noted however, that the dispersion decreases with σ and for an ideal conductor ($\sigma = \infty$) the dispersion is zero. Figures 4.20 and 4.21 show velocity and dispersion as a function of frequency and the conductivity of the medium.

EXAMPLE 4.11 Just outside a train compartment the field strength of a radio station is 0.1 V/m. What will be the approximate field strength inside the compartment? Assume the compartment to be a closed box of metal having conductivity 5×10^6 ℧/m. The thickness of the compartment wall is 5 mm. Frequency of radio station is 600 kHz.

Solution:

The attenuation constant for the compartment material (assuming the compartment material non-magnetic)

$$\alpha = \sqrt{\pi f \mu_o \sigma} = \sqrt{\pi \times 600 \times 10^3 \times 4\pi \times 10^{-7} \times 5 \times 10^6} = 3441.44 \text{ napers/m.}$$

The wave amplitude after passing through the compartment wall will be

$$E = E_o e^{-\alpha t} = 0.1 e^{-17.2} = 3.366 \times 10^{-9} \text{ V/m}$$

The wave, therefore, is attenuated by a factor $\sim 3 \times 10^7$.

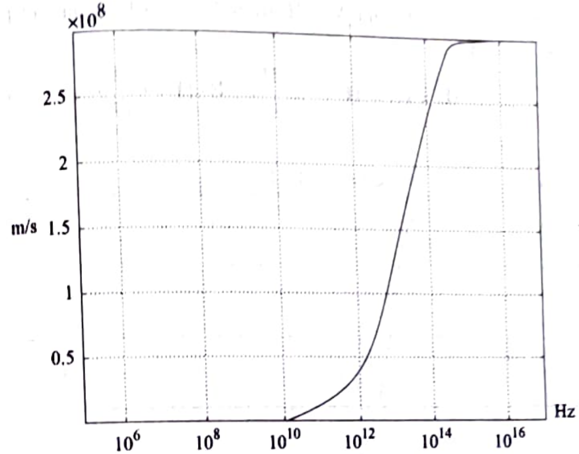


Fig. 4.20 *Variation of the phase velocity as a function of frequency.*

4.8 POWER FLOW AND POYNTING VECTOR

As seen above, the time varying electric and magnetic fields have to form an electromagnetic wave which propagates in the space. Naturally the wave carries some energy with it. It is then worthwhile to investigate, the quantity energy or power (i.e. energy per unit time) is carried by an electromagnetic wave? Note that in general, the electromagnetic wave need neither be a plane wave nor be travelling in an unbound medium. We, therefore, would like to develop a general frame work for power flow from arbitrary time varying fields. Since the electromagnetic phenomenon is completely governed by the Maxwell's equations, we again fall back upon the Maxwell's equations to find the power flow due to time varying fields.

Let us take the two Maxwell's equations

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\mu\frac{\partial \mathbf{H}}{\partial t} \tag{4.146}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J} + \epsilon\frac{\partial \mathbf{E}}{\partial t} \tag{4.147}$$

We assume here that μ, ϵ are not varying as a function of time. From the vector identity we have

$$\nabla \cdot (\mathbf{A} \times \mathbf{C}) = \mathbf{C} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot \nabla \times \mathbf{C} \tag{4.148}$$

where **A** and **C** are any two arbitrary vectors

Taking **A** = **E** and **C** = **H** we have the vector identity

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}) \tag{4.149}$$

Substituting for $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{H}$ from Eqns (4.146) and (4.147) in Eqn (4.149) we get

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (-\mu \frac{\partial \mathbf{H}}{\partial t}) - \mathbf{E} \cdot \{\mathbf{J} + \epsilon \frac{\partial \mathbf{E}}{\partial t}\} \tag{4.150}$$

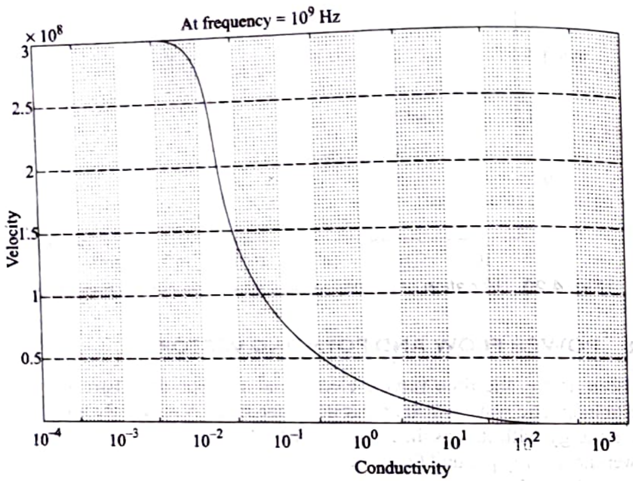


Fig. 4.21 *Variation of the phase velocity as a function of conductivity.*

Here we have assumed that the medium is isotropic and hence μ, ϵ and σ are scalar quantities. Also note that, for any two vectors **A** and **C** we have

$$\frac{\partial(\mathbf{A} \cdot \mathbf{C})}{\partial t} = \mathbf{A} \cdot \frac{\partial \mathbf{C}}{\partial t} + \mathbf{C} \cdot \frac{\partial \mathbf{A}}{\partial t} \tag{4.151}$$

$$\Rightarrow \quad \frac{\partial(\mathbf{A} \cdot \mathbf{A})}{\partial t} = 2\mathbf{A} \cdot \frac{\partial \mathbf{A}}{\partial t} \tag{4.152}$$

$$\Rightarrow \quad \mathbf{A} \cdot \frac{\partial \mathbf{A}}{\partial t} = \frac{1}{2}\frac{\partial(\mathbf{A} \cdot \mathbf{A})}{\partial t} = \frac{1}{2}\frac{\partial |\mathbf{A}|^2}{\partial t} \tag{4.153}$$

Noting that $\mathbf{J} = \sigma\mathbf{E}$, and taking σ not a function of time, we can write from Eqn (4.150)

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\frac{\mu}{2}\frac{\partial |\mathbf{H}|^2}{\partial t} - \frac{\epsilon}{2}\frac{\partial |\mathbf{E}|^2}{\partial t} - \sigma |\mathbf{E}|^2 \tag{4.154}$$

The Eqn (4.154) is essentially a point relation. That is, it should be valid at every point in the space at every instant of time.

If we now integrate Eqn (4.154) over a volume we get

$$\oint_V \nabla \cdot (\mathbf{E} \times \mathbf{H}) dv = \oint_V (-\frac{\mu}{2}\frac{\partial |\mathbf{H}|^2}{\partial t} - \frac{\epsilon}{2}\frac{\partial |\mathbf{E}|^2}{\partial t} - \sigma |\mathbf{E}|^2) dv \tag{4.155}$$

Application of the divergence theorem (see Eqn (3.16)) on the LHS yields

$$\oint_S (\mathbf{E} \times \mathbf{H}) \cdot \mathbf{da} = \oint_V -\frac{\mu}{2}\frac{\partial |\mathbf{H}|^2}{\partial t} dv - \oint_V \frac{\epsilon}{2}\frac{\partial |\mathbf{E}|^2}{\partial t} dv - \oint_V \sigma |\mathbf{E}|^2 dv \tag{4.156}$$

Taking the volume V constant as a function of time, we can interchange the integral and the $\frac{\partial}{\partial t}$ giving

$$\oint_S (\mathbf{E} \times \mathbf{H}) \cdot \mathbf{da} = -\frac{\partial}{\partial t}\oint_V \frac{\mu}{2}|H|^2 dv - \frac{\partial}{\partial t}\oint_V \frac{\epsilon}{2}|\mathbf{E}|^2 dv - \oint_V \sigma |E|^2 dv \tag{4.157}$$

$$= -\frac{\partial}{\partial t}\left(\oint_V \frac{\mu}{2}|H|^2 dv + \oint_V \frac{\epsilon}{2}|E|^2 dv\right) - \oint_V \sigma |E|^2 dv \tag{4.158}$$

In Eqn (4.158), the first term within the brackets gives the magnetic energy stored in the volume, whereas the second term represents the electric energy stored in the volume V. The quantity within the brackets therefore represents the total energy stored in the volume, and the first term on the RHS of Eqn (4.158) represents the rate of decrease of total energy stored in the volume V, i.e. decrease in power in the volume V.

The second term on the RHS gives the ohmic power loss in the volume V. This term then represents the electromagnetic energy converted to heat per unit time. The two terms on the RHS consequently represent the total decrease in the electromagnetic energy per unit time, i.e. the power loss from the volume. Since there are no other losses in the medium, the law of conservation of energy dictates that this power loss should be due to energy leaving the volume. The LHS of Eqn (4.158) hence gives the power coming out of the volume, i.e.

$$\text{Net Outward Power } W = \oint_S (\mathbf{E} \times \mathbf{H}) \cdot \mathbf{da} \tag{4.159}$$

Since the surface integral of $(\mathbf{E} \times \mathbf{H})$ gives the total power flow from the surface, the quantity $(\mathbf{E} \times \mathbf{H})$, therefore, represents the power density on the surface of the volume. The vector defined as

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} \tag{4.160}$$

is called the 'Poynting vector' and it gives the power flow density (power flow per unit area) at any point on the surface.

The statement that 'the surface integral of **P** over a closed surface (the LHS in Eqn (4.158)) is equal to the total power leaving the closed surface' is called the *Poynting theorem.* It should be emphasized however, that the Poynting theorem is strictly valid only for a closed surface. That is to say that the Eqn (4.160) is not a point relationship according to the Poynting theorem. The use of $(\mathbf{E} \times \mathbf{H})$ as power density at every point on the closed surface is arbitrary. There are certain classical cases which contradict the use of point relationship for the Poynting vector. For example, let us place a charge at one of the poles of a bar magnet as shown in Fig. 4.22. Now, let us take some point, say P, as shown in Fig. 4.22. At this point since $\mathbf{E}$ and $\mathbf{H}$ are not parallel, we have finite $\mathbf{E} \times \mathbf{H}$ which is pointing out of the plane of the paper. In other words, we have a power coming out of the paper at point P. Obviously, since in this case there is no sustained power flow, the answer obtained from the Poynting vector is absurd. If we consider however, a symmetrically located point P', the $\mathbf{E} \times \mathbf{H}$ vector here is of same magnitude as that at P but going inside the plane of the paper. The sum of the power flow at P and P' hence is zero. Extending the result to any closed surface around the magnet we find that the net power flow from the surface is zero even though the poynting vector $\mathbf{P}$ is non-zero on the surface. It might, therefore, appear that the use of $\mathbf{E} \times \mathbf{H}$ for power density at any location might give erroneous results at times. However, except for few pathological cases (like the one discussed above) the Poynting vector correctly gives the power density at a point in the space. The Eqn (4.160) hence is normally used as a point relation although it is not strictly a point relation.

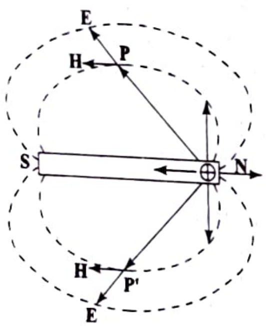


Fig. 4.22 *Poynting Vector for a system consisting of a charge placed on a bar magnet.*

The Poynting vector is a rather useful concept as it tells that the power flow is perpendicular to both the electric and the magnetic fields. The Poynting vector has the unit of Watts/m^2.

For a uniform plane wave travelling in $+z$ direction, the electric and the magnetic fields are oriented in $+x$ and $+y$ directions respectively. The $(\mathbf{E} \times \mathbf{H}) = (\hat{\mathbf{x}} \times \hat{\mathbf{y}})$ vector then is along $+z$ direction, i.e. the direction of the wave propagation. The Poynting vector hence correctly gives the direction of the power flow as the direction of the wave propagation.

EXAMPLE 4.12 A uniform plane wave has a power density of 20 W/m^2 and is travelling along $-y$ direction in the free space. If the electric field makes an angle of 30° with the $+x$ axis, find electric and magnetic field strengths and the direction of the magnetic field.

Solution:

For a uniform plane wave E and H are perpendicular to each other. Therefore, the poynting vector is

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} = |\mathbf{E}||\mathbf{H}|$$

$$= \frac{|\mathbf{E}|^2}{\eta_0} \qquad \text{since } |\mathbf{H}| = \frac{|\mathbf{E}|}{\eta_0}$$

$$\Rightarrow \qquad |E| = \sqrt{P\eta_0} = \sqrt{20 \times 120\pi} = 86.83 \text{ V/m}$$

$$\text{and} \qquad |H| = \frac{E}{120\pi} = 0.23 \text{ A/m}$$

Since, $\mathbf{E}$ and $\mathbf{H}$ are perpendicular to the direction of the wave, $-y$ axis, they must lie in xz plane. Also $\mathbf{E} \times \mathbf{H}$ should be in the direction of the wave. The $\mathbf{H}$ vector therefore will make an angle of 30° with the z-axis (see Fig. 4.23).

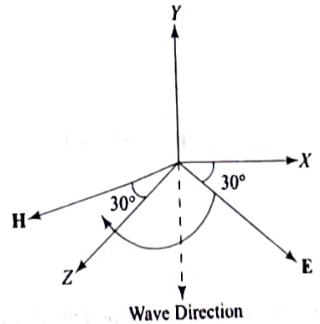


Fig. 4.23

4.8.1 Instantaneous and Average Poynting Vector

For time harmonic fields, it is rather useful to have the average power density (average over time) or average Poynting vector at any point in space. In the following sections, starting with the instantaneous $\mathbf{E}$, $\mathbf{H}$ and $\mathbf{P}$ we will derive the average power density or the average Poynting vector.

Writing $\mathbf{E}$ and $\mathbf{H}$ explicitly for the time harmonic function we have

$$\mathbf{E}(x, y, z, t) = \mathbf{E}_0(x, y, z)e^{j\omega t} \tag{4.161}$$

etc at the conductor surface, let us assume that after all transient adjustments have taken place, the field just inside the conductor surface has some value E_0. Due to high attenuation constant α, the field decreases rapidly as the wave propagates deeper in the conductor. The wave amplitude reduces to $1/e$ of its value at the surface, over a distance of $1/\alpha$ and within a distance of few times $1/\alpha$ the field reduces practically to zero. Figure 4.19 shows the amplitude of a wave as a function of depth inside a conductor. The field is, therefore, effectively confined to a layer which is $\sim 1/\alpha$ deep below the surface of the conductor. The thickness of the layer ($\sim 1/\alpha$) decreases as ω and σ increase. At a frequency of tens of MHz, and for conductivity of $\sim 10^7$ ℧/m (good conductor), the thickness lies in the range of μm. The field confinement then is just in the skin of the conductor. This effect is therefore called the 'skin effect', and the thickness of the layer is called the 'skin depth' or depth of penetration. The skin depth is given as

$$\delta = \frac{1}{\alpha} = \sqrt{\frac{2}{\omega\mu_0\sigma}} = \sqrt{\frac{1}{\pi f \mu_0 \sigma}} \tag{4.127}$$

The skin effect can be wisely exploited for shielding an electromagnetic wave. If a region with electromagnetic radiation is covered with a metal sheet of thickness much larger than the skin depth the fields outside the metal cover would be of negligibly small value.

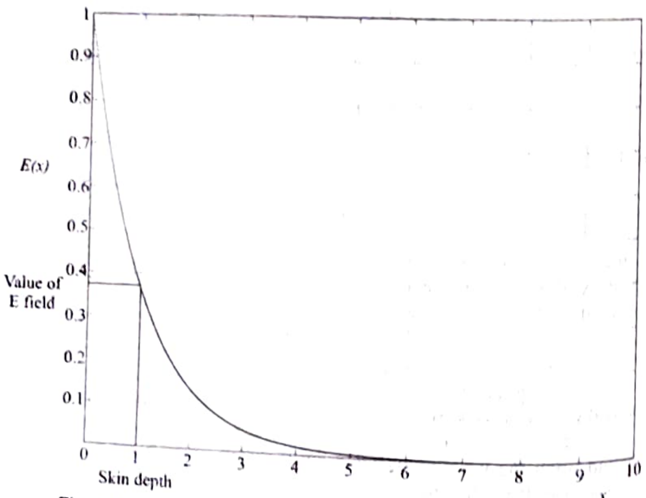


Fig. 4.19 *A decaying electrical field inside a good conductor.*

The skin effect which helps in isolating electromagnetic environments acts as a hinderance for the radio wave communication through the conducting walls. We all have noticed the poor reception of radio signals inside a train. The skin effect is the primary cause of that.

Table 4.1 shows the skin depth values for different conductors at a frequency of 1 MHz. The skin-depth is inversely proportional to the square-root of the frequency, and the square-root of the conductivity.

Table 4.1 Skin depth at 1 MHz

Material	Conductivity	Skin depth
Silver	6.17×10^7 ℧/m	63.87 μm
Copper	5.88×10^7 ℧/m	76.33 μm
Aluminium	3.65×10^7 ℧/m	83.17 μm
Doped Silicon	10^3 ℧/m	6.39 mm
Earth	5×10^{-3} ℧/m	7.12 m

The intrinsic impedance of a good conductor is

$$\eta_c \approx \sqrt{\frac{j\omega\mu_0}{\sigma}} = \sqrt{\frac{\omega\mu_0}{\sigma}}\angle 45^\circ \tag{4.128}$$

The magnetic field, therefore, lags the electric field by $\approx 45^\circ$ inside a good conductor.

EXAMPLE 4.10 A material has dielectric constant 25 and conductivity 2×10^6 ℧/m. What is the frequency above which the material cannot behave like a good conductor? If a plane wave of 10 MHz is incident on the material, effectively upto what depth can the wave penitrate the material, and what will be the wavelength of the wave inside the material?

Solution:

For any material to behave like a good conductor, we must have $\sigma >> \omega\epsilon_0\epsilon_r$. As a rule of thumb, we may say that when $\sigma > 10\omega\epsilon_0\epsilon_r$, the material is a good conductor. Therefore, the frequency above which the material does not behave like a good conductor, is

$$\omega = \frac{\sigma}{10\epsilon_0\epsilon_r} = \frac{2 \times 10^6}{10 \times \frac{1}{36\pi} \times 10^{-9} \times 25}$$

$$\Rightarrow \qquad f = 1.44 \times 10^{14} \ \text{Hz}.$$

Effective depth of wave penitration is the skin depth,

$$\delta = \frac{1}{\sqrt{\pi f \mu_0 \sigma}} = \frac{1}{\sqrt{\pi \times 10^7 \times 4\pi \times 10^{-7} \times 2 \times 10^6}} = 112.54 \ \mu\text{m}$$

To find the wavelength we first calculate the phase constant $\beta = \sqrt{\pi f \mu_0 \sigma} = 8885.5$ rad/m.
The wavelength $\lambda = \frac{2\pi}{\beta} = 0.707$ mm.

4.7 PHASE VELOCITY OF A WAVE

As we have seen above, the electric field of a plane wave travelling in $+z$ direction is written as

$$\mathbf{E}(z, t) = \mathbf{E}_0 e^{-\gamma z} \cdot e^{j\omega t} \tag{4.129}$$

For a general medium, γ is complex ($\gamma \equiv \alpha + j\beta$), and the electric field can be written as

$$\mathbf{E}(z, t) = E_0 e^{-\alpha z} \cdot e^{-j\beta z} \cdot e^{j\omega t} \tag{4.130}$$

$$= E_0 e^{-\alpha z} \cdot e^{j(\omega t - \beta z)} \tag{4.131}$$

As has been discussed in detail for the transmission lines, the phase ϕ of $\mathbf{E}$ is a composite function of space (z) and time (t). The constant phase point moves in $+z$ direction as a function of time. One can then ask a question, "with what speed does the constant phase point move in the space"

Let us consider an observer standing at some location in space and seeing the wave passing by him. Since $z = constant$ for the observer, he observes the phase of the wave to be ωt, i.e. the phase increases linearly as a function of time. Now, consider an observer who is holding on to a particular phase point. Obviously, he has to move with the speed with which the phase point is moving so that he never leaves the point. Then for this observer, the phase appears stationary (constant) as a function of time. The velocity of the wave is same as the velocity of the observer for whom the phase is constant. This velocity is called the 'phase velocity' of the wave. For a wave travelling in $+z$ direction, the phase of the wave is

$$\phi = \omega t - \beta z = constant \tag{4.132}$$

Differentiating Eqn (4.132) with respect to time we get

$$\omega - \beta \frac{\partial z}{\partial t} = 0 \tag{4.133}$$

$$\Rightarrow \quad \text{phase velocity} : v_p = \frac{\partial z}{\partial t} = \frac{\omega}{\beta} \tag{4.134}$$

Since β in general, is a function of the dielectric constant, ϵ_r of the medium, and the conductivity σ, the phase velocity also changes from medium to medium. In fact, the phase velocity is one of the parameters which characterizes the medium. Let us find the phase velocities of a wave in different media.

1. Free Space: For free space (or vacuum) we have $\epsilon = \epsilon_0 = \frac{1}{36\pi} \times 10^{-9}$ F/m $\mu = \mu_0 = 4\pi \times 10^{-7}$ H/m. We therefore have

$$\beta = \omega\sqrt{\mu_0\epsilon_0} = \frac{\omega}{3} \times 10^{-8} \text{ rad/m} \tag{4.135}$$

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu_0\epsilon_0}} = 3 \times 10^8 \text{ m/s} \tag{4.136}$$

This is the velocity of light in the free-space and is normally denoted by letter 'c'. One would appreciate the strength of the Maxwell's equations now which correctly predicted the wave phenomenon and provided correct estimate of the velocity of light in vacuum.

2. Pure Dielectric: For a pure dielectric we have $\sigma = 0$, $\epsilon = \epsilon_0\epsilon_r$, $\mu = \mu_0$

$$\Rightarrow \quad \beta = \omega\sqrt{\mu_0\epsilon_0\epsilon_r} = \omega\sqrt{\mu_0\epsilon_0}\sqrt{\epsilon_r} \tag{4.137}$$

$$\Rightarrow \quad v_p = \frac{\omega}{\omega\sqrt{\mu_0\epsilon_0}\sqrt{\epsilon_r}} = \frac{c}{\sqrt{\epsilon_r}} \tag{4.138}$$

Equation (4.138) gives the velocity of light in a medium with dielectric constant ϵ_r. From Eqn (4.138) we note that the velocity of an electromagnetic wave in a dielectric medium is always less than that in the vacuum and is independent of the frequency of the wave.

We may recall from our high school physics that the refractive index n of a medium is defined as

$$n = \frac{\text{velocity of light in vaccum}}{\text{velocity of light in the medium}} \tag{4.139}$$

Therefore, from Eqns (4.138) and (4.139) we get

$$n = \sqrt{\epsilon_r} \tag{4.140}$$

The refractive index of a dielectric material is the square root of its dielectric constant.

3. Dielectric Medium with Loss: For a lossy dielectric, $\sigma \neq 0$, and the phase constant is given as (from Eqn (4.119))

$$\beta = \omega\sqrt{\mu_0\epsilon_0\epsilon_r}\left\{1 + \frac{1}{8}\frac{\sigma^2}{\omega^2\epsilon^2}\right\} \tag{4.141}$$

hence the phase velocity is

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu_0\epsilon_0\epsilon_r}}\left\{1 + \frac{1}{8}\frac{\sigma^2}{\omega^2\epsilon^2}\right\}^{-1} = \frac{c}{\sqrt{\epsilon_r}}\left\{1 + \frac{1}{8}\frac{\sigma^2}{\omega^2\epsilon_0^2\epsilon_r^2}\right\}^{-1} \tag{4.142}$$

From Eqn (4.142) we note that for a lossy dielectric medium, the phase velocity is a function of ω, i.e. the waves of different frequencies travel with different velocities. This phenomenon is called the 'dispersion', and the medium is said to be a 'dispersive medium'. One can then say that a loss-less dielectric is non-dispersive, whereas a lossy medium is dispersive and the dispersion increases with the loss in the medium. For a low-loss medium however, the dispersion is generally small.

4. Good Conductor: For a good conductor $\sigma >> \omega\epsilon$ and the phase constant is (Eqn (4.126))

$$\beta = \sqrt{\frac{\omega\mu_0\sigma}{2}} \tag{4.143}$$

$$\Rightarrow \qquad v_p = \frac{\omega}{\sqrt{\frac{\omega\mu_0\sigma}{2}}} = \sqrt{\frac{2\omega}{\mu_0\sigma}} \tag{4.144}$$

Multiplying numerator and denominator within the square root sign of Eqn (4.144) by $\omega\epsilon_0$ and re-arranging we get

$$v_p = \frac{1}{\sqrt{\mu_0\epsilon_0}}\sqrt{\frac{2\omega\epsilon_0}{\sigma}} = c\sqrt{\frac{2\omega\epsilon_0}{\sigma}} \tag{4.145}$$

For a good conductor $\omega\epsilon_0/\sigma << 1$ and therefore $v_p << c$. The electromagnetic wave therefore slows down considerably in a conductor. As can be seen from the example, the phase velocity of an electromagnetic wave inside copper is few hundred meter/s which is of the order of the velocity of sound in copper.

Compared to a lossy dielectric, the conductor is much dispersive since v_p varies as $\sqrt{\omega}$. It should be noted however, that the dispersion decreases with σ and for an ideal conductor ($\sigma = \infty$) the dispersion is zero. Figures 4.20 and 4.21 show velocity and dispersion as a function of frequency and the conductivity of the medium.

EXAMPLE 4.11 Just outside a train compartment the field strength of a radio station is 0.1 V/m. What will be the approximate field strength inside the compartment? Assume the compartment to be a closed box of metal having conductivity 5×10^6 ℧/m. The thickness of the compartment wall is 5 mm. Frequency of radio station is 600 kHz.

Solution:

The attenuation constant for the compartment material (assuming the compartment material non-magnetic)

$$\alpha = \sqrt{\pi f \mu_0\sigma} = \sqrt{\pi \times 600 \times 10^3 \times 4\pi \times 10^{-7} \times 5 \times 10^6} = 3441.44 \text{ napers/m.}$$

The wave amplitude after passing through the compartment wall will be

$$E = E_0 e^{-\alpha t} = 0.1e^{-17.2} = 3.366 \times 10^{-9} \text{ V/m}$$

The wave, therefore, is attenuated by a factor $\sim 3 \times 10^7$.

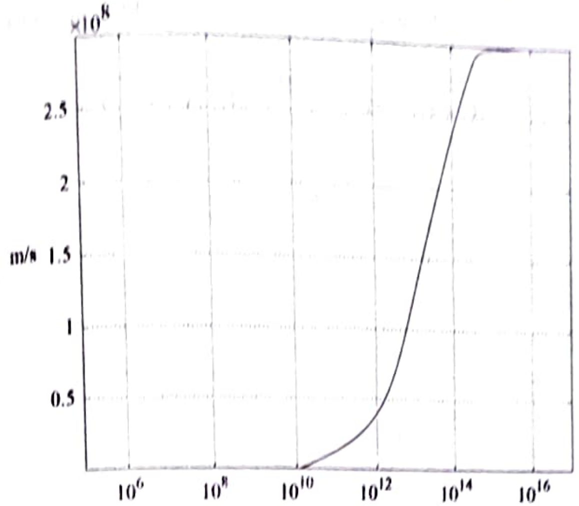


Fig. 4.20 *Variation of the phase velocity as a function of frequency.*

4.8 POWER FLOW AND POYNTING VECTOR

As seen above, the time varying electric and magnetic fields have to form an electromagnetic wave which propagates in the space. Naturally the wave carries some energy with it. It is then worthwhile to investigate, the quantity energy or power (i.e. energy per unit time) is carried by an electromagnetic wave? Note that in general, the electromagnetic wave need neither be a plane wave nor be travelling in an unbound medium. We, therefore, would like to develop a general frame work for power flow from arbitrary time varying fields. Since the electromagnetic phenomenon is completely governed by the Maxwell's equations, we again fall back upon the Maxwell's equations to find the power flow due to time varying fields.

Let us take the two Maxwell's equations

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\mu\frac{\partial \mathbf{H}}{\partial t} \tag{4.146}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J} + \epsilon\frac{\partial \mathbf{E}}{\partial t} \tag{4.147}$$

We assume here that μ, ϵ are not varying as a function of time. From the vector identity we have

$$\nabla \cdot (\mathbf{A} \times \mathbf{C}) = \mathbf{C} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot \nabla \times \mathbf{C} \tag{4.148}$$

where **A** and **C** are any two arbitrary vectors

Taking $\mathbf{A} = \mathbf{E}$ and $\mathbf{C} = \mathbf{H}$ we have the vector identity

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}) \tag{4.149}$$

Substituting for $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{H}$ from Eqns (4.146) and (4.147) in Eqn (4.149) we get

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (-\mu \frac{\partial \mathbf{H}}{\partial t}) - \mathbf{E} \cdot (\mathbf{J} + \epsilon \frac{\partial \mathbf{E}}{\partial t}) \tag{4.150}$$

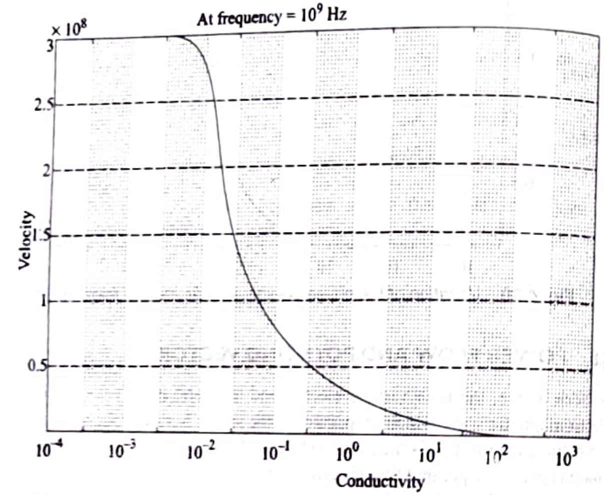


Fig. 4.21 *Variation of the phase velocity as a function of conductivity.*

Here we have assumed that the medium is isotropic and hence μ, ϵ and σ are scalar quantities. Also note that, for any two vectors **A** and **C** we have

$$\frac{\partial(\mathbf{A} \cdot \mathbf{C})}{\partial t} = \mathbf{A} \cdot \frac{\partial \mathbf{C}}{\partial t} + \mathbf{C} \cdot \frac{\partial \mathbf{A}}{\partial t} \tag{4.151}$$

$$\Rightarrow \frac{\partial(\mathbf{A} \cdot \mathbf{A})}{\partial t} = 2\mathbf{A} \cdot \frac{\partial \mathbf{A}}{\partial t} \tag{4.152}$$

$$\Rightarrow \mathbf{A} \cdot \frac{\partial \mathbf{A}}{\partial t} = \frac{1}{2}\frac{\partial(\mathbf{A} \cdot \mathbf{A})}{\partial t} = \frac{1}{2}\frac{\partial |\mathbf{A}|^2}{\partial t} \tag{4.153}$$

Noting that $\mathbf{J} = \sigma\mathbf{E}$, and taking σ not a function of time, we can write from Eqn (4.150)

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\frac{\mu}{2}\frac{\partial |\mathbf{H}|^2}{\partial t} - \frac{\epsilon}{2}\frac{\partial |\mathbf{E}|^2}{\partial t} - \sigma |\mathbf{E}|^2 \tag{4.154}$$

The Eqn (4.154) is essentially a point relation. That is, it should be valid at every point in the space at every instant of time.

If we now integrate Eqn (4.154) over a volume we get

$$\oint_V \nabla \cdot (\mathbf{E} \times \mathbf{H}) dv = \oint_V (-\frac{\mu}{2}\frac{\partial |\mathbf{H}|^2}{\partial t} - \frac{\epsilon}{2}\frac{\partial |\mathbf{E}|^2}{\partial t} - \sigma |\mathbf{E}|^2) dv \tag{4.155}$$

Application of the divergence theorem (see Eqn (3.16)) on the LHS yields

$$\oint_S (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{a} = \oint_V -\frac{\mu}{2}\frac{\partial |\mathbf{H}|^2}{\partial t} dv - \oint_V \frac{\epsilon}{2}\frac{\partial |\mathbf{E}|^2}{\partial t} dv - \oint_V \sigma |\mathbf{E}|^2 dv \tag{4.156}$$

Taking the volume V constant as a function of time, we can interchange the integral and the $\frac{\partial}{\partial t}$ giving

$$\oint_S (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{a} = -\frac{\partial}{\partial t}\oint_V \frac{\mu}{2}|H|^2 dv - \frac{\partial}{\partial t}\oint_V \frac{\epsilon}{2}|\mathbf{E}|^2 dv - \oint_V \sigma |E|^2 dv \tag{4.157}$$

$$= -\frac{\partial}{\partial t}\left(\oint_V \frac{\mu}{2}|H|^2 dv + \oint_V \frac{\epsilon}{2}|E|^2 dv\right) - \oint_V \sigma |E|^2 dv \tag{4.158}$$

In Eqn (4.158), the first term within the brackets gives the magnetic energy stored in the volume, whereas the second term represents the electric energy stored in the volume V. The quantity within the brackets therefore represents the total energy stored in the volume, and the first term on the RHS of Eqn (4.158) represents the rate of decrease of total energy stored in the volume V, i.e. decrease in power in the volume V.

The second term on the RHS gives the ohmic power loss in the volume V. This term then represents the electromagnetic energy converted to heat per unit time. The two terms on the RHS consequently represent the total decrease in the electromagnetic energy per unit time, i.e. the power loss from the volume. Since there are no other losses in the medium, the law of conservation of energy dictates that this power loss should be due to energy leaving the volume. The LHS of Eqn (4.158) hence gives the power coming out of the volume, i.e.

$$\text{Net Outward Power } W = \oint_S (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{a} \tag{4.159}$$

Since the surface integral of $(\mathbf{E} \times \mathbf{H})$ gives the total power flow from the surface, the quantity $(\mathbf{E} \times \mathbf{H})$, therefore, represents the power density on the surface of the volume. The vector defined as

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} \tag{4.160}$$

is called the 'Poynting vector' and it gives the power flow density (power flow per unit area) at any point on the surface.

The statement that 'the surface integral of **P** over a closed surface (the LHS in Eqn (4.158)) is equal to the total power leaving the closed surface' is called the *Poynting theorem*. It should be emphasized however, that the Poynting theorem is strictly valid only for a closed surface. That is to say that the Eqn (4.160) is not a point relationship according to the Poynting theorem. The use of (**E** × **H**) as power density at every point on the closed surface is arbitrary. There are certain classical cases which contradict the use of point relationship for the Poynting vector. For example, let us place a charge at one of the poles of a bar magnet as shown in Fig. 4.22. Now, let us take some point, say P, as shown in Fig. 4.22. At this point since **E** and **H** are not parallel, we have finite **E** × **H** which is pointing out of the plane of the paper. In other words, we have a power coming out of the paper at point P. Obviously, since in this case there is no sustained power flow, the answer obtained from the Poynting vector is absurd. If we consider however, a symmetrically located point P', the **E** × **H** vector here is of same magnitude as that at P but going inside the plane of the paper. The sum of the power flow at P and P' hence is zero. Extending the result to any closed surface around the magnet we find that the net power flow from the surface is zero even though the poynting vector **P** is non-zero on the surface. It might, therefore, appear that the use of **E** × **H** for power density at any location might give erroneous results at times. However, except for few pathological cases (like the one discussed above) the Poynting vector correctly gives the power density at a point in the space. The Eqn (4.160) hence is normally used as a point relation although it is not strictly a point relation.

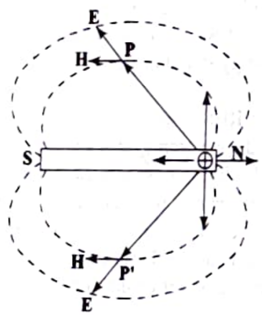


Fig. 4.22 *Poynting Vector for a system consisting of a charge placed on a bar magnet.*

The Poynting vector is a rather useful concept as it tells that the power flow is perpendicular to both the electric and the magnetic fields. The Poynting vector has the unit of Watts/m^2.

For a uniform plane wave travelling in $+z$ direction, the electric and the magnetic fields are oriented in $+x$ and $+y$ directions respectively. The $(\mathbf{E} \times \mathbf{H}) = (\hat{\mathbf{x}} \times \hat{\mathbf{y}})$ vector then is along $+z$ direction, i.e. the direction of the wave propagation. The Poynting vector hence correctly gives the direction of the power flow as the direction of the wave propagation.

EXAMPLE 4.12 A uniform plane wave has a power density of 20 W/m^2 and is travelling along $-y$ direction in the free space. If the electric field makes an angle of 30° with the $+x$ axis, find electric and magnetic field strengths and the direction of the magnetic field.

Solution:

For a uniform plane wave E and H are perpendicular to each other. Therefore, the poynting vector is

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} = |\mathbf{E}||\mathbf{H}|$$

$$= \frac{|\mathbf{E}|^2}{\eta_0} \qquad \text{since} \quad |\mathbf{H}| = \frac{|\mathbf{E}|}{\eta_0}$$

$$\Rightarrow \qquad |E| = \sqrt{P\eta_0} = \sqrt{20 \times 120\pi} = 86.83 \text{ V/m}$$

$$\text{and} \qquad |H| = \frac{E}{120\pi} = 0.23 \text{ A/m}$$

Since, **E** and **H** are perpendicular to the direction of the wave, $-y$ axis, they must lie in xz plane. Also **E** × **H** should be in the direction of the wave. The **H** vector therefore will make an angle of 30° with the z-axis (see Fig. 4.23).

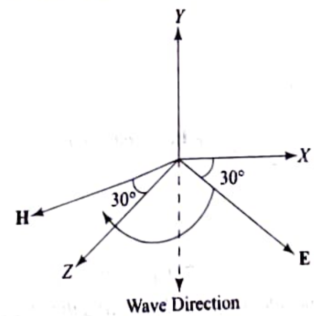


Fig. 4.23

4.8.1 Instantaneous and Average Poynting Vector

For time harmonic fields, it is rather useful to have the average power density (average over time) or average Poynting vector at any point in space. In the following sections, starting with the instantaneous **E**, **H** and **P** we will derive the average power density or the average Poynting vector.

Writing **E** and **H** explicitly for the time harmonic function we have

$$\mathbf{E}(x, y, z, t) = \mathbf{E}_0(x, y, z)e^{j\omega t} \tag{4.161}$$

and

$$\mathbf{H}(x, y, z, t) = \mathbf{H}_0(x, y, z)e^{j\omega t} \tag{4.162}$$

The instantaneous values of the electric and the magnetic fields are

$$\mathbf{E}(x, y, z, t) = Re\{\mathbf{E}_0(x, y, z)e^{j\omega t}\} \tag{4.163}$$

$$= Re\{E_0(x, y, z)e^{j\phi_e}e^{j\omega t}\}\hat{\mathbf{e}} \tag{4.164}$$

$$= E_0(x, y, z)\cos(\phi_e + \omega t)\hat{\mathbf{e}} \tag{4.165}$$

and

$$\mathbf{H}(x, y, z, t) = Re\{\mathbf{H}_0(x, y, z)e^{j\omega t}\} \tag{4.166}$$

$$= Re\{H_0(x, y, z)e^{j\phi_h}e^{j\omega t}\}\hat{\mathbf{h}} \tag{4.167}$$

$$= H_0(x, y, z)\cos(\phi_h + \omega t)\hat{\mathbf{h}} \tag{4.168}$$

where $\hat{\mathbf{e}}$ and $\hat{\mathbf{h}}$ are the unit vectors in the directions of **E** and **H** respectively and ϕ_e and ϕ_h are the time phases of the electric and the magnetic fields respectively at a point (x, y, z) in the space (see Fig. 4.24). The instantaneous Poynting vector is

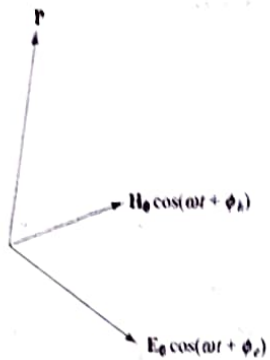


Fig. 4.24 Electric and magnetic fields, and the Poynting vector.

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} = E_0H_0\cos(\omega t + \phi_e)\cos(\omega t + \phi_h)(\hat{\mathbf{e}} \times \hat{\mathbf{h}}) \tag{4.169}$$

Equation (4.169) can be re-written as

$$\mathbf{P} = \frac{E_0H_0}{2}\{\cos(\phi_e - \phi_h)\hat{\mathbf{e}} \times \hat{\mathbf{h}} + \cos(2\omega t + \phi_e + \phi_h)\hat{\mathbf{e}} \times \hat{\mathbf{h}}\} \tag{4.170}$$

The instantaneous power density harmonically varies at a frequency 2ω, double of that of **E** or **H**. The average power density

$$\mathbf{P}_{av} = \frac{1}{T}\int_0^T \mathbf{P}dt \tag{4.171}$$

where T is the time period of the time harmonic fields, $T = 2\pi/\omega$. Substituting Eqn (4.170) in Eqn (4.171) and noting that the average of the second term is zero, we get

$$\mathbf{P}_{av} = \frac{1}{2}E_0H_0\cos(\phi_e - \phi_h)\hat{\mathbf{e}} \times \hat{\mathbf{h}} \tag{4.172}$$

$$= \frac{1}{2}Re\left(\{E_0e^{j\phi_e}e^{j\omega t}\hat{\mathbf{e}}\} \times \{H_0e^{-j\phi_h}e^{-j\omega t}\hat{\mathbf{h}}\}\right) \tag{4.173}$$

$$= \frac{1}{2}Re\left(\{E_0e^{j\phi_e + j\omega t}\hat{\mathbf{e}}\} \times \{H_0e^{j\phi_h + j\omega t}\hat{\mathbf{h}}\}^*\right) \tag{4.174}$$

$$\Rightarrow \quad \mathbf{P}_{av} = \frac{1}{2}Re(\mathbf{E} \times \mathbf{H}^*) \tag{4.175}$$

The average power density is a much meaningful quantity as it gives the actual power flow at that location. The instantaneous power on the other hand, does not correctly represent the power flow as it can be negative or positive. It can probably then give the amount of power oscillating back and forth around a point plus the actual power flow at that point.

From Eqns (4.172) and (4.175) it is clear that for a real power flow, two conditions should be satisfied.

1. **E** and **H** fields should cross each other
2. **E** and **H** should not be in time quardrature, i.e. 90° out of phase with each other.

E and **H** fields which are parallel and/or in time quardrature, do not constitute any power flow. In the later section, we will observe that in complex situations, this argument comes very handy in checking whether the fields carry any power along with them.

EXAMPLE 4.13 In a region the **E** and **H** fields are gives as

$$\mathbf{E} = 100(j\hat{\mathbf{x}} + 2\hat{\mathbf{y}} - j\hat{\mathbf{z}})e^{j\omega t}$$

$$\mathbf{H} = (-\hat{\mathbf{x}} + j\hat{\mathbf{y}} + \hat{\mathbf{z}})e^{j\omega t}$$

Find the average power flow density and direction of power flow in the region.

Solution:

The average poynting vector

$$\mathbf{P}_{av} = \frac{1}{2}Re\{\mathbf{E} \times \mathbf{H}^*\}$$

$$= \frac{100}{2}Re\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ j & 2 & -j \\ -1 & -j & 1 \end{vmatrix}$$

$$= 150\sqrt{2}\frac{(\hat{\mathbf{x}} + \hat{\mathbf{z}})}{\sqrt{2}}$$

The average power density is $150\sqrt{2}\ \mathrm{W/m^2} = 212.132\ \mathrm{W/m^2}$ and the power flows in the direction $\frac{(\hat{x}+\hat{z})}{\sqrt{2}}$, i.e. in the xz plane at an angle of 45° with respect to x and z axes.

4.8.2 Power Density of a Uniform Plane Wave

For a uniform plane wave, **E** and **H** are perpendicular to each other and the ratio of their magnitudes is equal to the intrinsic impedance of the medium, η. Without losing generality let us take the E-field oriented along the $+x$ direction and the H-field oriented along the $+y$ direction as

$$\mathbf{E} = E_0 e^{j\omega t}\hat{\mathbf{x}} \tag{4.176}$$

$$\mathbf{H} = H_0 e^{j\omega t}\hat{\mathbf{y}} = \frac{\mathbf{E_0}}{\eta}e^{j\omega t}\hat{\mathbf{y}} \tag{4.177}$$

The average power density of the wave is

$$\mathbf{P_{av}} = \frac{1}{2}Re\{\mathbf{E}\times\mathbf{H}^*\} \tag{4.178}$$

$$= \frac{1}{2}Re\{E_0 e^{j\omega t}[\frac{E_0}{\eta}e^{j\omega t}]^*\}\hat{\mathbf{x}}\times\hat{\mathbf{y}} \tag{4.179}$$

$$= \frac{1}{2}Re\{\frac{|E_0|^2}{\eta^*}\}\hat{\mathbf{z}} = \frac{1}{2}Re\{\eta|H_0|^2\}\hat{\mathbf{z}} \tag{4.180}$$

$$\Rightarrow \quad \boxed{P_{av} = \frac{|E_0|^2}{2}Re\{\frac{1}{\eta^*}\} \equiv \frac{|H_0|^2}{2}Re\{\eta\}} \tag{4.181}$$

(a) For a **loss-less dielectric medium** $\eta = \sqrt{\frac{\mu}{\epsilon}}$ = Real number
The average power density of the wave is

$$\boxed{P_{av} = \frac{1}{2}\frac{|E_0|^2}{\eta} = \frac{1}{2}|E_0|^2\sqrt{\frac{\epsilon}{\mu}}} \tag{4.182}$$

(b) For a **lossy medium** however, η is complex, and **E** and **H** are not in time phase. Consequently, P_{av} has to be calculated using exact Eqn (4.181).

(c) For a **good conductor** $\sigma >> \omega\epsilon$, and $\eta \approx \sqrt{\frac{\omega\mu}{2\sigma}} + j\sqrt{\frac{\omega\mu}{2\sigma}}$. The phase angle between **E** and **H** is approximately 45°, and the average power density is

$$P_{av} = \frac{1}{2}|E_0|^2 Re\{\frac{1}{\eta^*}\} = \frac{1}{2}\frac{|E_0|^2}{|\eta|^2}Re\{\eta\} \tag{4.183}$$

$$= \frac{1}{2}|E_0|^2\sqrt{\frac{\sigma}{2\omega\mu}} \tag{4.184}$$

It is interesting to note from Eqn (4.184) that as the frequency increases, the power density of the wave reduces and at very high frequency very little power penitrates the conducting medium.

EXAMPLE 4.14 At some location inside a lossy dieletric material the measured peak electric field of a wave is 10 V/m. The material has relative permittivity of 8 and conductivity of 100 ℧/m. Find the average power density of the wave at that location. Also find the power density at a distance of 1 cm in the direction of the wave propagation. The frequency of the wave is 300 MHz.

Solution:

$$\omega = 2\pi\times 300\times 10^6 = 1.8849\times 10^9\ \mathrm{rad/sec}$$

The intrinsic impedance of the medium is

$$\eta = \sqrt{\frac{j\omega\mu_0}{\sigma + j\omega\epsilon_0\epsilon_r}} = 3.44 + j3.44\ \Omega$$

From Eqn (4.181), the average power density of the wave is

$$P_{av} = \frac{|E_0|^2}{2}Re\{\frac{1}{\eta^*}\} = 7.27\ \mathrm{W/m^2}$$

Now, since the conductivity of the medium is non-zero, there is attenuation in the medium. The propagation constant of the medium is

$$\gamma = \alpha + j\beta = \sqrt{j\omega\mu_0(\sigma + j\omega\epsilon_0\epsilon_r)}$$
$$= 343.9 + j344.36 \text{ per meter}$$

$\Rightarrow$ Attenuation constant α =343.9 nepers/m

Power density at 1 cm distance is

$$P_{av}e^{-2\times 0.01\times\alpha} = P_{av}e^{-6.878} = 7.49\ \mathrm{mW/m^2}$$

4.9 SURFACE CURRENT AND POWER LOSS IN A CONDUCTOR

4.9.1 Surface Current

We have seen in the previous chapter that the tangential component of the electric field is zero at the surface of an ideal conductor. The tangential magnetic field is balanced by the surface current J_s giving,

$$\mathbf{J_s} = \hat{\mathbf{n}}\times\mathbf{H} = \mathbf{H_t} \tag{4.185}$$

where $\hat{\mathbf{n}}$ is the outward normal unit vector to the conductor surface and **H** is the total magnetic field at the surface. $\mathbf{H_t}$ denotes the tangential component of the magnetic field.

Inside an ideal conductor no time varying fields exist and hence the current is truely the surface current. One may wonder at this point, as to the force that drives this current, which needs electric field to move charges. Since the electric field along the conducting surface is zero, there is no force for driving the current. This is however, a steady state picture. To excite the current, some electric field must have been present momentarily. The electric field would have put the charges into motion and must have disappeared. Since the conductivity is infinite for an ideal conductor, once the charges are placed in motion, the current keeps flowing for infinite time without the electric force. It is, therefore, important that one clearly distinguishes between the two cases (i) no field and no current and (ii) no field but current. In steady state both cases have no electric field along the conductor surface.

The surface current is a concept of an ideal conductor. As we do not have ideal conductors in real life, one would wonder regarding the purpose of the concept of surface current. In the following sections we will observe that although there is nothing like surface current for a non-ideal conductor ($\sigma \neq \infty$), the concept can still be useful in analysing good conductors.

Let us consider an electric field $\mathbf{E_0}$ tangential to the conducting surface (along x-direction) and just inside it. It should be made clear that $\mathbf{E_0}$ is not the incident field but the field which would exist at the surface (why the field is not the same as the incident field will become clear later after we have discussed the reflection from the conducting surface). Since we are considering good conductors, the magnetic field $\mathbf{H_0}$ is almost tangential to the surface (along y-axis) and the ratio $|\mathbf{E_0}|/|\mathbf{H_0}|$ is the intrinsic impedance of the conductor η_c. The $\mathbf{E_0}$ and $\mathbf{H_0}$ constitute a wave going inside the conductor as shown in Fig. 4.25.

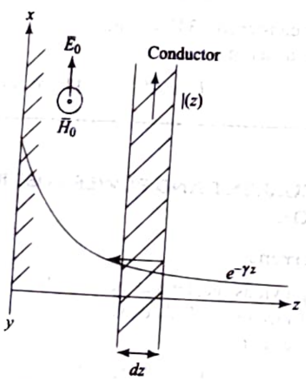


Fig. 4.25 *Electrical field variation inside a conductor.*

The propagation constant of the wave inside the conductor is

$$\gamma = \alpha + j\beta = \sqrt{\frac{\omega\mu_0\sigma}{2}} + j\sqrt{\frac{\omega\mu_0\sigma}{2}} \tag{4.186}$$

The field amplitude decreases exponentially inside the conductor. The field at a depth z from the surface then is

$$\mathbf{E}(z) = \mathbf{E_0}e^{-\gamma z} = \mathbf{E_0}e^{-\alpha z}.e^{-j\beta z} \tag{4.187}$$

Now, due to conductivity σ of the conductor, a conduction current flows in the same direction as the electric field $\mathbf{E}$. The density of the conduction current at a depth z is

$$\mathbf{J}(z) = \sigma\mathbf{E} = \sigma E_0 e^{-\alpha z}e^{-j\beta z}\hat{\mathbf{x}} \tag{4.188}$$

Let us now consider a slab parallel to the surface having unit width along y-direction, and thickness dz. The sheet is located at a distance of z from the surface. The current in the sheet can be written as

$$\mathbf{I}(z) = \mathbf{J}(z)dz = \sigma E_0 e^{-\gamma z}dz\hat{\mathbf{x}} \tag{4.189}$$

The current flows along x-direction (same as that of the electric field). If we now integrate $\mathbf{I}(z)$ along z, we get total current flow per unit width of the conductor surface (the width is perpendicular to the direction of the current and is along y-direction) as

$$\mathbf{J_s} = \int_0^\infty \mathbf{E_0}\sigma e^{-\gamma z}dz = \mathbf{E_0}\sigma\left[-\frac{e^{-\gamma z}}{\gamma}\right]_0^\infty \tag{4.190}$$

Since γ has +ve real part, $e^{-\gamma\infty} \to 0$ giving

$$\mathbf{J_s} = \frac{E_0\sigma}{\gamma}\hat{\mathbf{x}} \tag{4.191}$$

For a good conductor, since the field decays very rapidly inside the conductor, the current $\mathbf{J_s}$ effectively flows close to the surface, and we may treat this current as the surface current. Note that dimensionally $\mathbf{J_s}$ has dimensions Ampere/meter, which is same as the dimension of the surface current. To completely justify that $\mathbf{J_s}$ is equivalent to the surface current, it must be related to the tangential magnetic field $\mathbf{H_0}$ through Eqn (4.185). For the conductor, the intrinsic impedance is

$$\eta_c = \sqrt{\frac{j\omega\mu}{\sigma}} = \frac{\gamma}{\sigma} \tag{4.192}$$

since for a good conductor the propagation constant $\gamma = \sqrt{j\omega\mu\sigma}$.

The magnetic field at the surface of the conductor is

$$|\mathbf{H_0}| = \frac{|\mathbf{E_0}|}{|\eta_c|} = \frac{|\mathbf{E_0}|}{|\gamma|}\sigma = |\mathbf{J_s}| \tag{4.193}$$

From Fig. 4.25, we see that the normal vector $\hat{\mathbf{n}}$ to the conductor surface is in $-z$ direction, and the magnetic field is in y-direction. We, therefore, get

$$\hat{\mathbf{n}} \times \mathbf{H_0} = -\hat{\mathbf{z}} \times \hat{\mathbf{y}}|H_0| = \mathbf{J}_s \tag{4.194}$$

Hence, the Eqn (4.185) is indeed satisfied by $\mathbf{J}_s$ given by Eqn (4.191).

It is important to emphasize that for a non-ideal conductor, there is only a volume current density $\mathbf{J}$ and truly there is no surface current. However, the total integrated current $\mathbf{J}_s$ can be treated like the surface current. One can observe that for an ideal conductor, as $\sigma \to \infty$, the skin depth tends to be zero, and the current $\mathbf{J}_s$ truly becomes the surface current.

From Eqn (4.193), we can define a parameter called the surface impedance Z_s which is the ratio of the tangential electric field E_0 and the surface current J_s as

$$Z_s = \frac{|\mathbf{E}_{\tan}|}{|\mathbf{J}_s|} = \frac{E_0}{J_s} = \frac{\gamma}{\sigma} = \eta_c \tag{4.195}$$

Seperation of real and imaginary parts yields

$$Z_s = R_s + jX_s = \sqrt{\frac{\omega\mu_0}{2\sigma}} + j\sqrt{\frac{\omega\mu_0}{2\sigma}} \tag{4.196}$$

As will be seen later the surface impedance is a useful parameter in computation of the conductor losses.

4.9.2 Power Loss in a Conductor

Let us again consider the thin sheet in Fig. 4.25. Take a piece of this sheet which has unit length and unit width as shown in Fig. 4.26. If we treat this slab as a resistor of resistivity $\rho = 1/\sigma$, the area of cross-section of the resistor is $A = 1 \times dz = dz$ and the length of the resistor is $l = 1$.

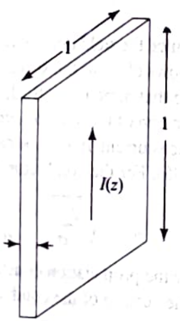


Fig. 4.26 *Power Loss in a thin sheet.*

The resistance of the slab is

$$dR = \frac{\rho l}{A} = \frac{1}{\sigma dz} \tag{4.197}$$

The ohmic loss in the slab is

$$dW = \frac{1}{2}|I(z)|^2 dR \tag{4.198}$$

Substituting for $I(z)$ from Eqn (4.189) we get

$$dW = \frac{1}{2}|\sigma E_0 e^{-\gamma z} dz|^2 \frac{1}{\sigma dz} \tag{4.199}$$

$$= \frac{1}{2}\sigma |E_0|^2 e^{-2\alpha z} dz \tag{4.200}$$

The total loss per unit area of the conductor surface can be obtained by integrating Eqn (4.200) from $z = 0$ to ∞ as

$$W = \frac{1}{2}\int_0^\infty \sigma |E_0|^2 e^{-2\alpha z} dz \tag{4.201}$$

$$= \frac{1}{2}\sigma |E_0|^2 \left[\frac{e^{-2\alpha z}}{-2\alpha}\right]_0^\infty \tag{4.202}$$

$$\Rightarrow \quad W = \frac{1}{2}\frac{\sigma |E_0|^2}{2\alpha} = \frac{1}{2}\frac{\sigma}{2\alpha}\frac{|\gamma|^2}{\sigma^2}|\mathbf{J}_s|^2 \tag{4.203}$$

Substituting for γ and α from Eqn (4.186) we get

$$W = \frac{1}{2}\frac{|\mathbf{J}_s|^2 \omega\mu_0\sigma}{2\sigma\sqrt{\omega\mu_0\sigma/2}} \tag{4.204}$$

$$= \frac{1}{2}|\mathbf{J}_s|^2\sqrt{\frac{\omega\mu_0}{2\sigma}} \tag{4.205}$$

$$= \frac{1}{2}R_s|\mathbf{J}_s|^2 \tag{4.206}$$

The power loss, therefore, is proportional to the surface resistance ($R_s = \sqrt{\omega\mu_0/2\sigma}$) which increases with frequency and decreases with the conductivity. It is then interesting to see that as the conductivity increases, the wave attenuates rapidly inside the conductor but this attenuation is not due to the ohmic loss. This case, therefore, is not similar to a lossy transmission line where the power is lost in the heating of the line due to ohmic loss. This means that, as the conductivity increases, the energy finds it difficult to enter the conducting surface. For ideal conductor, i.e. for $\sigma = \infty$, there is no penitration of the wave. The current flows only on the conductor surface and there is no power loss.

EXAMPLE 4.15 The magnetic field at the surface of a good conductor is 2 A/m. The frequency of the field is 600 MHz. If the conductivity of the conductor is 10^7 ℧/m, find the skin depth, surface impedance, and the power loss per unit area of the conductor.

Solution:

$$\omega = 2\pi f = 2\pi \times 600 \times 10^6 = 3.7698 \times 10^9 \text{ rad/s}$$

The skin depth

$$\delta = \frac{1}{\sqrt{\pi f \mu_0 \sigma}} = 6.5 \ \mu\text{m}$$

Surface impedance

$$Z_s = \sqrt{\frac{j\omega\mu_0}{\sigma}} = (15.3 + j15.3) \times 10^{-3} \ \Omega$$

The surface r.m.s. current density $|J_s| = |H| = 2$ A/m.
Therefore, Power loss per unit area

$$W = R_s|J_s|^2 = 15.3 \times 10^{-3} \times (2)^2 = 61.2 \quad \text{mW/m}^2$$

Note: The factor 1/2 is not used since the surface current has r.m.s. value of 2 A/m.

4.10 SUMMARY

In this chapter the basic characteristics of the uniform plane waves have been investigated. Although plane waves strictly exist only in an unbound medium, in practice the plane wave model is adequate in situations where the size of the medium is much larger than the wavelength. Starting from the Maxwell's equations it is shown that in Cartesian coordinate system the simplest phenomenon consistent with the Maxwell's equations is the uniform plane waves. The uniform plane wave is transverse in nature, and the electric and magnetic fields are related through the medium parameter, called the intrinsic impedance of the medium. An important characteristic called the wave polarization is investigated further. Concept of complex permittivity is introduced for dielectric media having finite conductivity. Plane wave propagation in dielectrics and conductors has been analysed and the concept of the skin depth has been introduced. Very important concepts of wave propagation like the phase velocity and the Poynting vector are introduced in this chapter. The chapter closes with discussion on the power loss in non-ideal conductors.

Review Questions

4.1 What is the meaning of source-free medium?

4.2 In what situation can μ and ϵ be taken outside the ∇ operator in Eqns (4.7) and (4.9)

4.3 If ϵ and μ are functions of space, what will be the wave equation?

4.4 Under what condition does a uniform electric or magnetic field will exist in a three dimensional space?

4.5 What is the simplest form of time harmonic field which can exist in an unbound medium ?

4.6 What is phase constant of a plane wave?

4.7 How does the phase constant vary with the frequency for a uniform plane wave?

4.8 What is intrinsic impedance of a medium? What is its significance?

4.9 What is the value of the characteristic impedance for the free space?

4.10 What is wave polarization?

4.11 Why is it not required to define magnetic field along with the electric field for a uniform plane wave?

4.12 What is linear, circular and elliptical polarization?

4.13 Under what conditions do two linear polarizations generate a circular polarization?

4.14 Under what conditions do two circular polarizations generate a linear polarization?

4.15 In a circularly polarized wave, how many rotations does the electric field vector make in a second?

4.16 What is the state of polarization?

4.17 What is Poincare' sphere of polarization and what is its utility?

4.18 What are orthogonal states of polarization?

4.19 By using combination of a linear and LC polarizations, what states of polarization can be generated?

4.20 Why is circular polarization preferred for satellite communication?

4.21 Can we use two elliptical polarizations to get circular polarization? How?

4.22 What is the meaning of complex permittivity?

4.23 Can a medium be a conductor at one frequency and dielectric at other?

4.24 Why does the earth which has very poor conductivity behave like a conductor at low frequencies.

4.25 Verify that copper is a good conductor even upto X-ray frequencies.

4.26 How does the attenuation constant vary as a function of conductivity of a non ideal dielectric medium?

4.27 What is loss tangent?

4.28 What is loss tangent for an ideal dielectric?

4.29 What is the phase difference between the electric and the magnetic fields of a plane wave in an ideal dielectric and in a lossy dielectric medium?

4.30 What is the phase of the intrinsic impedance of a good conductor?

4.31 What does the complex nature of intrinsic impedance of a medium mean?

4.32 What is skin depth?

4.33 What is phase velocity?

4.34 What is refractive index of a medium?

4.35 What happens to the speed of a wave in a dielectric medium compared to that in vaccum?

4.36 What is dispersion?

4.37 Can a loss-less medium be dispersive?

4.38 Can a lossy medium be non-dispersive?

4.39 What is Poynting theorem and Poynting vector?

4.40 Does Poynting vector always give correct measure of power flow at a point?

4.41 What are the essential conditions to have a power flow at a point in space?

4.42 What is the power density of a uniform plane wave?

4.43 What is surface current?

4.44 What is the driving force for the surface current?

4.45 Do we see surface currents in practice? Why?

4.46 What is the unit of surface current density?

4.47 How is power loss related to surface current?

4.48 How does the power loss on the surface of a good conductor vary with frequency and conductivity?

Problems

4.1 Find the characteristic impedance of a medium having $\epsilon_r = 10$ and $\mu_r = 25$. If a plane wave travelling along the $+x$-direction has 5 V/m $-z$-oriented electric field, find the magnitude and direction of the magnetic field of the wave.

4.2 The magnetic field of a plane wave in a dielectric medium is given as

$$\mathbf{H} = 5\hat{y} - 8\hat{z}$$

If the phase constant of the wave at 1 GHz is 25 rad/m, find the vector electric field of the wave.

4.3 For a uniform plane wave in a medium, the electric and magnetic fields are given as

$$\mathbf{E} = 5\hat{x} + 2\hat{y} + 6\hat{z}\ \text{V/m}$$

$$\mathbf{H} = \frac{1}{120\pi}(2\hat{x} - 2\hat{y} - \hat{z})\ \text{A/m}$$

Find the direction of the wave and the intrinsic impedance of the medium.

4.4 For a uniform plane wave travelling in $+z$-direction, the x and y components of the electric fields are $2\angle 0$ V/m and $5\angle 45^\circ$ V/m respectively. Find the equation of the ellipse of polarization.

4.5 For the ellipse of polarization in Problem 4.4 find the major and minor axes of the ellipse. Also find the angle which the major axis of the ellipse makes with the $+x$ direction.

4.6 For a wave travelling in $+y$ direction, the x-component of the electric field leads the z-component by 60°. If the two components have equal amplitudes, find the AR and τ for the ellipse of polarization. The tilt angle τ is measured from the $+x$ axis towards the $+z$ axis.

4.7 For a wave travelling in $+z$ direction, the electric field is given as

$$\mathbf{E} = (3 + j4)\hat{x} + (5 - j6)\hat{y}$$

Find the maximum value of the electric field.

4.8 For the ellipse of polarization in Problem 4.7, find the latitude and longitude of the point on the Poincare' sphere.

4.9 An elliptically polarized wave is to be transmitted along the $+z$ direction. The axial ratio of the ellipse is 2 and the tilt angle is 30°. Find the amplitudes and the phases of the two linearly polarized waves which can generate the desired wave.

4.10 An RH elliptically polarized wave with $AR = 4$ and $\tau = 40^\circ$ is to be genetared by combining two circularly polarized waves. Find the parameters of the circularly polarized waves.

4.11 An elliptically polarized wave with $AR = 3.5$ is received by a linearly polarized antenna aligned with the major axis of the ellipse. Find the power transfer efficiency from the wave to the antenna.

4.12 A horizontal linear and a LH circular polaization antenna are simultaneously excited. The peak electric field generated by each antenna is 1 V/m. At an instant when the electric field due to the circularly polarized antenna is along $+y$ direction, the electric field due to the linear antenna is passing through zero but rising towards $+x$ direction. Draw the polarization ellipse for the combined wave. Find the state of polarization of the wave and its orthogonal state. The waves are travelling in $+z$ direction.

4.13 Two identical linearly polarized antennas lie in the xy-plane. Antenna 1 along the x-axis and antenna 2 along $\phi = \phi_0$ line. The two antennas are excited with $1\angle 0$ and $A\angle\theta$ respectively. Find A and θ required to generate LHC and RHC polarized waves travelling in the $+z$-direction.

4.14 For a material $\epsilon_r = 10$ and $\sigma = 10^2$ ℧/m. Upto what frequency can the material be treated as a good conductor?

4.15 For the material in Problem 4.14, find the complex dielectric constant of the medium, and the loss-tangent at 10 GHz.

4.16 Find attenuation and phase constants of a plane wave in the material in Problem 4.14, at 6 GHz.

4.17 For the sea water the dielectric constant is 80 and the conductivity is 10^{-2} ℧/m. Find the wavelength of the 100 MHz signal and the distance over which the signal will reduce to 10% of its value ate the surface of the sea?

4.18 A material has dielectric constant 2.4 and loss-tangent 0.1 at 1 GHz. Find the complex propagation constant of a plane wave in the medium at 1 GHz and 10 GHz.

4.19 If the electric field inside the material in Problem 4.18 is 10 V/m at 6 GHz, find the intrinsic impedance of the medium and the complex magnetic field.

4.20 For a semiconducting material $\epsilon_r = 11$ and $\sigma = 10^{1}$ ℧/m. What is the skin depth of the material at 1 GHz?

4.21 What is the resistance of a copper rod of 1 mm diameter and 1cm length at 100 MHz. The conductivity of copper is 5.6×10^7 ℧/m.

4.22 Inside a conductor with $\sigma = 10^6$ ℧/m, the electric field at some location at 900 MHz is 1 μV/m. What is the magnetic field at that location?

4.23 What is the phase velocity and the wavelength inside the conductor in Problem 4.22 at 900 MHz?

4.24 A computer generated EMI is to be shielded by placing the computer inside an aluminium box. The electric field due to computer at 1 MHz is 1 V/m. Find the thickness of the box so as to get the EMI of less than 1 nV/outside the box.

4.25 Inside a dielectric material with $\epsilon_r = 9, 8$, the electric field of a wave if 100 V/m. Find the magnetic field and the Poynting vector of the wave.

4.26 On the surface of a aluminium sheet, the tangential electric field is 1 μV/m at 1 GHz. What is the powerloss per unit area of the sheet?

4.27 A circularly polarized wave in vacuum has the average power density of 50 W/m^2. Find the magnitudes of the electric and the magnetic fields.

4.28 A wave at 900 MHz is normally incident on a thick dielectric slab with $\epsilon_r = 4$ and $\tan\delta = 10^{-2}$. The power density of the wave just inside the slab is 100 W/m^2. Find the electric and the magnetic fields just inside the slab and at a depth of 100 m inside the slab.

4.29 The electric field at the surface of a copper sheet is 10 μV/m at 10 MHz. Find the magnetic field at the surface and the power loss per unit area of the sheet.

CHAPTER 5

Plane Waves at Media Interface

In the previous chapter, we studied the solution of Maxwell'szz equations for time varying fields in an unbound medium. We found that the time varying fields exist in the form of a plane transverse electromagnetic wave in an infinite medium. For our convenience, we oriented the coordinate system to align with the direction of the wave motion and studied the behavior of electric and magnetic fields in the plane transverse to the direction of the wave propagation. The arbitrary orientation of the coordinate axis was possible, because, in an infinite medium, there is no special direction as the medium looks same in all directions. This is obviously a hypothetical situation. In practice, we never see a medium which is uniform in all directions. In the next few chapters, we will gradually make the medium bound and try to capture the EM wave in a completely closed space. The first step towards this would be to make the medium semi-infinite, i.e. divide the space into two semi-infinite regions with different medium properties and study the behavior of a plane wave at the interface of the two regions. Specifically, we investigate the transfer of fields and power from one medium to another. In this chapter, we ponder over the question like "What happens to the plane wave nature of the wave at the interface ?" "What fraction of the wave energy is transported to the second medium?" "What happens to the polarization of the wave", etc. and try to find their answers.

It is very clear that since, space is no more symmetric, the coordinate axes cannot be oriented arbitrarily. If we orient the coordinate axes along the media interface, the wave will be travelling in an arbitrary direction (not along any of the axes) and if the coordinate axes are oriented to get one of the axes along the wave motion, the axes will have arbitrary orientation with respect to the media interface. Generally, the coordinate axes are aligned along the interface and the wave is assumed to travel at an angle with respect to the media interface. It is, therefore, necessary to formulate the wave function for a wave which is travelling at an arbitrary angle with respect to the coordinate axes.

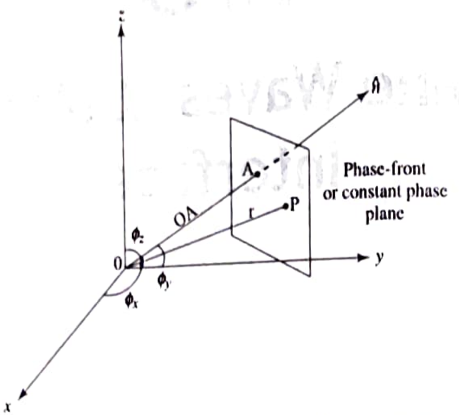


Fig. 5.1 *Wave travelling in an arbitrary direction with respect to the coordinate axes.*

5.1 PLANE WAVE IN ARBITRARY DIRECTION

As we have seen earlier, a plane wave is described by a phase front which is a plane perpendicular to the direction of the wave motion. As shown in Fig. 5.1, let us consider a wave travelling in some arbitrary direction, and let the unit vector in the direction of wave motion be denoted by $\hat{\mathbf{n}}$. If the unit vector $\hat{\mathbf{n}}$ makes angles ϕ_x, ϕ_y, ϕ_z respectively with the three axes x, y, z we have

$$\hat{\mathbf{n}} = \cos\phi_x\hat{\mathbf{x}} + \cos\phi_y\hat{\mathbf{y}} + \cos\phi_z\hat{\mathbf{z}} \tag{5.1}$$

where $\cos\phi_x$, $\cos\phi_y$ and $\cos\phi_z$ are called the direction cosines of the vector $\hat{\mathbf{n}}$.

The planes perpendicular to $\hat{\mathbf{n}}$ are then the phase fronts or the constant phase planes. Let us consider one of the phase fronts as shown in Fig. 5.1, and let any point P on the plane have coordinates (x, y, z). The vector **OP** can be written as

$$\mathbf{OP} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \equiv \mathbf{r} \tag{5.2}$$

From Fig. 5.1 we can see that the dot product

$$\hat{\mathbf{n}}\cdot\mathbf{OP} = \hat{\mathbf{n}}\cdot\mathbf{r} = |\mathbf{OA}| \tag{5.3}$$
$$= \text{Normal distance of the plane from the origin}$$

Equation (5.3) is valid for any point on the phase front and hence, we can write the equation of the phase front as

$$\hat{\mathbf{n}}\cdot\mathbf{OP} = \hat{\mathbf{n}}\cdot\mathbf{r} = \text{constant} \tag{5.4}$$

If we assume that phase of the plane which passes through the origin is zero and if the phase constant of the wave is β, the phase change over the distance OA will be

$$\beta|\mathbf{OA}| = \beta\hat{\mathbf{n}}\cdot\mathbf{r} \tag{5.5}$$

The electric field of a plane wave travelling in direction $\hat{\mathbf{n}}$ can then be written as

$$\mathbf{E} = \mathbf{E_0}e^{-j\beta\hat{\mathbf{n}}\cdot\mathbf{r}} \tag{5.6}$$

where $\mathbf{E_0}$ is a constant vector$(\equiv E_{0x}\hat{\mathbf{x}} + E_{0y}\hat{\mathbf{y}} + E_{0z}\hat{\mathbf{z}})$.

For a transverse electromagnetic wave, $\mathbf{E_0}$ lies in the constant phase plane and hence is perpendicular to $\hat{\mathbf{n}}$. We, therefore, have

$$\mathbf{E_0}\cdot\hat{\mathbf{n}} = 0 \tag{5.7}$$

If we now define the vector propagation constant **k** (also called the wave vector) as

$$\begin{aligned}\mathbf{k} = \beta\hat{\mathbf{n}} &= \beta\{\cos\phi_x\hat{\mathbf{x}} + \cos\phi_y\hat{\mathbf{y}} + \cos\phi_z\hat{\mathbf{z}}\}\\ &= \beta\cos\phi_x\hat{\mathbf{x}} + \beta\cos\phi_y\hat{\mathbf{y}} + \beta\cos\phi_z\hat{\mathbf{z}}\end{aligned} \tag{5.8}$$

the electric field can be written as

$$\mathbf{E} = \mathbf{E_0}e^{-j\mathbf{k}\cdot\mathbf{r}} = \mathbf{E_0}e^{-j(k_x x + k_y y + k_z z)} \tag{5.9}$$

with

$$\mathbf{E_0}\cdot\mathbf{k} = 0 \tag{5.10}$$

where we have defined $k_x = \beta\cos\phi_x$, $k_y = \beta\cos\phi_y$, and $k_z = \beta\cos\phi_z$.

From Eqns (5.9) and (5.10) we can verify that if the wave travels in $+z$ direction (as taken in the previous chapter) $\phi_x = \phi_y = \pi/2$ and $\phi_z = 0$, and the wave vector becomes

$$\mathbf{k} = \beta\{\cos\left(\frac{\pi}{2}\right)\hat{\mathbf{x}} + \cos\left(\frac{\pi}{2}\right)\hat{\mathbf{y}} + \cos(0)\hat{\mathbf{z}}\} = \beta\hat{\mathbf{z}} \tag{5.11}$$

The corresponding electric field can then be written as

$$\mathbf{E} = \mathbf{E_0}e^{-j\beta\hat{\mathbf{z}}\cdot\mathbf{r}} = \mathbf{E_0}e^{-j\beta z} \tag{5.12}$$

Since, $\mathbf{E_0}$ is perpendicular to $\hat{\mathbf{z}}$, it lies in the xy-plane. Equation (5.12) is identical to the plane wave Eqn (4.54).

The magnetic field of a wave can be obtained by substituting **E** in the Maxwell's equation

$$\begin{aligned}\mathbf{H} &= -\frac{1}{j\omega\mu}\nabla\times\mathbf{E}\\ &= -\frac{1}{j\omega\mu}\begin{vmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}}\\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z}\\ E_x & E_y & E_z\end{vmatrix}\end{aligned} \tag{5.13}$$

where $E_x = E_{0x}e^{-j\mathbf{k}\cdot\mathbf{r}}$, $E_y = E_{0y}e^{-j\mathbf{k}\cdot\mathbf{r}}$ and $E_z = E_{0z}e^{-j\mathbf{k}\cdot\mathbf{r}}$, and $\mathbf{k}\cdot\mathbf{r} = k_x x + k_y y + k_z z$.

Since E_{0x}, E_{0y}, E_{0z} are constants,

$$\frac{\partial}{\partial x}\{E_x, E_y, E_z\} = -jk_x\{E_x, E_y, E_z\}$$

$$\Rightarrow \qquad \text{Operator } \frac{\partial}{\partial x} \equiv -jk_x \tag{5.14}$$

Similarly, we can get

$$\frac{\partial}{\partial y} \equiv -jk_y$$

$$\frac{\partial}{\partial z} \equiv -jk_z$$

Substituting for partial derivative operators in Eqn (5.13), the magnetic field can be written as

$$\mathbf{H} = -\frac{1}{j\omega\mu}\nabla\times\mathbf{E}$$

$$= -\frac{1}{j\omega\mu}\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ -jk_x & -jk_y & -jk_z \\ E_x & E_y & E_z \end{vmatrix}$$

$$= -\frac{1}{j\omega\mu}\{-j\mathbf{k}\times\mathbf{E}\} = \frac{1}{\omega\mu}\mathbf{k}\times\mathbf{E} \tag{5.15}$$

From Eqn (5.15) we can note that $\mathbf{H}$ is perpendicular to both $\mathbf{k}$ and $\mathbf{E}$ and since $\mathbf{E}$ is perpendicular to $\mathbf{k}$, the three vectors $\mathbf{E}$, $\mathbf{H}$ and $\mathbf{k}$ are perpendicular to each other (Transverse electromagnetic wave).

Since, the wave vector $= \beta\hat{\mathbf{n}} = \omega\sqrt{\mu\epsilon}\,\hat{\mathbf{n}}$, and $\sqrt{\mu/\epsilon} = \eta$ (intrinsic impedance of the medium), Eqn (5.15) yields

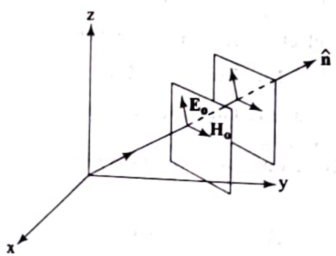


Fig. 5.2 *Transverse nature of electromagnetic wave and constant phase planes.*

$$\mathbf{H} = \frac{1}{\omega\mu}\beta\hat{\mathbf{n}}\times\mathbf{E} = \frac{\omega\sqrt{\mu\epsilon}}{\omega\mu}\hat{\mathbf{n}}\times\mathbf{E}$$

$$= \frac{\hat{\mathbf{n}}\times\mathbf{E}}{\eta} \tag{5.16}$$

$$= \frac{(\hat{\mathbf{n}}\times\mathbf{E}_0)e^{-j\mathbf{k}\cdot\mathbf{r}}}{\eta} = \mathbf{H}_0 e^{-j\mathbf{k}\cdot\mathbf{r}} \tag{5.17}$$

Since $|\hat{\mathbf{n}}\times\mathbf{E}_0| = |\mathbf{E}_0|$, the magnitude of the magnetic field $|\mathbf{H}_0|$ is equal to $|\mathbf{E}_0|/\eta$.

Figure 5.2 shows the constant phase planes for a transverse electromagnetic wave.

EXAMPLE 5.1 A 400 MHz uniform plane wave is travelling in free space along a direction which makes 60° angle with the x-axis and 45° angle with the y-axis. Find the expressions for the vector electric and magnetic fields. The electric field is linearly polarized and has a peak amplitude of 10 V/m. Also, the x-component of the field is twice the y-component of the field.

Solution:

In free space, the velocity of the wave is 3×10^8 m/s.

$$\Rightarrow \qquad \beta = \frac{\omega}{3\times10^8} = \frac{2\pi\times400\times10^6}{3\times10^8} = \frac{8\pi}{3}\text{ rad/m.}$$

The direction cosines of the wave normal are

$$\cos\phi_x = \cos 60^\circ = \frac{1}{2}$$

$$\cos\phi_y = \cos 45^\circ = \frac{1}{\sqrt{2}}$$

Since,

$$\cos^2\phi_x + \cos^2\phi_y + \cos^2\phi_z = 1$$

$$\text{we get,} \quad \cos^2\phi_z = 1 - \frac{1}{4} - \frac{1}{2} = \frac{1}{4}$$

$$\Rightarrow \qquad \cos\phi_z = \frac{1}{2}$$

The unit vector along the direction of the wave normal is

$$\hat{\mathbf{n}} = \frac{1}{2}\hat{\mathbf{x}} + \frac{1}{\sqrt{2}}\hat{\mathbf{y}} + \frac{1}{2}\hat{\mathbf{z}}$$

The electric field is given by

$$\mathbf{E} = \mathbf{E}_0 e^{-j\beta\hat{\mathbf{n}}\cdot\mathbf{r}}$$

$$= \mathbf{E}_0 e^{-j\frac{8\pi}{3}\left(\frac{x}{2}+\frac{y}{\sqrt{2}}+\frac{z}{2}\right)}$$

$\mathbf{E}_0$ is a constant vector perpendicular to the wave normal.

$$\mathbf{E}_0 = E_{0x}\hat{\mathbf{x}} + E_{0y}\hat{\mathbf{y}} + E_{0z}\hat{\mathbf{z}}$$

Since, the wave is linearly polarized E_{0x}, E_{0y} and E_{0z} should be in phase. Therefore, without losing generality, let us assume them to be real.
Since, for uniform plane wave $\mathbf{E}_0$ is perpendicular to $\hat{\mathbf{n}}$, we have

$$\hat{\mathbf{n}} \cdot \mathbf{E}_0 = \frac{E_{0x}}{2} + \frac{E_{0y}}{\sqrt{2}} + \frac{E_{0z}}{2} = 0 \tag{1}$$

It is given that,

$$E_{0x} = 2E_{0y} \tag{2}$$

and the peak amplitude $= \sqrt{E_{0x}^2 + E_{0y}^2 + E_{0z}^2} = 10$ V/m (3)

Solving (1), (2) and (3), we get,

$$E_{0x} = 4.9 \text{ V/m}$$

$$E_{0y} = 2.45 \text{ V/m}$$

$$E_{0z} = -8.37 \text{ V/m}$$

and the vector electric field is given as

$$\mathbf{E} = (4.9\hat{\mathbf{x}} + 2.45\hat{\mathbf{y}} - 8.37\hat{\mathbf{z}})e^{-j\frac{8\pi}{3}\left(\frac{x}{2}+\frac{y}{\sqrt{2}}+\frac{z}{2}\right)}$$

Using Eqn (5.15), the vector magnetic field can be obtained as

$$H = \frac{1}{\omega\mu_0}(\mathbf{k} \times \mathbf{E}) = \frac{\beta}{\omega\mu_0}(\hat{\mathbf{n}} \times \mathbf{E}_0)e^{-j\beta\hat{\mathbf{n}}\cdot\mathbf{r}}$$

$$= \sqrt{\frac{\epsilon_0}{\mu_0}} \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ 1.5496 & 0.7748 & -2.6454 \end{vmatrix} e^{j\beta\hat{\mathbf{n}}\cdot\mathbf{r}}$$

$$= \frac{1}{120\pi}(-7.14\hat{\mathbf{x}} + 6.63\hat{\mathbf{y}} - 2.24\hat{\mathbf{z}})e^{-j\frac{8\pi}{3}\left(\frac{x}{2}+\frac{y}{\sqrt{2}}+\frac{z}{2}\right)} \text{ A/m}$$

[Note: The intrinsic impedance of the free-space is is $\sqrt{\mu_0/\epsilon_0} = 120\pi$.]

5.1.1 Phase Velocity and Wavelength

As seen in the previous section, the choice of coordinate system is generally guided by the media boundaries, etc. and the wave travels in an arbitrary direction. It is, however, useful to find the velocity of the wave along the principal coordinate axes. As will be seen in the following sections, the phase velocity of a wave along the principal axes is not a simple vector resolution of the phase velocity of the wave. The phase velocity along an axis is the velocity of the constant phase point along that axis.

Let us consider a wave with wave vector $\mathbf{k}$. The electric field for this wave is given by (see Eqn (5.9)).

$$\mathbf{E} = \mathbf{E}_0 e^{-j\mathbf{k}\cdot\mathbf{r}}$$

$$= \mathbf{E}_0 e^{-j\beta x\cos\phi_x} e^{-j\beta y\cos\phi_y} e^{-j\beta y\cos\phi_z} \tag{5.18}$$

Let us now find the phase velocity along the z-axis. Rewriting Eqn (5.18) we get

$$\mathbf{E} = \mathbf{E}_0 e^{-j\beta(x\cos\phi_x + y\cos\phi_y)} e^{-j\beta z\cos\phi_z} \tag{5.19}$$

Two things should be noted from Eqn (5.19):

1. In the xy plane (plane perpendicular to z-direction) the phase is not constant. So xy-plane is not a constant phase plane.
2. The phase constant along z-direction is $k_z = \beta\cos\phi_z$.

Since, $\cos\phi_z$ is always less than or equal to unity, k_z is always $\leq \beta$. The phase velocity in the z-direction is therefore

$$v_{pz} = \frac{\omega}{k_z} = \frac{\omega}{\beta\cos\phi_z} = \frac{v_0}{\cos\phi_z} \tag{5.20}$$

where v_0 is the phase velocity of the wave in the direction $\mathbf{k}$, i.e. actual phase velocity of the wave.

Similarly, we can get the phase velocities of the wave along the x and y directions as

$$v_{px} = \frac{\omega}{k_x} = \frac{\omega}{\beta\cos\phi_x} = \frac{v_0}{\cos\phi_x} \tag{5.21}$$

$$v_{py} = \frac{\omega}{k_y} = \frac{\omega}{\beta\cos\phi_y} = \frac{v_0}{\cos\phi_y} \tag{5.22}$$

It is interesting to notice that since, $|\cos\phi_x|$, $|\cos\phi_y|$, $|\cos\phi_z| \leq 1$, the velocities v_{px}, v_{py}, v_{pz} are always greater than or equal to v_0. In fact when any of the angles ϕ_x, ϕ_y, $\phi_z \to \pi/2$, the cosines of these angles tend to 0 and the corresponding velocities approach infinity. The bounds on the phase velocity therefore are

$$v_0 \leq v_{px}, v_{py}, v_{pz} \leq \infty \tag{5.23}$$

The wavelength of the wave in x, y, z directions are respectively

$$\lambda_x = \frac{v_{px}}{f} = \frac{\lambda_0}{\cos\phi_x}$$

$$\lambda_y = \frac{v_{py}}{f} = \frac{\lambda_0}{\cos\phi_y}$$

$$\lambda_z = \frac{v_{pz}}{f} = \frac{\lambda_0}{\cos\phi_z} \tag{5.24}$$

where, $\lambda_0 = v_0/f$ is the wavelength of the wave.

From Eqns (5.23) and (5.24) we should note that the phase velocity and the wavelength are smallest in the direction of the wave motion.

If we consider the unbound medium as the free-space, the phase velocity of the wave is $v_0 = c$ (velocity of light in vaccum). Equations (5.23) and (5.24) then yield

$$c \le v_{px},\ v_{py},\ v_{pz} \le \infty \tag{5.25}$$

One would then wonder as to "how the velocity is greater than the velocity of light, and if we can achieve infinite velocity can we send information with infinite speed?" The answer to this question can be obtained from Fig. 5.3.

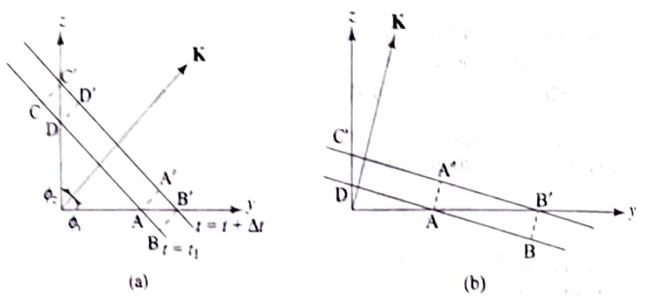


Fig. 5.3 *Phase fronts for a uniform plane wave in the yz-plane.*

For geometrical simplicity let us consider a plane wave travelling in the yz-plane (Fig. 5.3(a)). Let a wavefront intersect the y-axis at point A at time $t = t_1$. After a time Δt the point A moves to A′. The phase front now intersects the y-axis at B′ which corresponds to point B at the wavefront at time $t = t_1$. Similarly, the wave front intersects the z-axis at point C′ which corresponds to point C on the wavefront at $t = t_1$. While calculating the phase velocity along y-direction or z-direction, we do not actually find the distance travelled by a particular point on the wavefront but find the distance between two different points on the wavefront. Consequently, we get a velocity which appears to violate the physical laws (like velocity > c). The velocity of energy corresponds to the distance travelled by point A from A to A′. It is interesting to note from Fig. 5.3 that as the angle ϕ_y increases the distance AB′ increases but the distance DC′ decreases. Consequently, the phase velocity v_y increases but the phase velocity v_z decreases and for $\phi_y \to \pi/2$, $v_y \to \infty$. For $\Delta t \to 0$ the distance DC′ and consequently $v_z \to 0$ but in this case also $v_y \to \infty$. We can therefore conclude that the phase velocity is not a velocity of a physical point on the wavefront and hence is not guided by the physical laws. Obviously, then we can not send information with infinite speed or for that matter with a speed greater than the velocity of light.

It is worthwhile to ask a question at this juncture that "with what velocity does the point A move along the y-axis?" Since the point A moves with velocity v_0 in

the direction of **k**, its velocity in y-direction is

$$v_{gy} = v_0 \cos\phi_y \tag{5.26}$$

v_g is called the group velocity of the wave and v_{gy} denotes the group velocity in the y-direction. The group velocity corresponds to the physical motion of a point on the wavefront and therefore gives the speed with which the energy travels. As can be seen from Eqns (5.26), v_{gy} is always less than or equal to v_0. The same is true for v_{gx} and v_{gz}. From Eqns (5.22) and (5.26) we also note that

$$v_{gy}\, v_{py} = v_0^2 \tag{5.27}$$

In fact, Eqn (5.27) is true for any arbitrary direction and hence we can make a general statement that 'the product of the group and the phase velocities in any direction is equal to the square of the velocity of the wave in the medium'. Figure 5.4 shows the exclusive domains of v_p and v_g. v_p is always greater than or equal to the velocity of the wave in the medium, whereas, v_g is always less than or equal to the velocity of the wave in the medium. As $v_g \to 0$, $v_p \to \infty$, and as $v_g \to v_0$, v_p also $\to v_0$.

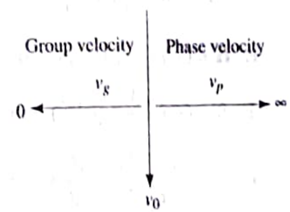


Fig. 5.4 *Domains of v_p and v_g*

EXAMPLE 5.2 The complex vector electric and magnetic fields in an ideal dielectric medium are given as

$$\mathbf{E} = (-j\sqrt{3}\hat{\mathbf{x}} + 2\hat{\mathbf{y}} + j\hat{\mathbf{z}})e^{-j0.02\pi(x+\sqrt{3}z)},$$

$$\mathbf{H} = \frac{1}{30\pi}(\sqrt{3}\hat{\mathbf{x}} + j2\hat{\mathbf{y}} - \hat{\mathbf{z}})e^{-j0.02\pi(x+\sqrt{3}z)}$$

Show that these fields correspond to a uniform plane wave. Find frequency and velocity of the wave. Also find the phase velocities along the x, y and z-directions. Find the state of polarization of the wave.

Solution:

For uniform plane wave **E**, **H** and the wave normal are perpendicular to each other. Therefore, $\mathbf{E}\cdot\mathbf{H}$, $\mathbf{E}\cdot\hat{\mathbf{n}}$ and $\mathbf{H}\cdot\hat{\mathbf{n}}$ should be zero.

We note from the field expressions that

$$\hat{\mathbf{n}}\cdot\mathbf{r} = \frac{1}{2}(x + \sqrt{3}z)$$

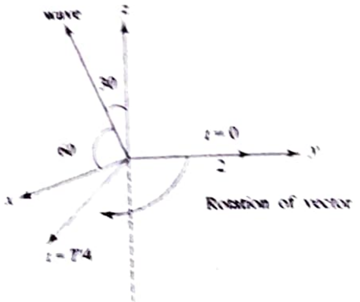


Fig. 5.5 Figure for Example 5.2

$$\Rightarrow \quad \hat{\mathbf{n}} = \frac{1}{2}\hat{\mathbf{x}} + \frac{\sqrt{3}}{2}\hat{\mathbf{z}}$$

$$\mathbf{E} \cdot \mathbf{H} = (-j3 + j4 - j) = 0$$

$$\mathbf{E} \cdot \hat{\mathbf{n}} = (-j\frac{\sqrt{3}}{2} + j\frac{\sqrt{3}}{2}) = 0$$

$$\mathbf{H} \cdot \hat{\mathbf{n}} = \frac{1}{30\pi}(\frac{\sqrt{3}}{2} - \frac{\sqrt{3}}{2}) = 0$$

The fields in Fig. 5.5, therefore, represent fields of a uniform plane wave. Now taking the coefficient of $-j$ in the exponent, we get,

$$\beta\hat{\mathbf{n}} \cdot \mathbf{r} = 0.02\pi(x + \sqrt{3}z) = 0.02\pi(\hat{\mathbf{x}} + \sqrt{3}\hat{\mathbf{z}})(x\hat{\mathbf{x}} + z\hat{\mathbf{z}})$$

$$\Rightarrow \beta\hat{\mathbf{n}} = 0.02\pi(\hat{\mathbf{x}} + \sqrt{3}\hat{\mathbf{z}}) = 0.04\pi\left(\frac{1}{2}\hat{\mathbf{x}} + \frac{\sqrt{3}}{2}\hat{\mathbf{z}}\right)$$

$$\Rightarrow \beta = 0.04\pi$$

The components of **E** and **H** are not in time phase and consequently the fields are not linearly polarized. For the fields the z-component leads the y-component by $\frac{\pi}{2}$ and as the x-component lags the y-component by $\pi/2$. At some instant when the y-component is at its positive peak, the x-component and the z-component are zero. At that instant we have

$$\mathbf{E} = 2\hat{\mathbf{y}} \Rightarrow |\mathbf{E}| = 2$$

and

$$\mathbf{H} = \frac{1}{30\pi}(\sqrt{3}\hat{\mathbf{x}} - \hat{\mathbf{z}}) \qquad \Rightarrow |H| = \frac{2}{30\pi}$$

Now for a uniform plane wave

$$\frac{|E|}{|H|} = \sqrt{\frac{\mu_0}{\epsilon_0 \times \epsilon_r}} = 30\pi = \frac{120\pi}{\sqrt{\epsilon_r}}$$

$$\Rightarrow \qquad \epsilon_r = 16$$

The velocity of the wave

$$v = \frac{1}{\sqrt{\mu_0\epsilon_0\epsilon_r}} = \frac{3 \times 10^8}{\sqrt{\epsilon_r}} = 0.75 \times 10^8 \ \text{m/s.}$$

Now we know that $\omega = 2\pi f = \beta \times$ velocity
$\Rightarrow$ frequency of the wave

$$f = \frac{\beta \times \text{velocity}}{2\pi} = \frac{0.04\pi \times 0.75 \times 10^8}{2\pi} = 1.5 \text{ MHz.}$$

The phase velocity along the x-direction

$$v_x = \frac{v}{\cos\phi_x} = \frac{0.75 \times 10^8}{1/2} = 1.5 \times 10^8 \text{ m/s.}$$

The phase velocity along y-direction is

$$v_y = \frac{v}{\cos\phi_y} = \frac{v}{\cos(\pi/2)} = \infty$$

The phase velocity along z-direction is

$$v_z = \frac{v}{\cos\phi_z} = \frac{0.75 \times 10^8}{(\sqrt{3})} = 0.866 \times 10^8 \text{ m/s.}$$

To find the state of polarization of the wave, let us trace the tip of the electric field vector. The three components of the electric fields are

$$E_x = \sqrt{3}\cos\left(\omega t - \frac{\pi}{2}\right) = \sqrt{3}\sin\omega t$$

$$E_y = 2\cos\omega t$$

$$E_z = \cos\left(\omega t + \frac{\pi}{2}\right) = -\sin\omega t$$

Let the time period of the signal be $T (T = 2\pi/\omega = 1/f)$.

At $t = 0$, $E_x = 0$, $E_y = 2$, $E_z = 0 \Rightarrow |E| = 2$

At $t = T/8$, $\omega t = \pi/4 \Rightarrow E_x = \frac{\sqrt{3}}{\sqrt{2}}, E_y = \frac{2}{\sqrt{2}}, E_z = \frac{-1}{\sqrt{2}} \quad \Rightarrow \quad |E| = 2$

At $t = T/4$, $\omega t = \pi/2 \Rightarrow E_x = \sqrt{3}, E_y = 0, E_z = -1 \quad \Rightarrow \quad |E| = 2$

Since, the $|E|$ is constant as a function time, it is the circular polarization. To get the sense of rotation draw the fields at different times as shown in Fig. 5.5. The vector rotation is towards the left hand if we face in the direction of wave. The wave polarization, therefore, is LHC.

5.2 PLANE WAVE AT DIELECTRIC INTERFACE

In the previous sections we discussed the propagation of a plane wave in an unbound medium. Let us now consider the propagation of a plane wave across a dielectric interface. Let the space be divided into two semi-infinite regions with different medium parameters like the permeability and the permittivity. Let us also assume that both the regions are loss-less, i.e. the conductivity for both regions is zero. Without losing generality let us orient the coordinate system such that the dielectric interface is along the xy-plane passing through $z = 0$ as shown in Fig. 5.6.

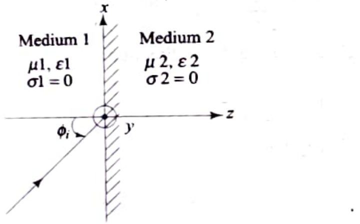


Fig. 5.6 *Electromagnetic wave at a dielectric interface.*

Let us denote the region on the left of the interface ($z < 0$) as region 1, and let all its parameters be denoted with suffix 1. Similarly, let the region on the right of the interface ($z > 0$) be denoted as region 2 and let its parameters be denoted with suffix 2. Then, μ_1, ϵ_1 represent permeability and permittivity of medium 1 respectively and μ_2, ϵ_2 represent corresponding parameters for medium 2.

Let us now consider a plane wave incident on the interface from medium 1 side. For simplicity let us assume that the wave vector **k** lies in the xz-plane and makes an angle θ_i with the normal to the interface (in this case z-direction). The angle θ_i is called the angle of incidence. The plane which contains the wave vector **k** and the normal to the interface is called the *plane of incidence*. The angles which the wave vector makes with the three axes are therefore

$$\phi_x = \frac{\pi}{2} - \theta_i, \qquad \phi_y = \frac{\pi}{2}, \qquad \phi_z = \theta_i \tag{5.28}$$

We can then write the field (electric or magnetic) for this wave as

$$\mathbf{F}_i = \mathbf{F}_{i0} e^{-j\mathbf{k}_i \cdot \mathbf{r}} \tag{5.29}$$

$$= \mathbf{F}_{i0} e^{-j\beta_1 (x\cos\phi_x + y\cos\phi_y + z\cos\phi_z)} \tag{5.30}$$

The suffix i indicates the incident field, $\mathbf{F}_{i0}$ is a constant vector, and β_1 is the phase constant of the wave in medium 1, $\beta_1 = \omega\sqrt{\mu_1\epsilon_1}$.

Substituting for ϕ_x, ϕ_y, ϕ_z into Eqn (5.30) we get

$$\mathbf{F}_i = \mathbf{F}_{i0} e^{-j\beta_1 (x\cos(\pi/2-\theta_i) + y\cos(\pi/2) + z\cos\theta_i)}$$

$$= \mathbf{F}_{i0} e^{-j\beta_1 (x\sin\theta_i + z\cos\theta_i)} \tag{5.31}$$

From Eqn (5.31) it is clear that the field is independent of y.

At the interface i.e. at $z = 0$, the phase is constant along y-direction and it linearly increases along the x-direction. The field magnitude in the xy-plane therefore is

$$A = Re\{\mathbf{F}_i e^{-j\beta_1 x \sin\theta_i}\}$$

$$= \mathbf{F}_{i0} \cos(\beta_1 x \sin\theta_i) \tag{5.32}$$

The field variation is like a sinusoidally corrugated sheet with corrugations oriented along the y-axis as shown in Fig. 5.7.

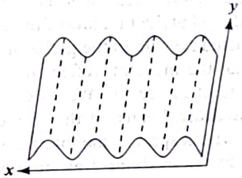


Fig. 5.7 *Spatial variation of electric field amplitude.*

When the sinusoidally corrugated field is incident at the interface, a field with similar corrugated variation is induced on the other side of the interface. This is due to the boundary conditions which require continuity of the fields at the dielectric boundary. One can additionally show that both magnetic field and electric field in general, cannot satisfy the boundary conditions without modifying the incident fields. In other words, we can say that when a plane wave is incident on an interface, the fields with similar phase variation are induced on both sides of the interface. In region 1 then the total field is the combination of the incident field and the induced field, where as in region 2 there is only the induced field. Since, the induced fields are also time varying, they also constitute waves. The induced fields in region 1 then form a wave moving away from the interface, i.e. back in region 1, whereas the induced fields in region 2 form a wave going away from the interface in region 2 as shown in Fig. 5.8. Since, both induced waves see infinite medium ahead of them they are in the form of plane waves. These waves are called the reflected and transmitted (or refracted) waves respectively.

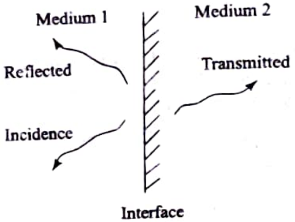


Fig. 5.8 *Incident, reflected and refracted waves.*

We can now note that since the phase variation for the induced fields is same as that of the incident field at the interface (constant along y-direction), the wave vectors of the reflected and transmitted waves also lie in the plane containing the incident wave vector and the interface normal, i.e. the plane of incidence. One can then conclude that, the wave vectors for the incident, reflected and refracted (transmitted) waves, and the normal to the interface lie in the same plane. In bulk optics the wave vectors are called the light rays. The reader would then note that the above conclusion is nothing but the first law of reflection of light which states that 'the reflected and refracted rays lie in the plane of incidence'.

We have seen earlier that the incident, reflected and refracted waves travel in the same plane, the plane of incidence. We, however, do not know in what directions they would be travelling. Let us say that the reflected wave travels back in medium 1 at an angle θ_r with respect to the interface normal and the refracted (transmitted) wave travels in the medium 2 at an angle θ_t with respect to the interface normal, as shown in Fig. 5.9.

For reflected wave $\phi_x = \pi/2 - \theta_r$, $\phi_y = \pi/2$ and $\phi_z = \pi - \theta_r$, and the field $\mathbf{F_r}$ for the reflected wave can be written as

$$\mathbf{F_r} = \mathbf{F_{r0}} e^{-j\beta_1\{x\cos(\pi/2-\theta_r)+y\cos(\pi/2)+z\cos(\pi-\theta_r)\}}$$

$$= \mathbf{F_{r0}} e^{-j\beta_1\{x\sin\theta_r - z\cos\theta_r\}} \tag{5.33}$$

Similarly, for the transmitted wave $\phi_x = \pi/2 - \theta_t$, $\phi_y = \pi/2$, and $\phi_z = \theta_t$ and its field can be written as

$$\mathbf{F_t} = \mathbf{F_{t0}} e^{-j\beta_2\{x\cos(\pi/2-\theta_t)+y\cos(\pi/2)+z\cos\theta_t\}}$$

$$= \mathbf{F_{t0}} e^{-j\beta_2\{x\sin\theta_t + z\cos\theta_t\}} \tag{5.34}$$

where, $\beta_2 = \omega\sqrt{\mu_2\epsilon_2}$ is the phase constant in medium 2. At the interface, i.e. at $z = 0$ the fields must satisfy the boundary conditions. If fields $\mathbf{F_i}$, $\mathbf{F_r}$, $\mathbf{F_t}$ represent electric fields, their tangential components must be continuous across the interface. On the other hand, if we take $\mathbf{F_i}$, $\mathbf{F_r}$, $\mathbf{F_t}$ as magnetic fields, generally

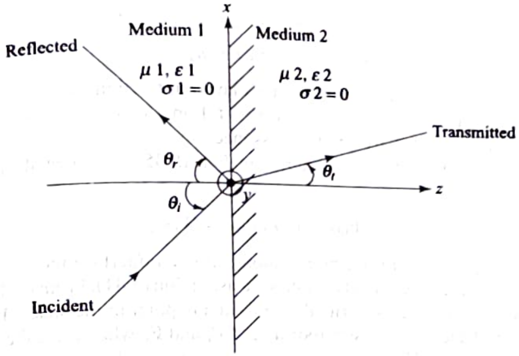


Fig. 5.9 *Incident, transmitted and reflected waves at a media interface.*

their normal components must be continuous. However, in the present case, their tangential components also must be continuous at the interface since for ideal dielectrics there is no surface current. So, without worrying about whether **F** represents **E** or **H**, we can say in general that their tangential components must be continuous at the interface. We therefore have

$$(\mathbf{F_{i0}})_{\tan} e^{-j\beta_1 x\sin\theta_i} + (\mathbf{F_{r0}})_{\tan} e^{-j\beta_1 x\sin\theta_r} = (\mathbf{F_{t0}})_{\tan} e^{-j\beta_2 x\sin\theta_t} \tag{5.35}$$

Suffix 'tan' represents the tangential component of the vector.

Equation (5.35) is true for every point on the interface, i.e. for every value of x and y. This can happen only if the phases of the three terms are same i.e.

$$\beta_1 x\sin\theta_i = \beta_1 x\sin\theta_r = \beta_2 x\sin\theta_t \tag{5.36}$$

$$\Rightarrow \quad \sin\theta_i = \sin\theta_r$$

$$\Rightarrow \quad \theta_i = \theta_r \tag{5.37}$$

and

$$\beta_1 \sin\theta_i = \beta_2 \sin\theta_t \tag{5.38}$$

$$\Rightarrow \quad \sqrt{\mu_1\epsilon_1}\sin\theta_i = \sqrt{\mu_2\epsilon_2}\sin\theta_t \tag{5.39}$$

From Eqn (5.37) we get the second law of reflection that is, 'the angle of reflection is equal to the angle of incidence'.

For ideal dielectrics $\mu_1 = \mu_2 = \mu_0$ (free space permeability), $\epsilon_1 = \epsilon_0\epsilon_{r1}$ and $\epsilon_2 = \epsilon_0\epsilon_{r2}$ where ϵ_{r1} and ϵ_{r2} are the dielectric constants of the two media respectively. Noting from eqn (4.140) that $\sqrt{\epsilon_{r1}} = n_1$ (refractive index of medium 1) and $\sqrt{\epsilon_{r2}} = n_2$ (refractive index of medium 2), Eqn (5.39) can be written as

$$\sqrt{\mu_0\epsilon_0\epsilon_{r1}}\sin\theta_i = \sqrt{\mu_0\epsilon_0\epsilon_{r2}}\sin\theta_t$$

$$\Rightarrow \qquad \sqrt{\epsilon_{r1}}\sin\theta_i = \sqrt{\epsilon_{r2}}\sin\theta_t$$

$$\Rightarrow \qquad n_1 \sin\theta_i = n_2 \sin\theta_t \tag{5.40}$$

Equation (5.40) is the well known Snell's law of refraction. One should however note that Eqn (5.40) is a special case of Eqn (5.39) which represents the generalized Snell's law for any loss-less medium.

Once the phases of the three terms in Eqn (5.35) are equated the equation reduces to

$$(\mathbf{F}_{i0})_{\tan} + (\mathbf{F}_{r0})_{\tan} = (\mathbf{F}_{t0})_{\tan} \tag{5.41}$$

That is, the problem of wave propagation across an interface reduces to finding the fields of the three waves at the interface using Eqn (5.41). From the knowledge of $\mathbf{F}_{i0}$, $\mathbf{F}_{r0}$ and $\mathbf{F}_{t0}$ we can write the fields at any point in the space. If the point is in region 1 the field is a superposition of $\mathbf{F}_i$ and $\mathbf{F}_r$ whereas, if the point is in region 2 the field is $\mathbf{F}_t$.

EXAMPLE 5.3 A light beam is incident from air to a medium with a dielectric constant 4 and relative permeability 100. If the angle of incidence is 60°. Find the angle of reflection and angle of refraction.

Solution:

The angle of reflection θ_r = angle of incidence $\theta_i = 60°$. From the Snell's law,

$$\sqrt{\mu_1\epsilon_1}\sin\theta_i = \sqrt{\mu_2\epsilon_2}\sin\theta_t$$

$$\Rightarrow \sqrt{\mu_0\epsilon_0}\sin 60° = \sqrt{\mu_0(100)\epsilon_0(4)}\sin\theta_t$$

$$\sin\theta_t = \frac{\sin 60°}{20}$$

$$= 0.0433$$

$$\Rightarrow \qquad \text{Angle of refraction} \quad \theta_t = 2.48°$$

5.3 REFLECTION AND REFRACTION OF WAVES AT DIELECTRIC INTERFACE

In the previous sections, we had a general discussion on the propagation of a plane wave across a loss-less dielectric interface. Let us now analyze the specific cases of field orientation with respect to the plane of incidence. As can be seen from Fig. 5.10, **E** (and consequently **H**) vector can make an arbitrary angle with respect to the plane of incidence.

Geometrically, it is quite complicated to handle the fields with arbitrary orientation. Instead, one can elegantly handle the case by decomposing the field into its components, one in the plane of incidence and other perpendicular to it and then combining the results for the two components. In the following, therefore,

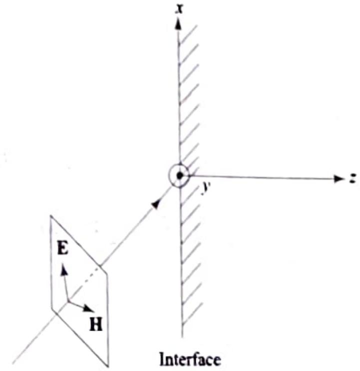


Fig. 5.10 *A wave incident at a media interface with its field at an angle with respect to the plane of incidence.*

we discuss the two cases (i) Electric field in the plane of incidence, also called *parallel polarization* (ii) Electric field normal to the plane of incidence also called *perpendicular polarization*. Specifically, we are interested in finding out the field amplitudes of the transmitted and the reflected waves for a given incident wave. Normally, we carry out the analysis for the **E** field and as discussed earlier we can find the magnetic field as and when needed from the property of the uniform plane wave. We, therefore, define two parameters of interest namely

$$\text{Reflection coefficient: } \Gamma = \frac{\mathbf{E}_r}{\mathbf{E}_i} \tag{5.42a}$$

$$\text{Transmission coefficient: } \tau = \frac{\mathbf{E}_t}{\mathbf{E}_i} \tag{5.42b}$$

and find their values for the parallel and perpendicular polarizations.

5.3.1 Reflection and Refraction with Perpendicular Polarization

Let us consider a plane wave with perpendicular ($\perp$) polarization incident at a dielectric interface as shown in Fig. 5.11. The angle of incidence is say θ_i (angle which the wave vector makes with the interface normal) and the plane of incidence is the plane of the paper. Without losing generality let us assume that the incident electric field $\mathbf{E}_i$ is pointing out of the plane of the paper (oriented in y-direction). Then, from the property of the plane wave the magnetic field $\mathbf{H}_i$ will be perpendicular to $\mathbf{E}_i$ and will lie in the plane of the paper perpendicular to the wave vector as shown in Fig. 5.11. The $\mathbf{H}_i$, therefore, lies in the xz-plane. Also, the ratio of the magnitudes of $\mathbf{E}_i$ and $\mathbf{H}_i$ is η_1, the intrinsic impedance of medium 1.

As discussed in the previous sections, the wave vectors of reflected and transmitted waves also lie in the plane of incidence (in this case the plane of

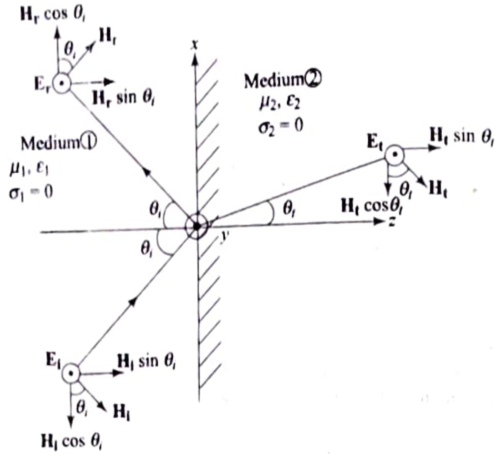


Fig. 5.11 *Uniform plane wave with perpendicular polarization at a media interface.*

the paper), and the angle of reflection is same as the angle of incidence. One may further argue that since the tangential component of **E** should be continuous at the interface and since, the incident wave has only y-directed electric field, the reflected and transmitted waves must also have y-directed **E** field only. Of course, the fields might point into the plane of paper or point out of it. However, the direction reversal can easily be accommodated by assigning a +ve or -ve sign to a field. Therefore, without losing generality we assume that both fields $\mathbf{E_r}$ and $\mathbf{E_t}$ are along $+y$ direction, i.e. pointing outwards from the plane of the paper. If any of them or both of them were directed in the opposite direction (along $-y$-direction) their signs would come negative automatically.

Now for each wave, the Poynting vector $\mathbf{E} \times \mathbf{H}$ should give the direction of the wave vector. Since, the directions of the incident, reflected and transmitted waves are known, the directions of the magnetic field for the three waves $\mathbf{H_i}$, $\mathbf{H_r}$ and $\mathbf{H_t}$ can be easily obtained as shown in Fig. 5.11. Note that, the incident and transmitted waves are travelling from left to right upwards and consequently the magnetic fields point downwards. Whereas, the reflected wave is travelling from right to left upwards and hence its magnetic field is oriented upwards.

We can write the electric fields for the three waves as (see Eqns (5.31), (5.33) and (5.34))

$$\text{Incident wave:} \quad \mathbf{E_i} = \mathbf{E_{i0}} e^{-j\beta_1(x\sin\theta_i + z\cos\theta_i)} \tag{5.43}$$

$$\text{Reflected wave:} \quad \mathbf{E_r} = \mathbf{E_{r0}} e^{-j\beta_1(x\sin\theta_i - z\cos\theta_i)} \tag{5.44}$$

$$\text{Transmitted wave:} \quad \mathbf{E_t} = \mathbf{E_{t0}} e^{-j\beta_2(x\sin\theta_t + z\cos\theta_t)} \tag{5.45}$$

where $\beta_1 = \omega\sqrt{\mu_1\epsilon_1}$, $\beta_2 = \omega\sqrt{\mu_2\epsilon_2}$ and $\mathbf{E_{i0}}$, $\mathbf{E_{r0}}$, and $\mathbf{E_{t0}}$ are the vector amplitudes of the three waves. For all the fields, a time variation of $e^{j\omega t}$ is implicitly assumed. Since, **E** and **H** are related through intrinsic impedance of the medium we also have

$$|\mathbf{H_i}| = \frac{|\mathbf{E_i}|}{\eta_1} \tag{5.46}$$

$$|\mathbf{H_r}| = \frac{|\mathbf{E_r}|}{\eta_1} \tag{5.47}$$

$$|\mathbf{H_t}| = \frac{|\mathbf{E_r}|}{\eta_2} \tag{5.48}$$

where $\eta_1 = \sqrt{\mu_1/\epsilon_1}$ and $\eta_2 = \sqrt{\mu_2/\epsilon_2}$ are the intrinsic impedances of medium 1 and medium 2 respectively.

Now, applying the boundary conditions, i.e. tangential component of the electric field is continuous at the interface, and since there is no surface current even the tangential component of magnetic field is continuous at the interface, we get at $z = 0$

$$\mathbf{E_i} + \mathbf{E_r} = \mathbf{E_t} \tag{5.49}$$

$$\mathbf{H_i}\cos\theta_i - \mathbf{H_r}\cos\theta_i = \mathbf{H_t}\cos\theta_t \tag{5.50}$$

Note that, the **E** fields are in y-direction and hence are tangential to the interface whereas for the magnetic field we have to take the component along the interface. Substituting Eqns (5.46) to (5.48) in (5.49) and (5.50) we get

$$\mathbf{E_{i0}} + \mathbf{E_{r0}} = \mathbf{E_{t0}} \tag{5.51}$$

$$\frac{\mathbf{E_{i0}}}{\eta_1}\cos\theta_i - \frac{\mathbf{E_{r0}}}{\eta_1}\cos\theta_i = \frac{\mathbf{E_{t0}}}{\eta_2}\cos\theta_t \tag{5.52}$$

Solving Eqns (5.51) and (5.52) we can obtain the reflection and transmission coefficients as

$$\text{Reflection coeff:} \quad \Gamma_\perp = \frac{\mathbf{E_{r0}}}{\mathbf{E_{i0}}} = \frac{\eta_2\cos\theta_i - \eta_1\cos\theta_t}{\eta_2\cos\theta_i + \eta_1\cos\theta_t} \tag{5.53}$$

$$\text{Transmission coeff:} \quad \tau_\perp = \frac{\mathbf{E_{t0}}}{\mathbf{E_{i0}}} = \frac{2\eta_2\cos\theta_i}{\eta_2\cos\theta_i + \eta_1\cos\theta_t} \tag{5.54}$$

From Eqns (5.53), (5.54) or dividing Eqn (5.51) by $\mathbf{E_{i0}}$ we get relation between the transmission and the reflection coefficients as

$$1 + \frac{\mathbf{E_{r0}}}{\mathbf{E_{i0}}} = \frac{\mathbf{E_{t0}}}{\mathbf{E_{i0}}} \tag{5.55}$$

$$\Rightarrow \quad 1 + \Gamma_\perp = \tau_\perp \tag{5.56}$$

From Eqns (5.53) and (5.54) we can make following observations:

1. The reflection and transmission coefficients are real. That means there is no arbitrary phase change in the reflected or transmitted wave at the interface.

Depending upon the sign of $\Gamma_\perp$ and $\tau_\perp$, the phase change could be either 0 or π (0 for +ve sign and π for negative sign).

2. Magnitude of Reflection coefficient $\Gamma_\perp$ is always less than unity whereas the magnitude of the transmission coefficient could be greater or lesser than unity. It means that the amplitude of the transmitted **E** field could be greater than the amplitude of the incident electric field. Obviously, this would happen when $\Gamma_\perp$ is positive (see Eqn (5.56)), i.e. when

$$\eta_2 \cos\theta_i > \eta_1 \cos\theta_t \tag{5.57}$$

From Snell's law since $\beta_1 \sin\theta_i = \beta_2 \sin\theta_t$ we get

$$\cos\theta_t = \pm\sqrt{1-\sin^2\theta_t} = \pm\sqrt{1-\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i} \tag{5.58}$$

Substitution of $\cos\theta_t$ in Eqn (5.57) yields

$$\eta_2 \cos\theta_i > \eta_1 \cos\theta_t = \eta_1\sqrt{1-\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i} \tag{5.59}$$

$$\Rightarrow \quad \frac{\eta_2^2}{\eta_1^2}\cos^2\theta_i > 1-\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i \tag{5.60}$$

$$\Rightarrow \quad \frac{\mu_2\epsilon_2}{\mu_1\epsilon_1}\cos^2\theta_i + \frac{\mu_1\epsilon_1}{\mu_2\epsilon_2}\sin^2\theta_i > 1 \tag{5.61}$$

Therefore, if Eqn (5.61) is satisfied the transmitted electric field is greater than the incident electric field.

It should be noted here that higher transmitted electric field does not mean higher transmitted power. The magnetic field reduces in the medium 2 appropriately to give the transmitted power less than or equal to the incident power. After all, there has to be conservation of power making sum of the transmitted and the reflected power equal to the incident power, i.e.

$$\frac{|\mathbf{E}_{i0}|^2}{\eta_1}\cos\theta_i = \frac{|\mathbf{E}_{r0}|^2}{\eta_1}\cos\theta_i + \frac{|\mathbf{E}_{t0}|^2}{\eta_2}\cos\theta_t$$

$$\Rightarrow \quad |\Gamma_\perp|^2 + \frac{\eta_1}{\eta_2}|\tau_\perp|^2\frac{\cos\theta_t}{\cos\theta_i} = 1 \tag{5.62}$$

EXAMPLE 5.4 A uniform plane wave having power density 20 W/m^2 is incident from air at air-dielectric interface at the angle of incidence 45°. The electric field vector for the wave lies perpendicular to the plane of incidence. Find the power density of the transmitted and the reflected wave. The relative permittivity of the dielectric medium is 25. If the frequency of the wave is 100 MHz, find the amplitude of the electric and magnetic field at a distance of 1 m on either side of the interface.

Solution:

The wavelength in air

$$\lambda_0 = \frac{c}{f} = \frac{3\times10^8}{100\times10^6} = 3.0 \text{ m}$$

In air the power density

$$= \frac{|E_{i0}|^2}{2\eta_0} = 20$$

$$\Rightarrow |E_{i0}| = \sqrt{40\eta_0} = \sqrt{40(377)} = 122.8 \text{ V/m}$$

$$\eta_1 = \eta_{i0} = 120\pi$$

$$\eta_2 = \frac{\eta_0}{\sqrt{\epsilon_r}} = \frac{120\pi}{\sqrt{25}} = 24\pi$$

$$\beta_1 = \beta_0 = \frac{2\pi}{\lambda_0} = \frac{2\pi}{3}$$

$$\beta_2 = \beta_0\sqrt{\epsilon_r} = \frac{2\pi}{3}\sqrt{25} = \frac{10\pi}{3}$$

From the Snell's law we get

$$\sin\theta_t = \frac{\sin\theta_i}{\sqrt{\epsilon_r}} = \frac{\sin 45^\circ}{\sqrt{25}} = \frac{1}{5\sqrt{2}}$$

$$\Rightarrow \quad \cos\theta_t = 0.9899$$

From Eqns (5.53) and (5.54), we have,

$$\Gamma_\perp = \frac{24\pi\cos45^\circ - 120\pi\times0.9899}{24\pi\cos45^\circ + 120\pi\times0.9899} = -0.75$$

$$\tau_\perp = \frac{2(24\pi)\cos45^\circ}{24\pi\cos45^\circ + 120\pi\times0.9899} = 0.25$$

We, therefore, get

$$E_{r0} = \Gamma_\perp E_{i0} = -0.75\times122.8 = -92.1\text{V/m}$$

and $$E_{t0} = \tau_\perp E_{i0} = 0.25\times122.8 = 30.7\text{V/m}$$

Following the coordinate system in Fig. 5.11, we have

$$\mathbf{E}_i = 122.8e^{-j\beta_1(x\sin(45^\circ)+z\cos(45^\circ))}\hat{\mathbf{y}}$$

$$= 122.8e^{-j\frac{2\pi}{3\sqrt{2}}(x+z)}\hat{\mathbf{y}}$$

$$\mathbf{E}_r = -92.1e^{-j\frac{2\pi}{3\sqrt{2}}(x-z)}\hat{\mathbf{y}}$$

and $$\mathbf{E}_t = 30.7e^{-j\beta_2(x\sin\theta_t+z\cos\theta_t)}\hat{\mathbf{y}}$$

$$= 30.7e^{-j\frac{10\pi}{3}\left(\frac{x}{5\sqrt{2}}+0.9899z\right)}\hat{\mathbf{y}}$$

Power density of the transmitted wave = [illegible] W/m²
Power density of the reflected wave = [illegible] W/m²
In Region 1, the electric field is (at $z = 0$ m)

$$\mathbf{E}_1 = \mathbf{E}_i + \mathbf{E}_r = \left([illegible] \right)$$

$$= [illegible]$$

$$\Rightarrow E_1 = [illegible] \text{ V/m}$$

In Region 2, the electric field is

$$\mathbf{E}_2 = \mathbf{E}_t$$

$$\Rightarrow \qquad \text{peak amplitude} = [illegible] \text{ V/m}$$

The magnetic fields for the three waves are (see Eqns (5.46), (5.47) and (5.48))

$$|\mathbf{H}_i| = [illegible] = 0.4[illegible] \text{ A/m}$$

$$|\mathbf{H}_r| = [illegible] = 0.3[illegible] \text{ A/m}$$

$$|\mathbf{H}_t| = [illegible] = 0.3[illegible] \text{ A/m}$$

Hence, total magnetic field component in x direction in Region 1 can be written as

$$H_x = |\mathbf{H}_r|\cos\theta_r - |\mathbf{H}_i|\cos\theta_i$$

$$= 0.3[illegible]\sqrt{3} - 0.4[illegible]\sqrt{3}$$

$$= 0.1[illegible] - 0.2[illegible] = -0.0[illegible] \text{ A/m}$$

The negative sign shows reversal of the field direction.
Similarly, the total magnetic field in z direction in Region 1 can be written as

$$H_z = |\mathbf{H}_r|\sin\theta_r + |\mathbf{H}_i|\sin\theta_i$$

$$= 0.3[illegible]\sqrt{3} + 0.4[illegible]\sqrt{3}$$

$$= 0.1[illegible] + 0.2[illegible] = 0.4[illegible] \text{ A/m}$$

Hence, peak amplitude of magnetic field in Region 1 is

$$|\mathbf{H}| = \sqrt{|H_x|^2 + |H_z|^2}$$

$$= \sqrt{(0.0[illegible])^2 + (0.4[illegible])^2} = 0.40[illegible] \text{ A/m}$$

Total magnetic field in Region 2 is same as $\mathbf{H}_t$, i.e. 0.3[illegible] A/m

5.3.2 Reflection and Refraction with Parallel Polarisation

On the lines similar to that for the perpendicular polarisation, we can now analyse reflection and refraction of a plane wave having parallel polarisation. In this case

the E vector lies in the plane of incidence. Again, assuming that plane of incidence is same as the plane of the paper, both wave vector and the E vector lie in the plane of the paper. Since, the magnetic field is perpendicular to both, it is now perpendicular to the plane of the paper. Therefore, without losing generality let us assume that the magnetic field is pointing out of the paper (oriented along +y direction) (see Fig. 5.13).

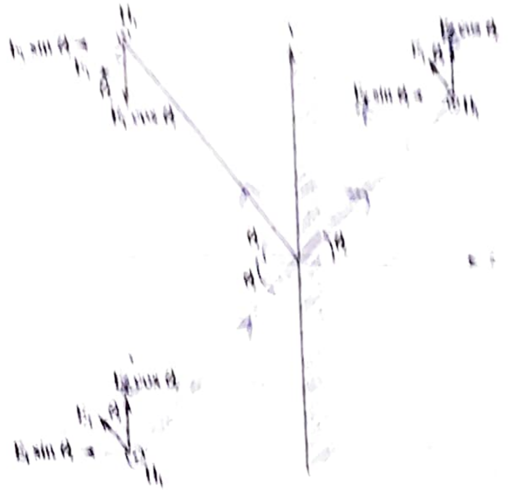


Fig. 5.13 Uniform plane wave with parallel polarisation at a media interface

Using similar arguments as in previous case, we can assume that the magnetic fields for the reflected and transmitted waves are also oriented along +y direction. Using the Poynting vector argument we can then find the directions of the respective electric fields. The electric fields for the incident, reflected and transmitted waves have exactly same form as given by Eqns (5.43) to (5.45) except that the vector $\mathbf{E}_{i0}$, $\mathbf{E}_{r0}$ and $\mathbf{E}_{t0}$ are no more oriented along the +y direction but lie in the xz-plane. The amplitudes of the magnetic fields for the three waves are also given by Eqns (5.46) to (5.48), but their direction is perpendicular to the plane of incidence (+y direction).

Applying continuity of the tangential components of the electric and magnetic fields at the interface we get

$$E_{i0}\cos\theta_i - E_{r0}\cos\theta_r = E_{t0}\cos\theta_t \tag{5.63}$$

$$H_{i0} + H_{r0} = H_{t0} \tag{5.64}$$

Again using Eqns (5.43) to (5.48) and solving Eqns (5.63) and (5.64) we get the reflection and the transmission coefficients for the parallel polarisation as

Reflection coefficient:

$$\Gamma_{\parallel} = \frac{\mathbf{E}_{r0}}{\mathbf{E}_{i0}} = \frac{\eta_1 \cos\theta_i - \eta_2 \cos\theta_t}{\eta_1 \cos\theta_i + \eta_2 \cos\theta_t} \tag{5.65}$$

Transmission coefficient:

$$\tau_{\parallel} = \frac{\mathbf{E}_{t0}}{\mathbf{E}_{i0}} = \frac{2\eta_2 \cos\theta_i}{\eta_2 \cos\theta_t + \eta_1 \cos\theta_i} \tag{5.66}$$

From Eqn (5.63) we note that

$$(1 + \Gamma_{\parallel}) \cos\theta_i = \tau_{\parallel} \cos\theta_t \tag{5.67}$$

In the case of parallel polarization also we can make similar observations as that in the previous case, i.e.

1. The magnitude of the reflection coefficient $\Gamma_{\parallel}$ is always less than unity and there is a phase difference of 0 or π between the incident and the reflected waves.
2. The transmission coefficient $\tau_{\parallel}$ could be greater or less than unity, i.e. the transmitted electric field could be greater or less than the incident electric field.

5.3.3 Reflection and Refraction—Normal Incidence

The reflection and refraction for the normal incidence, i.e. incidence along the normal to the interface can be obtained by making the angle of incidence θ_i zero either for perpendicular or for parallel polarization. Figure 5.13 shows the incident, reflected and transmitted waves with their respective fields for $\theta_i = 0$ in the two polarizations.

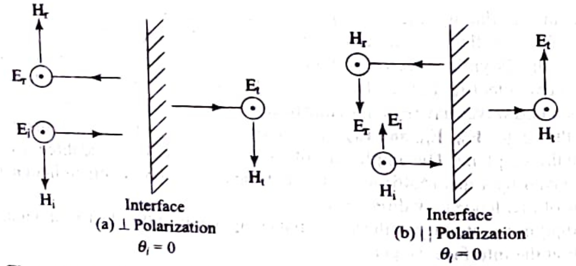


Fig. 5.13 *Uniform plane wave normally incident at a media interface.*

Substituting $\theta_i = 0$ and consequently $\theta_t = 0$ due to Snell' law, in Eqns (5.53) and (5.65) we get

$$\Gamma_{\perp}|_{(\theta_i=0)} = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \tag{5.68}$$

$$\Gamma_{\parallel}|_{(\theta_i=0)} = \frac{\eta_1 - \eta_2}{\eta_1 + \eta_2} \tag{5.69}$$

$$\Rightarrow \quad \Gamma_{\perp}|_{(\theta_i=0)} = -\Gamma_{\parallel}|_{(\theta_i=0)}$$

Since, for $\theta_i = 0$, the $\perp$ and $\parallel$ polarizations represent same case one would wonder why the two reflection coefficients are negative of each other! The reason for this lies in the original directions assumed for the E-fields. In perpendicular polarization case as $\theta_i = 0$, $\mathbf{E}_i$ and $\mathbf{E}_r$ point in the same direction whereas for parallel polarization when $\theta_i = 0$, $\mathbf{E}_i$ and $\mathbf{E}_r$ point in opposite directions (Fig. 5.13(b)). It is then clear that, for normal incidence if we assume the $\mathbf{E}_i$ and $\mathbf{E}_r$ pointing in the same direction the reflection coefficient is same as that given by Eqn (5.68). We, therefore, have for normal incidence,

$$\text{Reflection coefficient}: \Gamma = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \tag{5.70}$$

$$\text{Transmission coefficient}: \tau = \frac{2\eta_2}{\eta_2 + \eta_1} \tag{5.71}$$

The case of normal incidence can also be analysed in a different way. For normal incidence since $\theta_i = \theta_t = 0$ the electric fields for the three waves can be written as (put $\theta_i = \theta_t = 0$ in Eqns (5.43) to (5.45))

$$\mathbf{E}_i = \mathbf{E}_{i0} e^{-j\beta_1 z} \tag{5.72}$$

$$\mathbf{E}_r = \mathbf{E}_{r0} e^{+j\beta_1 z} \tag{5.73}$$

$$\mathbf{E}_t = \mathbf{E}_{t0} e^{-j\beta_2 z} \tag{5.74}$$

$$|\mathbf{H}_i| = \frac{|\mathbf{E}_i|}{\eta_1} = \frac{|\mathbf{E}_{i0}|}{\eta_1} e^{-j\beta_1 z} \tag{5.75}$$

$$|\mathbf{H}_r| = \frac{|\mathbf{E}_r|}{\eta_1} = \frac{|\mathbf{E}_{r0}|}{\eta_1} e^{+j\beta_1 z} \tag{5.76}$$

$$|\mathbf{H}_t| = \frac{|\mathbf{E}_t|}{\eta_2} = \frac{|\mathbf{E}_{t0}|}{\eta_2} e^{-j\beta_2 z} \tag{5.77}$$

Equations (5.72) to (5.77) are identical to the transmission line equations with E replacing V, H replacing I and η_1 and η_2 replaced by the characteristic impedances of the two media respectively. One can, therefore, conveniently analyse the normal incidence case using transmission line concept.

Figure 5.14 shows the equivalence of the two cases. Since, line 2 is infinitely long towards the right, at the junction it appears like an impedance η_2 (input impedance of an infinitely long line is always the characteristic impedance). The voltage reflection coefficient on line 1 is then

$$\Gamma = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \tag{5.78}$$

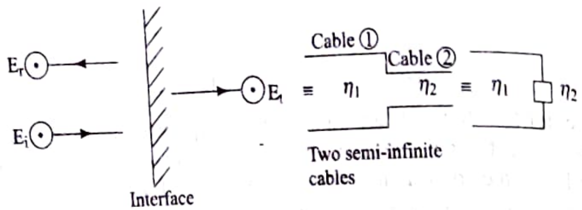


Fig. 5.14 *Analogy between normal incidence of a uniform plane wave at a media interface and the transmission line.*

which is same as Eqn (5.68) and the transmission coefficient is

$$\tau = \frac{2\eta_2}{\eta_2 + \eta_1} \tag{5.79}$$

The utility of transmission line analogy cannot be appreciated until we analyse the propagation of a plane wave through a multi-layer medium. In the following sections, we analyze normal incidence of a plane wave in a layered medium using wave as well as transmission line model. To develop the concepts clearly, we take only a three media problem here and leave the multi-media problem to the interested readers.

EXAMPLE 5.5 A uniform plane wave with 25 V/m electric field is normally incident on an infinitely thick slab of a material of dielectric constant 5. Find the electric and magnetic fields just inside the slab surface. How much power penetrates the material slab?

Solution:

The wave is incident from air to the material. Hence, we have

$$\eta_1 = \eta_0$$

$$\eta_2 = \frac{\eta_0}{\sqrt{5}}$$

The transmission coefficient

$$\tau = \frac{2\eta_2}{\eta_2 + \eta_1}$$

$$= \frac{\frac{2}{\sqrt{5}}}{\frac{1}{\sqrt{5}} + 1} = 0.618$$

The electric field inside the slab $E_t = \tau E_i = 0.618 \times 25 = 15.45$ V/m. The magnetic field inside the slab

$$H_t = \frac{E_t}{\eta_2} = 0.091 \text{ A/m}$$

The power transferred to the slab is $P = E_t H_t = 1.416 \text{ W/m}^2$.

5.4 NORMAL INCIDENCE ON A LAYERED MEDIUM

Let us consider a layered three media geometry as shown in Fig. 5.15. Media 1 and 2 are semi-infinite in size whereas media 2 is of finite thickness d. The parameters for media 1, 2 and 3 are denoted by suffices 1, 2, 3 respectively. Let us consider a plane wave with electric field $\mathbf{E}_i$ normally incident on the interface (1/2) from medium 1. Since, the electric and magnetic fields are oriented along the coordinate axes, without writing explicitely we will assume that all **E** vectors are along $+y$ direction and **H** vectors are along $\pm x$-directions depending upon the direction of the wave vector. If the wave is moving along $+z$ direction the H will be along $-x$ direction and if the wave is moving along $-z$ direction the H vector will be in $+x$ direction. (Note, that these signs are chosen to give correct directions of the Poynting vectors for each wave).

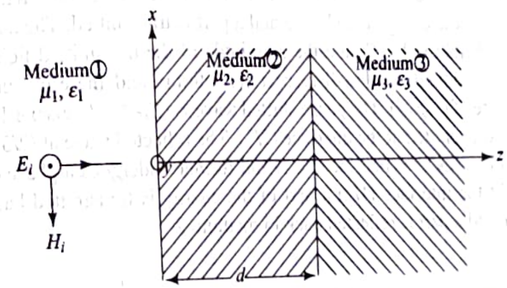


Fig. 5.15 *Normal incidence of a uniform plane wave at a layered medium.*

In the following analysis, we will drop the vector notation for the writing simplicity. The incident electric and magnetic fields, therefore, can be written as

$$E_i = E_{i0} e^{-j\beta_1 z} \tag{5.80}$$

$$H_i = \frac{E_{i0}}{\eta_1} e^{-j\beta_1 z} \tag{5.81}$$

Our primary interest now is to find:

(i) Fields for reflected wave in region 1.

(ii) Fields for transmitted wave in region 3.

As mentioned above, the problem can be solved in two ways (i) by multiple reflections and transmissions at the two interfaces (1/2) and (2/3) (ii) by transmission line analogy.

5.4.1 Multiple Reflections/Transmissions of Waves

Let us first create a notation for the transmission and reflection coefficients at the interfaces in forward as well as backward directions. Let Γ_{ij} represent the reflection coefficient at interface (i/j) for a wave incident from medium i. Similarly, let τ_{ij} represent the transmission coefficient from medium i to medium j. For example Γ_{12} will represent reflection coefficient at interface (1/2) for a wave incident from medium 1 and τ_{32} will represent transmission coefficient from medium 3 to medium 2 for a wave travelling backwards in medium 3.

From Eqns (5.78) and (5.79) we derive

$$\Gamma_{ij} = \frac{\eta_j - \eta_i}{\eta_j + \eta_i} \tag{5.82}$$

and

$$\tau_{ij} = \frac{2\eta_j}{\eta_j + \eta_i} \tag{5.83}$$

From Eqn (5.82) it is clear that $\Gamma_{ij} = -\Gamma_{ji}$. Now let a wave with electric field E_i be normally incident on the interface (1/2) from medium 1 side. At the interface (1/2) a part of the energy is reflected and a part is transmitted. The amplitude of the reflected field is $\Gamma_{12}E_i$ and the amplitude of the transmitted field is $\tau_{12}E_i$. The transmitted field travels a distance d in medium 2 and, therefore, undergoes a phase change of $(-\beta_2 d)$. The phase shifted wave $\tau_{12}E_i e^{-j\beta_2 d}$ gets partly reflected and partly transmitted at the interface (2/3). The reflected wave at (2/3) interface travels backwards upto interface (1/2) and further undergoes a phase change of $(-\beta_2 d)$. At the interface (1/2) a part of the energy is transmitted backward in medium 1 and a part is reflected back in medium 2.

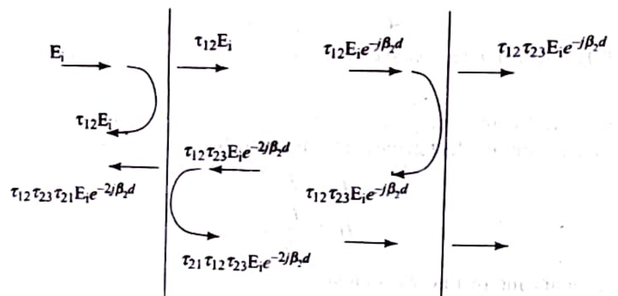


Fig. 5.16 *Multiple reflections of uniform plane waves in a layered medium.*

This process continues for infinite reflections and transmissions. In steady state, then we have a reflected wave in medium 1 which is a superposition of first reflected wave $\Gamma_{12}E_i$ and subsequent backward transmitted waves. In medium 2, we see two sets of waves one travelling from left to right and other from right to left. The total field at any point is equal to the sum of all the forward and backward travelling waves. In region 3 we have superposition of transmitted waves travelling from left to right.

From Fig. 5.16 we can write reflected field in medium 1 as

$$E_r = \Gamma_{12}E_i + \tau_{12}\Gamma_{23}\tau_{21}E_i e^{-j2\beta d} + \tau_{12}\Gamma_{23}\Gamma_{21}\Gamma_{23}\tau_{21}E_i e^{-j4\beta_2 d} + \tau_{12}\Gamma_{23}\Gamma_{21}\Gamma_{23}\Gamma_{21}\Gamma_{23}\tau_{21}E_i e^{-j6\beta_2 d} + \cdots \tag{5.84}$$

i.e.

$$E_r = \Gamma_{12}E_i + \tau_{12}\Gamma_{23}\tau_{21}e^{-j2\beta_2 d}E_i\{1 + \Gamma_{21}\Gamma_{23}e^{-j2\beta_2 d} + (\Gamma_{21}\Gamma_{23})^2 e^{-j4\beta_2 d} + \cdots\} \tag{5.85}$$

The terms in {} form a geometric series with progression ratio $\Gamma_{21}\ \Gamma_{23}\ exp(-j2\beta_2 d)$. Since $|\Gamma_{21}|$ and $|\Gamma_{23}|$ are less than 1, and $|exp(-j2\beta_2 d)| = 1$ the series is convergent and we can write its sum. The reflected field and the reflection coefficient, therefore, can be written as

$$E_r = \Gamma_{12}E_i + \tau_{12}\Gamma_{23}\tau_{21}e^{-j2\beta_2 d}\frac{E_i}{1 - \Gamma_{12}\Gamma_{23}e^{-j2\beta_2 d}} \tag{5.86}$$

giving

$$\Gamma \equiv \frac{E_r}{E_i} = \Gamma_{12} + \frac{\tau_{12}\Gamma_{23}\tau_{21}e^{-j2\beta_2 d}}{1 - \Gamma_{12}\Gamma_{23}e^{-j2\beta_2 d}} \tag{5.87}$$

Similarly, from Fig. 5.16 we can write the transmitted wave in medium 3 as

$$E_t = \tau_{12}\tau_{23}e^{-j\beta_2 d}E_i + \tau_{12}\Gamma_{23}\Gamma_{21}\tau_{23}e^{-j3\beta_2 d}E_i + \tau_{12}\Gamma_{23}\Gamma_{21}\Gamma_{23}\Gamma_{21}\tau_{23}e^{-j5\beta_2 d}E_i + \cdots$$

$$= \tau_{12}\tau_{23}e^{-j\beta_2 d}E_i\{1 + \Gamma_{23}\Gamma_{21}e^{-j2\beta_2 d} + (\Gamma_{21}\Gamma_{23})^2 e^{-j4\beta_2 d}\ldots\} \tag{5.88}$$

Again the terms in curly brackets form the same geometric series as that in Eqn (5.85), and we get the transmission coefficient as

$$\tau \equiv \frac{E_t}{E_i} = \frac{\tau_{12}\tau_{23}e^{-j\beta_2 d}}{1 - \Gamma_{21}\Gamma_{23}e^{-j2\beta_2 d}} \tag{5.89}$$

The procedure explained above can be extended to any number of layers with of course increased algebraic complexity.

5.4.2 Transmission Line Analogy

Let us now try to solve the problem of wave propagation in a multi-layer medium using the transmission line analogy. The geometry in Fig. 5.16 is equivalent to three transmission lines of characteristic impedances η_1, η_2, η_3 respectively.

The transmission lines 1 and 3 are infinitely long and the transmission line 2 is of length d. Also note that the phase constants of the three lines are β_1, β_2, β_3 respectively as shown in Fig. 5.17.

Now, since line 3 is infinitely long, it sees an impedance η_3 (characteristic impedance of line 3) at the junction (2/3). This impedance can be transformed to

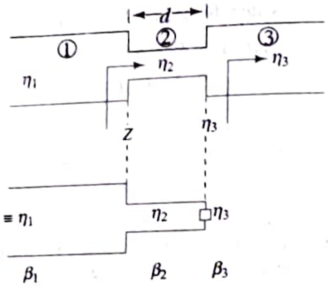


Fig. 5.17 *Transmission line analogy for a layered medium.*

an impedance (say) Z at (1/2) junction as

$$Z = \eta_2 \left[\frac{\eta_3 \cos \beta_2 d + j\eta_2 \sin \beta_2 d}{\eta_2 \cos \beta_2 d + j\eta_3 \sin \beta_2 d} \right] \tag{5.90}$$

Note that the characteristic impedance of the transforming line 2 is η_2.

The reflection coefficient on the line 1 then can be written as

$$\Gamma = \frac{Z - \eta_1}{Z + \eta_1} \tag{5.91}$$

$$= \frac{\eta_2(\eta_3 \cos \beta_2 d + j\eta_2 \sin \beta_2 d) - \eta_1(\eta_2 \cos \beta_2 d + j\eta_3 \sin \beta_2 d)}{\eta_2(\eta_3 \cos \beta_2 d + j\eta_2 \sin \beta_2 d) - \eta_1(\eta_2 \cos \beta_2 d + j\eta_3 \sin \beta_2 d)} \tag{5.92}$$

Equations (5.87) and (5.92) although look algebraically different, are infact the same. It is only a matter of algebraic manipulation to reduce one to the other.

The magnitude of the transmission coefficient can be obtained by applying conservation of power.

The electric field E_i in medium 1 has a power density of $|E_i|^2/\eta_1$. The power density of the reflected wave will be $|\Gamma E_i|^2/\eta_1$. Since, all the media are lossless, the difference of the two power densities is equal to the power density of the transmitted wave in medium 3, $|E_t|^2/\eta_3$, giving

$$\frac{|E_i|^2}{\eta_1} - \frac{|\Gamma E_i|^2}{\eta_1} = \frac{|E_t|^2}{\eta_3} = \frac{|\tau E_i|^2}{\eta_3} \tag{5.93}$$

$$\Rightarrow \qquad |\tau| = \sqrt{\frac{|E_t|^2}{|E_i|^2}} = \sqrt{\frac{\eta_3}{\eta_1}(1 - |\Gamma|^2)} \tag{5.94}$$

The above analysis clearly shows that the pr.blem of normal incidence on a multi-layer medium can be elegently solved u..ing the transmission line concepts.

The analysis of layered medium f...s many practical applications. For example, one may be interested in sending electromagnetic energy efficiently from one medium to another, or one may be interested in providing a protective sheet

of a material around a radiating instrument. In both type of applications, of course the primary objective is the efficient power transfer with minimal reflections. One can then ask, "under what condition does the reflected power reduces to zero", i.e. the reflection coefficient goes to zero.

From transmission line view point the answer to this question is rather simple, 'Use a quarter wavelength transformer to match two transmission lines'. We can get the same result by equating the reflection coefficient Γ in Eqn (5.92) to zero, giving

$$\eta_2(\eta_3 \cos \beta_2 d + j\eta_2 \sin \beta_2 d) = \eta_1(\eta_2 \cos \beta_2 d + j\eta_3 \sin \beta_2 d) \tag{5.95}$$

Separation of real and imaginary parts yields

$$\text{Real:} \qquad \eta_2(\eta_3 - \eta_1) \cos \beta_2 d = 0 \tag{5.96}$$

$$\text{Imag:} \qquad (\eta_2^2 - \eta_1\eta_3) \sin \beta_2 d = 0 \tag{5.97}$$

Since, both sin and cos cannot simultaneously become zero for any angle, we have two conditions under which there is no reflection.

1. $\sin \beta_2 d = 0$ **and** $\eta_3 = \eta_1$ This implies

$$\beta_2 d = m\pi \qquad m = 1, 2, 3 \cdots$$

$$\Rightarrow \qquad d = \frac{m\pi}{\beta_2} = \frac{m\pi}{2\pi/\lambda_2} = \frac{m\lambda_2}{2} \tag{5.98}$$

λ_2 is the wavelength in medium 2. ($\lambda_2 = 1/f\sqrt{\mu_2\epsilon_2}$)

This condition suggests that if a slab of any material which is multiples of half wavelength thick is inserted normally in the path of an electromagnetic wave, there is no reflection, i.e. the propagation on either side of the slab remains unaffected. This is interesting and useful, as it suggests that if an antenna is covered with a half wavelength thick sheet, the flow of the radiated wave is unaffected. This is the principle of radom design. Radom is a cover which protects an antenna from the weather. Since, the purpose of radom is mechanical, it should affect the wave propagation in the least possible fashion. Multiple dielectric layers can be used for designing proper radoms.

2. $\cos \beta_2 d = 0$ **and** $\eta_2^2 = \eta_1\eta_3$. This is the famous quarter wavelength transformer. The conditions imply

$$\beta_2 d = (2m+1)\frac{\pi}{2} \tag{5.99}$$

$$\Rightarrow \qquad d = (2m+1)\frac{\lambda_2}{4} \qquad m = 1, 2, 3 \cdots \tag{5.100}$$

$$\text{and} \qquad \eta_2 = \sqrt{\eta_1\eta_3} \tag{5.101}$$

This condition suggests that an electromagnetic wave can be fully transmitted across a media-interface without reflection by inserting a sheet of matching medium which is quarter wavelength thick and has intrinsic impedance equal to the geometric mean of the intrinsic impedances of the two media. This technique is frequently used in realizing anti-reflecting coatings in optical components.

EXAMPLE 5.6 A uniform plane wave having 10 W/m^2 power density is normally incident on a 5 cm thick dielectric sheet with $\epsilon_r = 9$. If the frequency of the wave is 1 GHz. Find the power density of the wave transmitted through the sheet.

Solution:

Intrinsic impedance of the dielectric is

$$\eta_d = \frac{\eta_0}{\sqrt{9}} = \frac{\eta_0}{3}$$

$$\Gamma_{12} = \frac{\eta_d - \eta_0}{\eta_d + \eta_0} = \frac{\frac{\eta_0}{3} - \eta_0}{\frac{\eta_0}{3} + \eta_0} = -\frac{1}{2}$$

$$\Gamma_{23} = -\Gamma_{12} = \frac{1}{2}$$

$$\tau_{12} = \frac{2\eta_d}{\eta_d + \eta_0} = \frac{1}{2}$$

$$\tau_{23} = \frac{2\eta_0}{\frac{\eta_0}{3} + \eta_0} = \frac{3}{2}$$

$$\beta_2 = \omega\sqrt{\mu_0\epsilon_0\epsilon_r} = \frac{2\pi \times 10^9}{3 \times 10^8} \times \sqrt{9} = 20\pi \text{ rad/m}$$

$$\beta_2 d = 20\pi(0.05) = \pi$$

The transmission coefficient from Eqn (5.89) is

$$\tau = \frac{(\frac{1}{2})(\frac{3}{2})e^{-j\pi}}{1 - (-\frac{1}{2})(\frac{1}{2})e^{-j2\pi}} = \frac{-\frac{3}{4}}{\frac{5}{4}} = -\left(\frac{3}{5}\right)$$

Power density of the transmitted wave

$$= |\tau|^2 \times \text{Power density of the incident wave}$$

$$= \frac{9}{25} \times 10 = \frac{18}{5} \text{ W/m}^2$$

5.5 TOTAL INTERNAL REFLECTION

Let us now investigate an interesting case of the wave propagation known as the 'total internal reflection (TIR)'.

From Eqns (5.38) and (5.58) we note that if

$$\frac{\beta_1}{\beta_2}\sin\theta_i \geq 1 \tag{5.102}$$

the angle of transmission does not exist, since $\sin\theta_t > 1$ and $\cos\theta_t$ is imaginary. It means that, if a plane wave is launched at a media interface at an angle which satisfies Eqn (5.102) there is no transmitted wave. The angle for which the equality in Eqn (5.102) holds is called the *critical angle*, θ_c. This is the angle of incidence $(\theta_i = \theta_c)$ at which

$$\sin\theta_t = \frac{\beta_1}{\beta_2}\sin\theta_i = 1 \tag{5.103}$$

$$\Rightarrow \qquad \theta_t = \frac{\pi}{2} \tag{5.104}$$

meaning that the transmitted wave travels just along the interface. The critical angle is therefore given by

$$\sin\theta_c = \frac{\beta_2}{\beta_1}$$

For $\theta > \theta_c$, the reflection coefficients for the two polarizations (prependicular and parallel) are re-written such that the quantity in the square root sign is positive, giving

$$\Gamma_{\parallel} = \frac{\eta_1\cos\theta_i - j\eta_2\sqrt{\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i - 1}}{\eta_1\cos\theta_i + j\eta_2\sqrt{\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i - 1}} \tag{5.105}$$

$$\Gamma_{\perp} = \frac{\eta_2\cos\theta_i - j\eta_1\sqrt{\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i - 1}}{\eta_2\cos\theta_i + j\eta_1\sqrt{\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i - 1}} \tag{5.106}$$

where we have written $\cos\theta_t$ explicitly in terms of θ_i, and have taken -1 common within the square root sign. It is now easy to note that both reflection coefficients are of the form

$$\frac{(a - jb)}{(a + jb)}, \qquad a, b \text{ are real,}$$

therefore, their magnitudes are unity. That is, for $\theta > \theta_c$, $|\Gamma_\perp| = 1$ and $|\Gamma_\parallel| = 1$. The reflection coefficients, however, have different phase angles. The phase angles for the two polarizations are

$$\phi_{\parallel} = -2\tan^{-1}\left(\frac{\eta_2\sqrt{\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i - 1}}{\eta_1\cos\theta_i}\right) \tag{5.107}$$

$$\phi_{\perp} = -2\tan^{-1}\left(\frac{\eta_1\sqrt{\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i - 1}}{\eta_2\cos\theta_i}\right) \tag{5.108}$$

Since, the magnitudes of the reflection coefficients are unity, the entire incident power is reflected back into medium 1. This phenomenon is therefore called the **Total Internal Reflection (TIR)**.

There are certain important things which are worth mentioning about the total internal reflection.

1. The total internal reflection can take place only if the wave travels from denser to rarer medium. From Eqn (5.102) we can see that the angle of incidence for which the total internal reflection can occur is given by

$$\sin\theta_i \geq \frac{\beta_2}{\beta_1} = \frac{\omega\sqrt{\mu_2\epsilon_2}}{\omega\sqrt{\mu_1\epsilon_1}} = \frac{\sqrt{\mu_2\epsilon_2}}{\sqrt{\mu_1\epsilon_1}} \tag{5.109}$$

Since, $\sin\theta_i$ has to be ≤ 1, for TIR to take place we must have

$$\mu_2\epsilon_2 \leq \mu_1\epsilon_1 \tag{5.110}$$

i.e. medium 2 is rarer compared to medium 1.

A special case of this is an ideal dielectric interface, for which $\mu_1 = \mu_2 = \mu_0$, and

$$\epsilon_1 = \epsilon_0\epsilon_{r1} = \epsilon_0 n_1^2$$

$$\epsilon_2 = \epsilon_0\epsilon_{r2} = \epsilon_0 n_2^2$$

n_1 and n_2 are refractive indices of the two dielectric media. Condition Eqn (5.110) then reduces to the well known condition

$$n_2 \leq n_1 \tag{5.111}$$

2. Wave undergoes a phase change during total internal reflection From Eqn (5.105) and (5.106) it is clear that the reflection coefficient has a phase which implies that the reflected wave lags with respect to the incident wave. The phase change depends upon the medium parameters as well as the angle of incidence. At critical angle the phase lag in the reflected wave is zero (quantity in the square root sign goes to zero) and it increases as the angle of incidence increases beyond θ_c. It is also important to note that the two polarizations, parallel and perpendicular, undergo different phase changes at total internal reflection. In

the later sections, this differential phase change for the two polarizations will be exploited to alter the state of polarization of an electromagnetic wave.

3. At TIR the fields do not vanish in the second medium. Substituting Eqn (5.58) with TIR condition Eqn (5.102) in Eqn (5.54) and Eqn (5.66) we find that even when $|\Gamma_{\perp}|$ and $|\Gamma_{\parallel}|$ become unity, the transmission coefficients $\tau_{\parallel}$ and $\tau_{\perp}$ do not become zero. In other words, total reflection of power in medium 1, i.e. no transmission of power to medium 2, does not mean no transmission of the fields to medium 2. The fields can very well exist in the second medium provided they do not constitute any power flow. These fields are called the 'evanescent fields'. Distribution of these fields in medium 2 can be readily obtained by substituting for θ_t in the expression for the transmitted field Eqn (5.45). Substituting

$$\sin\theta_t = \frac{\beta_1}{\beta_2}\sin\theta_i \tag{5.112}$$

and

$$\cos\theta_t = j\sqrt{\left(\frac{\beta_1^2}{\beta_2^2}\right)\sin^2\theta_i - 1} \tag{5.113}$$

in Eqn (5.45) we get

$$\begin{aligned}\mathbf{E}_t &= \mathbf{E}_{t0}e^{-j\beta_2\left\{\frac{\beta_1}{\beta_2}x\sin\theta_i \pm jz\sqrt{\frac{\beta_1^2}{\beta_2^2}(\sin^2\theta_i - 1)}\right\}} \\ &= \mathbf{E}_{t0}e^{-j\left\{\beta_1 x\sin\theta_i \pm jz\sqrt{\beta_1^2\sin^2\theta_i - \beta_2^2}\right\}} \\ &= \mathbf{E}_{t0}\underbrace{e^{-j\beta_1 x\sin\theta_i}}_{I}\underbrace{e^{\mp z\sqrt{\beta_1^2\sin^2\theta_i - \beta_2^2}}}_{II}\end{aligned} \tag{5.114}$$

In Eqn (5.114), the exponential term I represents a travelling wave in x-direction. Term II however represents only an exponentially varying field in the z-direction and not a wave as there is no phase variation along z-direction. Since, the field cannot grow indefinitely away from the interface, we should choose the −ve sign for the exponent which represents exponentially decaying fields away from the interface in medium 2. The field in medium 2, therefore, is correctly written as

$$\mathbf{E}_t = \mathbf{E}_{t0}e^{-j\beta_1 x\sin\theta_i}e^{-z\sqrt{\beta_1^2\sin^2\theta_i - \beta_2^2}} \tag{5.115}$$

Note that, the exponential decay constant in z-direction is a function of the angle of incidence besides the media parameters. Figure 5.17 shows the field variation as function of distance from the interface for different angles of incidence.

At the critical angle, the decay constant is zero and the field is constant along the z-direction in medium 2. As the angle of incidence increases beyond the critical angle, the field gets more confined to the interface. However, it is important to note that in no circumstance the field in medium 2 goes to zero. The boundary conditions demand that the fields must be continuous at the interface. Since, the

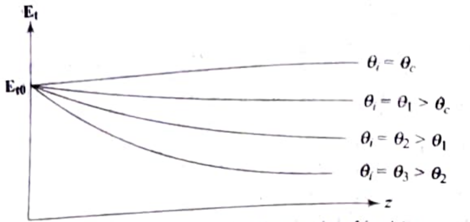


Fig. 5.18 *Electric field variation in medium 2 for angle of incidence greater than or equal to the critical angle.*

fields in medium 1 at the interface are non-zero, the fields in medium 2 are also finite at the interface. For TIR, where the energy is totally reflected in medium 1, one should not undermine the importance of the evanescent fields in medium 2. Though, these fields do not constitute any power flow, there presence is as important as the fields in medium 1 which have power flow associated with them. Any disturbance to the evanescent fields eventually disturbs the total internal reflection.

EXAMPLE 5.7 A uniform plane wave travelling in a dielectric of refractive index 3 is incident at the dielectric air interface. The angle of incidence is 60°. The electric field of incident wave is 10 V/m. Find the phase velocity of the field and the magnitude of the field in air at a distance of 0.5 cm from the dielectric air interface. Assume perpendicular polarization for the wave. Frequency of wave is 10 GHz.

Solution:

The critical angle at the interface is

$$\theta_c = \sin^{-1}\left(\frac{\sqrt{\epsilon_2}}{\sqrt{\epsilon_1}}\right) = \sin^{-1}\left(\frac{1}{n}\right)$$

$$= \sin^{-1}\left(\frac{1}{3}\right) = 19.47^\circ$$

Since, the angle of incidence is 60°, the wave is total internally reflected at the interface. In dielectric medium then we have superposition of the incident and the reflected wave and in air we have exponentially decaying fields.

From Eqn (5.106) reflection coefficient can be written as

$$\Gamma_\perp = \frac{\eta_2 \cos\theta_i - j\eta_1\sqrt{\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i - 1}}{\eta_2 \cos\theta_i + j\eta_1\sqrt{\frac{\beta_1^2}{\beta_2^2}\sin^2\theta_i - 1}}$$

$$= \frac{\sqrt{\frac{\epsilon_1}{\epsilon_2}}\cos\theta_i - j\sqrt{\frac{\epsilon_1}{\epsilon_2}\sin^2\theta_i - 1}}{\sqrt{\frac{\epsilon_1}{\epsilon_2}}\cos\theta_i + j\sqrt{\frac{\epsilon_1}{\epsilon_2}\sin^2\theta_i - 1}}$$

$$= \frac{3(\frac{1}{2}) - j\sqrt{9(\frac{\sqrt{3}}{2})^2 - 1}}{3(\frac{1}{2}) + j\sqrt{9(\frac{\sqrt{3}}{2})^2 - 1}}$$

$$= -0.4375 - 0.8922j$$

Hence, total electric field in dielectric medium is given by

$$E_1 = E_i + \Gamma_\perp E_i$$

$$= (0.5625 - 0.8922j) \times 10$$

$$\Rightarrow \quad \text{Peak amplitude} = 11.31 \text{ V/m}$$

In the air, electrical field decays as we go away from the interface. In air, β is along the x-axis. New β_x will be given by

$$\beta_x = \beta_1 \sin\theta_i$$

Hence, $$v_p = \frac{\omega}{\beta_1 \sin\theta_i} = \frac{c}{\sqrt{\epsilon_{r1}}\sin\theta_i}$$

$$= \frac{3\times 10^8}{3} \times \left(\frac{2}{\sqrt{3}}\right) = 1.155 \times 10^8 \text{ m/s}$$

$$\beta_2 = \frac{2\pi \times 10^{10}}{3\times 10^8} = \frac{200\pi}{3}$$

$$\beta_1 \sin\theta_i = \frac{2\pi \times 10^{10}}{3 \times 10^8} \times 3 \times \frac{\sqrt{3}}{2} = 100\sqrt{3}\pi$$

Electic field in air is given by Eqn (5.115)

$$E_{air} = 10 \times e^{-5\times 10^{-3}\sqrt{(100\sqrt{3}\pi)^2 - (200\pi/3)^2}}$$

$$= 0.811 \text{ V/m}$$

5.6 WAVE POLARIZATION AT MEDIA INTERFACE

In Section 4.5, we have shown that any arbitary polarization can be decomposed into any two orthogonal states of polarization. We can, therefore, analyse the polarization of the reflected and the transmitted waves at an interface by decomposing the incident electric field into its components, one in the plane of incidence and other perpendicular to it. In other words, we decompose the polarization of the incident wave into two orthogonal linear polarizations, one in the plane of incidence and other normal to it, find reflection and transmission for the two polarizations separately and combine them. Let us, therefore, write the incident electric field as

$$\mathbf{E}_i = \mathbf{E}_{i\parallel} + \mathbf{E}_{i\perp} e^{j\phi} \tag{5.116}$$

Note that locally the coordinate system is oriented such that $\mathbf{E}_{i\parallel}$ is along x'-axis and $\mathbf{E}_{i\perp}$ is along y'-axis and z' is the direction of the wave vector (see Fig. 5.19).

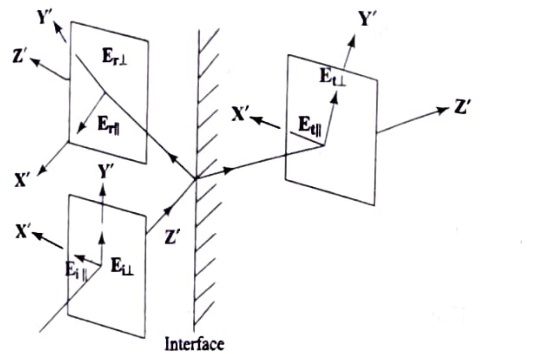


Fig. 5.19 *Polarization of incident, reflected and transmitted plane waves at media interface.*

If the reflection and transmission coefficients for parallel and perpendicular polarizations are denoted by $\Gamma_\parallel, \tau_\parallel$ and $\Gamma_\perp, \tau_\perp$ respectively, the reflected and transmitted electric fields are

$$\mathbf{E}_r = \mathbf{E}_{r\parallel} + \mathbf{E}_{r\perp} = \Gamma_\parallel \mathbf{E}_{i\parallel} + \Gamma_\perp \mathbf{E}_{i\perp} e^{j\phi} \tag{5.117}$$

$$\mathbf{E}_t = \mathbf{E}_{t\parallel} + \mathbf{E}_{t\perp} = \tau_\parallel \mathbf{E}_{i\parallel} + \tau_\perp \mathbf{E}_{i\perp} e^{j\phi} \tag{5.118}$$

since the reflection and transmission coefficients are different for parallel and perpendicular polarizations, we can say that in general the states of polarization of the reflected and transmitted waves are not same as that of the incident wave. However, it is worthwhile to investigate here a few simple but important cases.

5.6.1 Change in Polarization at Simple Reflection

1. Linearly Polarized Incident Wave If the incident wave is linearly polarized, $\mathbf{E}_{i\parallel}$ and $\mathbf{E}_{i\perp}$ are in phase, i.e. $\phi = 0$. For simple reflection (not total internal reflection) since the reflection coefficients $\Gamma_\parallel$ and $\Gamma_\perp$ are real (though they could be positive or negative), the components of the reflected wave, $\mathbf{E}_{r\parallel}$ and $\mathbf{E}_{r\perp}$ are either in phase or 180° out of phase (depending upon the sign of $\Gamma_\parallel$ and $\Gamma_\perp$). The polarization of the reflected wave hence remains linear. The orientation

of the linear polarization however will be different since

$$\frac{|\mathbf{E}_{r\parallel}|}{|\mathbf{E}_{r\perp}|} \neq \frac{|\mathbf{E}_{i\parallel}|}{|\mathbf{E}_{i\perp}|}$$

The argument is equally applicable to the transmitted wave as well since $\tau_\parallel$ and $\tau_\perp$ are also real in this case.

We, therefore, conclude that a linearly polarized wave remains linearly polarized after simple reflection but the plane of polarization changes as shown in Fig. 5.20.

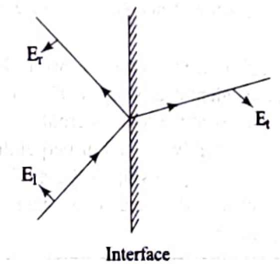


Fig. 5.20 *Linearly polarized incident wave.*

2. Circularly Polarized Incident Wave For a circularly polarized incident wave, we have

$$|\mathbf{E}_{i\parallel}| = |\mathbf{E}_{i\perp}|$$

$$\phi = \pm\frac{\pi}{2}$$

Since the reflection and transmission coefficients are real, the reflected (and also transmitted) wave has a phase difference of $\pm\pi/2$ between its components. However, the two components no more have same magnitudes, since $\Gamma_\parallel \neq \Gamma_\perp$ and $\tau_\parallel \neq \tau_\perp$. The reflected and transmitted waves therefore become elliptically polarized with the major axis of each ellipse either lying in the plane of incidence or perpendicular to it as shown in Fig. 5.21.

Depending upon the magnitudes of the reflection coefficients, the polarization ellipse may degenerate into a line. A circular polarization, therefore, may change even to the linear polarization after reflection. This aspect will be discussed later in Section 5.7.

5.6.2 Polarization Change at TIR

For the total internal reflection, there is no transmitted wave, so we need to discuss the polarization of the reflected wave only. For the TIR the magnitudes of the two reflection coefficients $\Gamma_\parallel$ and $\Gamma_\perp$, are unity but their phases are different.

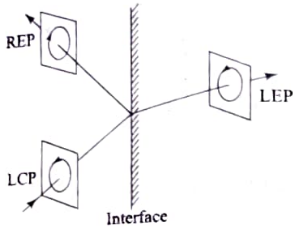


Fig. 5.21 *Circularly polarized incident wave at a dielectric boundary.*

Consequently, the phase difference between the two reflected components is not same as that between the incident components. Then it is evident that the state of polarization always changes at the total internal reflection, except when the incident polarization is either purely parallel or perpendicular. We can also note that the change in polarization depends upon the angle of incidence. The TIR can, therefore, be cleverly exploited to change the state of polarization of a wave without affecting its magnitude.

EXAMPLE 5.8 A circularly polarized plane wave is incident on a thick glass slab at an angle of incidence of 45°. Find the state of polarization of the reflected and transmitted waves. The frequency of the wave is 10^{14} Hz and the refractive index of glass is 1.5.

Solution:

Since, the incident wave is circularly polarized, the parallel and perpendicular components of the electric field are equal in magnitude but 90° out of phase. Let the perpendicular component be leading the parallel component, that is

$$E_{i\perp} = E_{i\parallel} e^{+j\frac{\pi}{2}} = jE_{i\parallel}$$

$$|E_{i\perp}| = |E_{i\parallel}| = A$$

From snell's Law

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

$$\Rightarrow \quad \sin\theta_t = \frac{n_1}{n_2}\sin\theta_i = \frac{1}{1.5}\sin 45^\circ = 0.4714$$

$$\Rightarrow \quad \cos\theta_t = 0.8819$$

Also we have

$$\eta_1 = \sqrt{\frac{\mu_0}{\epsilon_0}} = \eta_0$$

$$\eta_2 = \sqrt{\frac{\mu_0}{\epsilon_0\epsilon_r}} = \sqrt{\frac{\mu_0}{\epsilon_0}} \cdot \frac{1}{\sqrt{\epsilon_r}} = \frac{\eta_0}{n_2}$$

The reflection and transmission coefficients for perpendicular and parallel polarizations are (from Eqns (5.53), (5.54), (5.65) and (5.66))

$$\Gamma_\perp = \frac{\frac{\eta_0}{n_2}\cos 45^\circ - \eta_0\cos\theta_t}{\frac{\eta_0}{n_2}\cos 45^\circ + \eta_0\cos\theta_t}$$

$$= \frac{0.4714 - 0.8819}{0.4714 + 0.8819} = -0.3033$$

$$\tau_\perp = 1 + \Gamma_\perp = 0.6966$$

$$\Gamma_\parallel = \frac{\eta_0\cos 45^\circ - \frac{\eta_0}{n_2}\cos\theta_t}{\eta_0\cos 45^\circ + \frac{\eta_0}{n_2}\cos\theta_t}$$

$$= \frac{0.707 - \frac{0.8819}{1.5}}{0.707 + \frac{0.8819}{1.5}} = 0.09189$$

$$\tau_\parallel = \frac{2\frac{\eta_0}{n_2}\cos(45^\circ)}{\eta_0\cos(45^\circ) + \frac{\eta_0}{n_2}\cos\theta_t} = 0.728$$

The reflected and transmitted fields are

$$E_{r\parallel} = A\Gamma_\parallel \quad \text{and} \quad E_{r\perp} = jA\Gamma_\perp$$

$$\Rightarrow \frac{E_{r\perp}}{E_{r\parallel}} = \frac{j\Gamma_\perp}{\Gamma_\parallel} = \frac{-0.3033}{0.09189}\angle\pi/2 = -3.3\angle\pi/2$$

And

$$E_{t\parallel} = A\tau_\parallel \quad \text{and} \quad E_{t\perp} = jA\tau_\perp$$

$$\Rightarrow \frac{E_{t\perp}}{E_{t\parallel}} = \frac{j\tau_\perp}{\tau_\parallel} = \frac{0.6966}{0.728}\angle\pi/2 = 0.96\angle\pi/2$$

$$\mathbf{E}_t = A\tau_\parallel + jA\tau_\perp$$

The reflected wave is elliptically polarized and the transmitted wave is almost circularly polarized.

5.7 BREWSTER ANGLE

The Brewster angle is the angle of incidence for which there is no reflection, i.e. the reflection coefficient is zero. Equating (5.53) and (5.65) to zero we get

$$\text{For Perpendicular Polarization:} \quad \eta_2\cos\theta_{B\perp} - \eta_1\cos\theta_t = 0 \qquad (5.119)$$

$$\text{For Parallel Polarization:} \quad \eta_1\cos\theta_{B\parallel} - \eta_2\cos\theta_t = 0 \qquad (5.120)$$

where $\theta_{B\perp}$ and $\theta_{B\parallel}$ are the Brewster angles for perpendicular and parallel polarizations respectively.

Substitution for $\cos\theta_t$ from Eqn (5.58) in Eqn (5.119) and Eqn (5.120) yields

$$\eta_2 \cos\theta_{B\perp} = \eta_1\sqrt{1 - \frac{\beta_1^2}{\beta_2^2}\sin^2\theta_{B\perp}} \tag{5.121}$$

$$\eta_1 \cos\theta_{B\parallel} = \eta_2\sqrt{1 - \frac{\beta_1^2}{\beta_2^2}\sin^2\theta_{B\parallel}} \tag{5.122}$$

After a little algebraic manipulation we get

$$\tan\theta_{B\perp} = \frac{\beta_2}{\eta_1}\left(\frac{\eta_2^2 - \eta_1^2}{\beta_2^2 - \beta_1^2}\right)^{1/2} \tag{5.123}$$

$$\tan\theta_{B\parallel} = \frac{\beta_2}{\eta_2}\left(\frac{\eta_1^2 - \eta_2^2}{\beta_2^2 - \beta_1^2}\right)^{1/2} \tag{5.124}$$

Substituting $\beta_1 = \omega\sqrt{\mu_1\epsilon_1}$, $\beta_2 = \omega\sqrt{\mu_2\epsilon_2}$, $\eta_1 = \sqrt{\mu_1/\epsilon_1}$, and $\eta_2 = \sqrt{\mu_2/\epsilon_2}$, the Brewster angles for the two polarizations can be obtained as

$$\theta_{B\perp} = \tan^{-1}\left(\sqrt{\frac{\mu_2}{\mu_1}}\left[\frac{\mu_2\epsilon_1 - \mu_1\epsilon_2}{\mu_2\epsilon_2 - \mu_1\epsilon_1}\right]^{1/2}\right) \tag{5.125}$$

$$\theta_{B\parallel} = \tan^{-1}\left(\sqrt{\frac{\epsilon_2}{\epsilon_1}}\left[\frac{\mu_1\epsilon_2 - \mu_2\epsilon_1}{\mu_2\epsilon_2 - \mu_1\epsilon_1}\right]^{1/2}\right) \tag{5.126}$$

In principle, the Brewster angle can exist for both polarizations. However, for non-magnetic materials (dielectrics), $\mu_1 = \mu_2 = \mu_0$ (free-space permeability) and consequently the Brewster angle does not exist for the perpendicular polarization. This can be seen by substituting $\mu_1 = \mu_2 = \mu_0$ in Eqn (5.125). The quantity in the square bracket becomes negative and hence there is no real angle $\theta_{B\perp}$. The Brewster angle for the parallel polarization exists and is given by,

$$\theta_{B\parallel} = \tan^{-1}\left(\sqrt{\frac{\epsilon_2}{\epsilon_1}}\right) \tag{5.127}$$

Note: From Eqn (5.127) it might appear that when $\epsilon_1 = \epsilon_2$, i.e. when there is no medium discontinuity at the interface (in fact in that case there is no interface as there is only one medium), the Brewster angle is 45°. But this is absurd because, if there is no medium discontinuity, no-reflection condition should be true for all angles and not only for 45°. This absurdity, however, is purely mathematical. If $\epsilon_1 = \epsilon_2$, the quantity in square brackets of Eqn (5.126) is 0/0, i.e. $\theta_{B\parallel}$ is indeterminate and not 45°.

From the above discussion, it is clear that at the Brewster angle, a particular polarization has no reflection. Now, if the incident wave has some arbitrary polarization (which is a combination of parallel and perpendicular polarizations), after reflection at the Brewster angle the wave will have only one polarization component. If the angle of incidence is $\theta_{B\parallel}$, the reflected wave has only perpendicular polarization, and if the angle of incidence is $\theta_{B\perp}$, the reflected wave has only parallel polarization. In both the cases, the polarization of the reflected wave is linear (in the plane of incidence or perpendicular to it) irrespective of the polarization of the incident wave. The Brewster angle therefore is also called the 'Polarizing angle'. The Brewster angle concept is used to obtain linearly polarized light from an unpolarized one. In many applications, like optical interferometry, lasers, etc where the light has to be linearly polarized, the Brester angle concept can be conviniently used.

EXAMPLE 5.9 An elliptically polarized wave is incident on an air dielectric interface at the Brewster angle. The axial ratio of the polarization ellipse is 2 and the major axis of ellipse is perpendicular to the plane of incidence. Find the axial ratio and the orientation of the major axis of the reflected and transmitted waves.

Solution:

Since, the interface is dielectric interface, at the Brewster angle, parallel polarization is completely transmitted. The reflected wave has only perpendicular polarization. The reflected wave therefore has linear polarization perpendicular to the plane of incidence.

For transmitted wave, we have to obtain the field components.

The angle of incidence

$$\theta_i = \theta_{B\parallel} = \tan^{-1}\left(\sqrt{\frac{\epsilon_2}{\epsilon_1}}\right) = \tan^{-1}(n_2)$$

$$\Rightarrow \qquad \sin\theta_i = \frac{n_2}{\sqrt{1+n_2^2}}$$

$$\text{and} \qquad \cos\theta_i = \frac{1}{\sqrt{1+n_2^2}}$$

Using Snell's Law we get

$$\sin\theta_t = \frac{\sin\theta_i}{n_2} = \cos\theta_i$$

$$\text{and} \qquad \cos\theta_t = \sqrt{1-\sin^2\theta_t} = \sin\theta_i$$

For the ellipse since the AR is 2 we have,

$$E_{i\parallel} = A \qquad (say)$$

$$E_{i\perp} = 2E_{i\parallel}e^{j\pi/2} = 2Ae^{j\pi/2}$$

the phase difference between $E_{i\perp}$ and $E_{i\parallel}$ is $\pi/2$, since, the major axis is along the normal to the plane of incidence. Substituting $\cos\theta_i$ and $\cos\theta_t$ in expressions for the transmission coefficients, Eqns (5.54) and (5.66), we get

$$\tau_{\parallel} = \frac{1}{n_2}$$

and $$\tau_{\perp} = \frac{2}{1+n_2^2}$$

The transmitted fields are therefore,

$$E_{t\parallel} = \tau_{\parallel} E_{i\parallel} = \frac{A}{n_2}$$

$$E_{t\perp} = \tau_{\perp} E_{i\perp} = \frac{2Ae^{j\pi/2}}{1+n_2^2}$$

The phase between $E_{t\perp}$ and $E_{t\parallel}$ is still $\pi/2$ as was in the original wave. The transmitted wave, therefore, has same sense of rotation (LH or RH) as the incident wave. The ratio of the amplitudes of two components is

$$\frac{|E_{t\perp}|}{|E_{t\parallel}|} = \frac{\frac{2A}{(1+n_2^2)}}{A/n_2} = \frac{2n_2}{1+n_2^2}$$

Since, $n_2 \neq 1$ (because if $n_2 = 1$, there is no interface) $|E_{t\perp}| \neq |E_{t\parallel}|$ and consequently the transmitted wave also is elliptically polarized. If $|E_{t\perp}| > |E_{t\parallel}|$, i.e. if $2n_2 > 1 + n_2^2$, the major axis of the ellipse will be $\perp$ to the plane of incidence and AR will be $\frac{1+n_2^2}{2n_2}$.

However, if $|E_{t\perp}| < |E_{t\parallel}|$, i.e. if $2n_2 < 1 + n_2^2$, the major axis of the ellipse will be in the plane of incidence and AR will be $\frac{1+n_2}{2n_2}$.

For any medium n_2 is always greater than 1 and consequently $1 + n_2^2 > 2n_2$ always, i.e. we have, $|E_{t\parallel}| > |E_{t\perp}|$. The transmitted wave therefore is elliptically polarized with major axis in the plane of incidence.

5.8 FIELDS AND POWER FLOW AT MEDIA INTERFACE

From the discussions of the reflection and refraction of the plane waves at a media interface, it is clear that in medium 1, we have superposition of the incident and the reflected waves, and in medium 2 we have only the transmitted wave. It is then worthwhile to investigate the distribution of the fields in the two media, and the direction of the net power flow. Let us investigate the fields in the two media for a wave with perpendicular polarization. In the two media, we have,

Medium 1 : $$\mathbf{E}_1 = \mathbf{E}_i + \mathbf{E}_r \qquad (5.128)$$

Medium 2 : $$\mathbf{E}_2 = \mathbf{E}_t \qquad (5.129)$$

Substituting for $\mathbf{E}_i$, $\mathbf{E}_r$ and $\mathbf{E}_t$ from Eqns (5.43) to (5.45), we get

$$\mathbf{E}_1 = \mathbf{E}_{i0} e^{-j\beta_1(x\sin\theta_i + z\cos\theta_i)} + \mathbf{E}_{r0} e^{-j\beta_1(x\sin\theta_i - z\cos\theta_i)} \qquad (5.130)$$

$$\mathbf{E}_2 = \mathbf{E}_{t0} e^{-j\beta_2(x\sin\theta_t + z\cos\theta_t)} \qquad (5.131)$$

For perpendicular polarization $\mathbf{E}_{i0}$, $\mathbf{E}_{r0}$ and $\mathbf{E}_{t0}$ are in the same direction. Substituting for $\mathbf{E}_{r0}$ and $\mathbf{E}_{t0}$ from Eqns (5.53) and (5.54), we can get the fields in the two media as

Medium 1

$$\mathbf{E}_1 = \mathbf{E}_{i0} e^{-j\beta_1 x\sin\theta_i}\left[e^{-j\beta_1 z\cos\theta_i} + \Gamma_{\perp} e^{+j\beta_1 z\cos\theta_i}\right] \qquad (5.132)$$

$$= \mathbf{E}_{i0} e^{-j\beta_1 x\sin\theta_i}\left[(1+\Gamma_{\perp})\cos(\beta_1 z\cos\theta_i) + j(\Gamma_{\perp} - 1)\sin(\beta_1 z\cos\theta_i)\right] \qquad (5.133)$$

The expression inside the square bracket can be written in polar form (magnitude and phase form) giving

$$\mathbf{E}_1 = |\mathbf{E}_{i0}| e^{-j\beta_1 x\sin\theta_i}(Ae^{j\Theta}) \qquad (5.134)$$

where,

$$A = \left[(1+\Gamma_{\perp})^2\cos^2(\beta_1 z\cos\theta_i) + (\Gamma_{\perp} - 1)^2\sin^2(\beta_1 z\cos\theta_i)\right]^{1/2} \qquad (5.135)$$

$$= \left([1+\Gamma_{\perp}^2] + 2\Gamma_{\perp}\cos[2\beta_1 z\cos\theta_i]\right)^{1/2} \qquad (5.136)$$

and

$$\Theta \equiv \tan^{-1}\left(\frac{(\Gamma_{\perp} - 1)}{(\Gamma_{\perp} + 1)}\tan(\beta_1 z\cos\theta_i)\right) \qquad (5.137)$$

The amplitude of the total field in medium 1 is

$$|\mathbf{E}_1| = |\mathbf{E}_{i0}|\left([1+\Gamma_{\perp}^2] + 2\Gamma_{\perp}\cos[2\beta_1 z\cos\theta_i]\right)^{1/2} \qquad (5.138)$$

and the phase of the total field, including the time phase, is

$$\phi_1 = \omega t - \beta_1 x\sin\theta_i + \tan^{-1}\left(\frac{(\Gamma_{\perp} - 1)}{(\Gamma_{\perp} + 1)}\tan(\beta_1 z\cos\theta_i)\right) \qquad (5.139)$$

From Eqn (5.138), we can note that the field amplitude $|\mathbf{E}_1|$ is a function of z now. Since, $\cos(2\beta_1 z\cos\theta_i)$ varies between -1 and +1, the field amplitude varies from $|(1-\Gamma_{\perp})||\mathbf{E}_{i0}|$ to $|(1+\Gamma_{\perp})||\mathbf{E}_{i0}|$. This is a partial standing wave created by superposition of the incident and the reflected waves. A plot of the field amplitude appears like a corrugated surface with corrugations running parallel to the interface as shown in Fig. 5.22.

The field amplitude is constant in a plane parallel to the interface ($z =$ *constant*), but it varies from plane to plane. For example, plane P_1 corresponds to the minimum field whereas P_2 corresponds to the maximum field in Fig. 5.22. We can then call these planes as 'constant amplitude planes'.

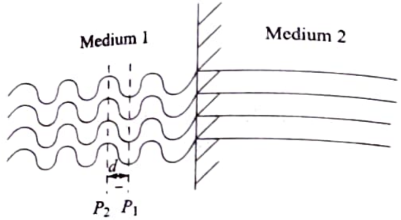


Fig. 5.22 *Electric field amplitude as function of space. Constant amplitu[de] planes.*

The distance between the two adjacent constant amplitude planes, one passi[ng] through the maximum amplitude points and other passing through the minim[um] amplitude points is given by

$$d = \frac{\pi}{2\beta_1 \cos\theta_i} = \frac{\pi}{2(2\pi/\lambda_1)\cos\theta_i} = \frac{\lambda_1}{4\cos\theta_i} \quad (5.14$$

where λ_1 is the wavelength in medium 1. Note that when $\theta_i = 0$, $d = \lambda_1/4$ whereas when $\theta_i \to \pi/2$, $d \to \infty$.

Like the constant amplitude planes, the constant phase surfaces are not plan[es]. Depending upon the angle of incidence, reflection coefficient etc, we may g[et] different constant phase surfaces.

The phase constant of the composite wave in x-direction is $\beta_1 \sin\theta_i$ and hen[ce] the phase velocity in x-direction is

$$v_{px} = \frac{\omega}{\beta_1 \sin\theta_i} \quad (5.14$$

In the z-direction, however, the phase constant cannot be written as explicit[ly] as that in the x-direction. We, therefore, have to obtain the phase velocity i[n] z-direction from the first principles. Making the phase stationary along z-directio[n] (and also making x constant) we get

$$\frac{\partial\phi_1}{\partial t} = 0 = \omega + \frac{\left(\frac{\Gamma_\perp - 1}{\Gamma_\perp + 1}\right)\sec^2(\beta_1 z\cos\theta_i)\beta_1\cos\theta_i \frac{dz}{dt}}{1 + \left\{\frac{\Gamma_\perp - 1}{\Gamma_\perp + 1}\tan(\beta_1 z\cos\theta_i)\right\}^2} \quad (5.14$$

$$\Rightarrow \quad v_{pz} = \frac{dz}{dt} = -\frac{\omega\left[1 + \left\{\frac{\Gamma_\perp - 1}{\Gamma_\perp + 1}\tan(\beta_1 z\cos\theta_i)\right\}^2\right]}{\left(\frac{\Gamma_\perp - 1}{\Gamma_\perp + 1}\right)\sec^2(\beta_1 z\cos\theta_i)\beta_1\cos\theta_i} \quad (5.14$$

From Eqns (5.139) and (5.143) two things are worth making note of:

1. The constant phase surfaces are no more planes.
2. The phase velocity in z-direction is not constant but is a function of z. Consequently, the total phase velocity becomes a function of space.

Figure 5.23 shows the phase as a function of space and Fig. 5.24 shows the variation of the phase velocity as function of space.

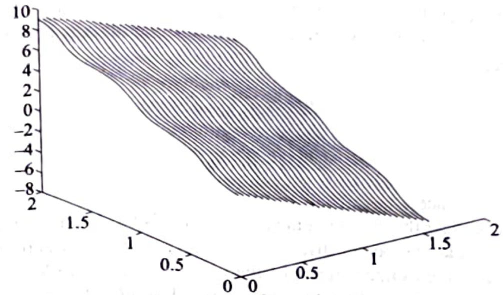


Fig. 5.23 *Spatial variation of phase for the waves at a media interface.*

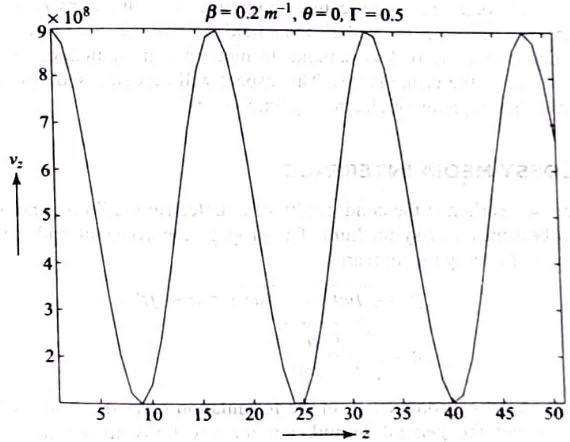


Fig. 5.24 *Variation of phase velocity in z-direction, v_z as a function of z.*

Medium 2 In medium 2 the fields are rather simple as there is only one travelling wave. The amplitude of the field is

$$|\mathbf{E}_2| = \tau_\perp |\mathbf{E}_{i0}| \quad (5.144)$$

and the phase of the wave (including the time phase) is

$$\phi_2 = \omega t - \beta_2 x \sin\theta_t + \beta_2 z \cos\theta_t \quad (5.145)$$

The amplitude of the field is constant every where in medium 2 (see Fig. 5.22) and the phase varies linearly with x and z. Consequently, the phase velocities in x and z directions

$$v_{px} = \frac{\omega}{\beta_2 \sin\theta_t} \tag{5.146}$$

$$v_{pz} = \frac{\omega}{\beta_2 \cos\theta_t} \tag{5.147}$$

are constant in space.

It should be clear that although the above analysis has been carried out for the perpendicular polarization, the discussion is valid for any polarization. The things to note are that due to interference of the incident and the reflected waves, we do not have constant field strength in medium 1. For a field receiver, like a radio receiver or a mobile phone, moving in medium 1, the receiver output therefore varies since it encounters stronger field at some place and weaker field at other. This phenomenon is called 'signal fading'. For a single media interface, as taken here, the field pattern is quite regular and predictable. In an environment where there are many reflecting boundaries, the field patterns are quite complex and the fading becomes more or less random. In modern communication systems, the fading is one of the prime issues. This aspect will be discussed in detail in the chapter on 'propagation of electromagnetic waves'.

5.9 LOSSY MEDIA INTERFACE

As discussed earlier, if the conductivity of a dielectric medium is non-zero, the medium becomes a lossy medium. The propagation constant and the intrinsic impedance of a lossy medium are

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} = \alpha + j\beta \tag{5.148}$$

$$\eta = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}} \tag{5.149}$$

If we replace $j\beta$ by complex γ, in the formulation carried out in the previous section, we get the general formulation for any lossy media interface. The formulation is valid for any arbitrary but finite value of conductivity. For $\sigma = \infty$ (ideal conductor), however, we have to reformulate the problem. This is due to the fact that for an ideal conductor, one has to include surface current in the boundary conditions. The continuity of the tangential component of the magnetic field is not valid for an ideal conductor as we have taken in our boundary conditions. Therefore, the case of ideal conducting boundary will be analyzed separately.

Although the extension of results from a loss-less media to lossy media appears straight forward, one may encounter certain conceptual difficulties regarding origin of the incident wave in medium 1. In the loss-less case, we never asked where the incident wave was originated and how much distance it had travelled before reaching the interface. This question was rather unimportant because, the amplitude of the wave was constant, and no matter how much the wave travelled (even infinite), it had the same amplitude. This, however, is not true for a lossy medium. If we assume that the incident wave was originated at an infinite distance from the interface, it would attenuate to zero for any finite value of α (that is any finite value of σ). On the other hand if we assume that an incident wave of finite amplitude is present at the interface, and if it was originated at an infinite distance from the interface, its energy would be infinite at its origin. Well, we need not worry about this as in practice neither the waves would have travelled infinite distance, nor will they be ideal plane waves. In our analysis, therefore, we start with an incident wave whose amplitude is known at the interface and find the fields in the two media in the vicinity of the interface with proper incorporation of the attenuation constant α.

EXAMPLE 5.10 A uniform plane is normally incident on an infinitely thick dielectric slab, having dielectric constant 10 and loss tangent 10^{-2} at $\omega = 10^{10}$ rad/sec. If the power density of the incident wave is 100 W/m^2. Find the power density of the wave in the dielectric at a distance of 10 m from the surface.

Solution:

$$\text{Loss tangent} = \frac{\sigma}{\omega\epsilon_0\epsilon_r} = 10^{-2}$$

$$\Rightarrow \qquad \sigma = 10^{-2}\omega\epsilon_0\epsilon_r = 0.01 \times 10^{10} \times \frac{10^{-9}}{36\pi} \times 10 = \frac{1}{36\pi}$$

For the incident wave the power density is

$$P = \frac{|E_i|^2}{\eta_0} = 100\ \text{W/m}^2$$

(All the quantities are assumed to be RMS here.)

$$\Rightarrow \qquad |E_i| = \sqrt{100\eta_0} = 194.16\ \text{V/m}$$

The transmitted field at the surface is

$$|E_t| = \tau|E_i| = \frac{2\eta_2}{\eta_1 + \eta_2}|E_i|$$

Here $$\eta_1 = \eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 120\pi$$

and $$\eta_2 = \sqrt{\frac{j\omega\mu_0}{\sigma + j\omega\epsilon_0\epsilon_r}} = 119.2 + j0.6$$

Therefore, $|E_t|_{\text{surface}} = 93.3$ V/m

Since, the dielectric slab is lossy, the field amplitude reduces exponentially ($e^{-\alpha z}$) as a function of distance where,

$$\alpha = Re\left\{\sqrt{j\omega\mu_0(\sigma + j\omega\epsilon_0\epsilon_r)}\right\} = 0.527 \text{ nepers/m}$$

The magnitude at 10 m distance from the surface is

$$|E_t| = |E_t|_{surface}e^{-0.527(10)} = 0.48 \text{ V/m}$$

The power density of the wave is

$$P = \frac{|E_t|^2}{Re\{\eta_2^*\}} = \frac{(0.48)^2}{119.2} = 1.93 \text{ mW/m}^2$$

EXAMPLE 5.11 A plane wave is normally incident on an infinitely thick dielectric slab of refractive index 2. Find the phase velocities of the total waves in air and in the dielectric slab. If the velocity varies with distance, find the upper and lower bounds on the phase velocity.

Solution:

We can note that in air we have incident as well as the reflected waves whereas in the dielectric we have only transmitted wave. In air, therefore, we get standing wave and in the dielectric we have only a traveling wave.

Let the phase constant of a wave in air be β_0. Then the phase constant in the dielectric is $\beta = \beta_0 n = 2\beta_0$. In the dielectric the phase velocity

$$v_p = \frac{\omega}{\beta} = \frac{\omega}{2\beta_0} = v_p = c/2$$

(Note: $\omega/\beta_0 = c$, the velocity of light in air.) In the air, we have to find the phase velocity from the first principles. Writing the total electric field for the standing wave in the air we have

$$E = \left(E_i e^{-j\beta_0 z} + \Gamma E_i e^{+j\beta_0 z}\right) e^{j\omega t}$$

Here, the time variation is written explicitly and Γ is the reflection coefficient.

$$\Gamma = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} = \frac{\eta_0/n - \eta_0}{\eta_0/n - \eta_0} = \frac{1-n}{1+n} = \frac{1-2}{1+2} = -\frac{1}{3}$$

Substituting for Γ we get,

$$E = E_i\left\{e^{-j\beta_0 z} - \frac{1}{3}e^{+j\beta_0 z}\right\} e^{j\omega t}$$

$$= \frac{E_i}{3}\{3[\cos\beta_0 z - j\sin\beta_0 z] - [\cos\beta_0 z + j\sin\beta_0 z]\} e^{j\omega t}$$

$$= \frac{E_i}{3}\{2\cos\beta_0 z - j4\sin\beta_0 z\} e^{j\omega t}$$

The phase of the standing wave, therefore, is

$$\phi = \omega t - \tan^{-1}\left\{\frac{4\sin\beta_0 z}{2\cos\beta_0 z}\right\}$$

$$= \omega t - \tan^{-1}(2\tan\beta_0 z)$$

The phase velocity can be obtained by making the phase stationary as a function of time, i.e.

$$\frac{d\phi}{dt} = 0 = \omega - \frac{d}{dt}\{\tan^{-1}(2\tan\beta_0 z)\}$$

$$= \omega - \frac{2\beta_0 \sec^2\beta_0 z(dz/dt)}{1 + 4\tan^2\beta_0 z}$$

$$\Rightarrow \quad \text{phase velocity} \quad v_p = \frac{dz}{dt} = \frac{\omega}{\beta_0}\frac{1 + 4\tan^2\beta_0 z}{2\sec^2\beta_0 z}$$

$$= c\left\{\frac{1}{2}\cos^2\beta_0 z + 2\sin^2\beta_0 z\right\}$$

The phase velocity in air, therefore, varies as a function z. The maximum and minimum values of phase velocity are $2c$ and $c/2$.

5.10 REFLECTION FROM A CONDUCTING BOUNDARY

Let us investigate the reflection of a plane wave from an ideal conducting boundary ($\sigma = \infty$). In this case, we have a dielectric-conductor interface and the plane wave is incident from the dielectric side. Since, no fields can exist inside an ideal conductor, there is no question of transmission (i.e. refraction) of the wave to medium 2. We, therefore, have only reflection at the dielectric-conductor interface. As in previous case, here also, we analyse reflection of two polarizations, namely parallel and perpendicular.

5.10.1 Perpendicular Polarization

Let us consider a plane wave with perpendicular polarization as shown in Fig. 5.25.

Let the angle of incidence be θ_i, and let the **E**-fields for both incident and reflected waves be oriented along $+y$-direction. The directions of the **H**-fields can be obtained using the Poynting vector, i.e. **E** × **H** should give the direction of the wave propagation. The fields for the incident wave hence can be written as

$$\mathbf{E}_i = E_{i0}e^{-j\beta_1(x\sin\theta_i + z\cos\theta_i)}\hat{\mathbf{y}} \quad (5.150)$$

$$\mathbf{H}_i = \frac{E_{i0}}{\eta_1}e^{-j\beta_1(x\sin\theta_i + z\cos\theta_i)}(-\cos\theta_i\hat{\mathbf{x}} + \sin\theta_i\hat{\mathbf{z}}) \quad (5.151)$$

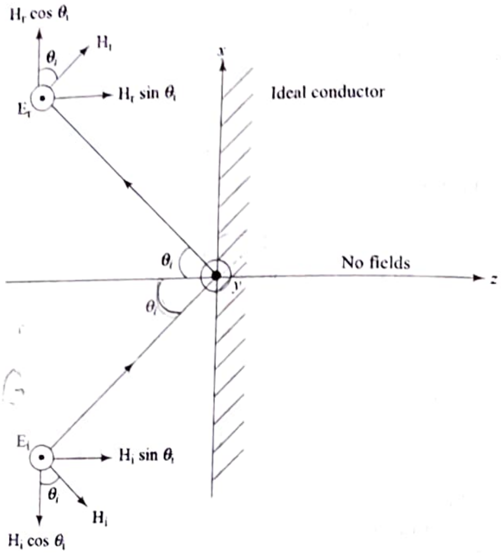


Fig. 5.25 *Reflection from a conducting boundary: Perpendicular polarization.*

and the reflected fields can be written as

$$\mathbf{E_r} = E_{r0}e^{-j\beta_1(x\sin\theta_i - z\cos\theta_i)}\hat{\mathbf{y}} \tag{5.152}$$

$$\mathbf{H_r} = \frac{E_{r0}}{\eta_1}e^{-j\beta_1(x\sin\theta_i - z\cos\theta_i)}(\cos\theta_i\hat{\mathbf{x}} + \sin\theta_i\hat{\mathbf{z}}) \tag{5.153}$$

Applying boundary conditions at the interface, i.e. continuity of the tangential components of the electric field and the normal components of the magnetic field, and noting that the fields are zero inside an ideal conductor, we get at $z = 0$,

$$\mathbf{E}_{1tan} = (\mathbf{E_i} + \mathbf{E_r})_{tan} = 0 \tag{5.154}$$

$$\mathbf{H}_{1nor} = (\mathbf{H_i} + \mathbf{H_r})_{nor} = 0 \tag{5.155}$$

Noting that the **E** fields are y-oriented and hence are tangential to the interface, we get from Eqn (5.154),

$$E_{i0}e^{-j\beta_1 x\sin\theta_i} + E_{r0}e^{-j\beta_1 x\sin\theta_i} = 0 \tag{5.156}$$

For normal component (z-component) of the magnetic field we have,

$$\frac{E_{i0}}{\eta_1}\sin\theta_i e^{-j\beta_1 x\sin\theta_i} + \frac{E_{r0}}{\eta_1}\sin\theta_i e^{-j\beta_1 x\sin\theta_i} = 0 \tag{5.157}$$

Equations (5.156) and (5.157) both give the same condition, i.e.

$$E_{r0} = -E_{i0} \tag{5.158}$$

From Eqn (5.158) it is clear that the reflected electric field is equal in magnitude to the incident field but 180° out of phase. In other words, the reflection coefficient

$$\Gamma_\perp \equiv \frac{E_{r0}}{E_{i0}} = -1 \tag{5.159}$$

We may recall that on a transmission line, the voltage reflection coefficient becomes −1 for the short circuit load. The ideal conducting boundary therefore is analogous to the short circuit on the transmission line.

Substituting for E_{r0} from Eqn (5.158), the total field in medium 1 is

$$\mathbf{E_1} = \mathbf{E_i} + \mathbf{E_r}$$

$$= E_{i0}e^{-j\beta_1 x\sin\theta_i}\left(e^{-j\beta_1 z\cos\theta_i} - e^{j\beta_1 z\cos\theta_i}\right)\hat{\mathbf{y}} \tag{5.160}$$

$$= -2jE_{i0}\underbrace{\sin(\beta_1 z\cos\theta_i)}_{\text{standing wave in x dr}^n}\overbrace{e^{-j\beta_1 x\sin\theta_i}\hat{\mathbf{y}}}^{\text{travelling wave in z direc}^n} \tag{5.161}$$

Similarly, H-field can be written as:

$$\mathbf{H_1} = \mathbf{H_i} + \mathbf{H_r}$$

$$= \frac{E_{i0}}{\eta_1}e^{-j\beta_1 x\sin\theta_i}[(-\cos\theta_i\hat{\mathbf{x}} + \sin\theta_i\hat{\mathbf{z}})e^{-j\beta_1 z\cos\theta_i}$$
$$-(\cos\theta_i\hat{\mathbf{x}} + \sin\theta_i\hat{\mathbf{z}})e^{j\beta_1 z\cos\theta_i}] \tag{5.162}$$

$$= -\frac{E_{i0}}{\eta_1}\cos\theta_i[e^{-j\beta_1 z\cos\theta_i} + e^{j\beta_1 z\cos\theta_i}]e^{-j\beta_1 x\sin\theta_i}\hat{\mathbf{x}}$$
$$-\frac{E_{i0}}{\eta_1}\sin\theta_i[e^{j\beta_1 z\cos\theta_i} - e^{-j\beta_1 z\cos\theta_i}]e^{-j\beta_1 x\sin\theta_i}\hat{\mathbf{z}} \tag{5.163}$$

$$= -2\frac{E_{i0}}{\eta_1}\cos\theta_i\cos(\beta_1 z\cos\theta_i)e^{-j\beta_1 x\sin\theta_i}\hat{\mathbf{x}}$$
$$-2j\frac{E_{i0}}{\eta_1}\sin\theta_i\sin(\beta_1 z\cos\theta_i)e^{-j\beta_1 x\sin\theta_i}\hat{\mathbf{z}} \tag{5.164}$$

Equations (5.161) and (5.164) represent a travelling wave in +x-direction and a standing wave iz z-direction. Figure 5.26 shows variation of $|E_y|$, $|H_x|$ and $|H_z|$ components as a function of distance in z-direction.

From Fig. 5.26 following points can be noted:

1. $|E_y|$ and $|H_z|$ standing wave patterns are aligned in space whereas, the $|H_x|$ standing wave pattern is shifted by quardrature with respect to the $|H_z|$ (or $|E_y|$) pattern. That is, where $|H_z|$ is maximum, $|H_x|$ is zero and vice versa.

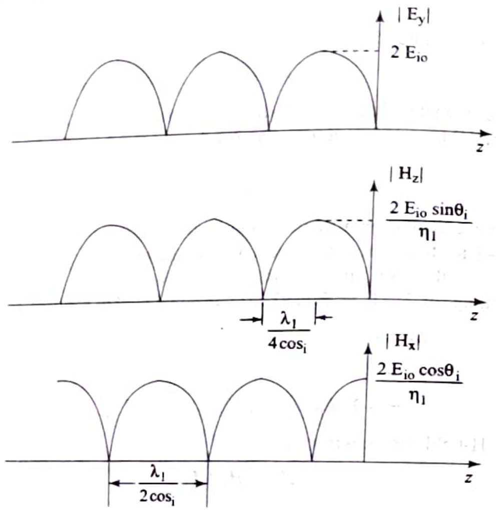


Fig. 5.26 *Variation of amplitudes of the field components as function of distance from the conducting boundary.*

2. $|E_y|$ and $|H_z|$ are zero at the interface as required by the boundary conditions.
3. The electric field E_y also becomes zero, when

$$\beta_1 z\cos\theta_i = \pm m\pi \qquad m = 0, 1, 2, \cdots \tag{5.165}$$

$$\Rightarrow \quad z = -\frac{m\pi}{\beta_1\cos\theta_i} = -\frac{m\lambda_1}{2\cos\theta_i} \tag{5.166}$$

Here, we have used $\beta_1 = 2\pi/\lambda_1$ and only negative sign has been chosen since z is negative in medium 1.

Equation (5.166) is the equation of a plane parallel to the xy-plane, i.e. parallel to the interface. We, therefore, see that the electric field is zero in all planes which are parallel to the interface and are at distances $\lambda_1/2\cos\theta_i$, $2\lambda_1/2\cos\theta_i$, $3\lambda_1/2\cos\theta_i$, and so on. Also, from Eqn (5.164) the z-component of the magnetic field, i.e. the component of the magnetic field which is normal to the planes, is also zero in these planes. In other words, the magnetic field becomes tangential on these planes.

4. The magnetic field component H_x is zero when

$$\beta_1 z\cos\theta_i = \pm(2m+1)\frac{\pi}{2} \qquad m = 0, 1, 2, \cdots \tag{5.167}$$

$$\Rightarrow \quad z = -(2m+1)\frac{\pi}{2\beta_1\cos\theta_i} = -(2m+1)\frac{\lambda_1}{4\cos\theta_i} \tag{5.168}$$

5. The tangential component of the magnetic field, H_x is maximum at the interface. Now, since there is no magnetic field inside the conductor, there has to be a surface current at the interface

$$\mathbf{J_s} = \hat{\mathbf{n}} \times \mathbf{H} \tag{5.169}$$

where $\hat{\mathbf{n}}$ unit outward normal to the interface. For our coordinate system $\hat{\mathbf{n}} = \hat{\mathbf{z}}$. We, therefore, have

$$\mathbf{J_s} = \hat{\mathbf{z}} \times \mathbf{H} = H_x\hat{\mathbf{y}} \tag{5.170}$$

The surface current has a magnitude of $(2E_{i0}\cos\theta_i)/\eta_1$ and it flows along $+y$ direction.

6. The net power flow is in x-direction, that is along the interface. This can be seen in two ways:
   (a) The fields represent a travelling wave along $+x$-direction with a phase constant $\beta_{1x} = \beta_1\sin\theta_i$.
   (b) If we calculate the average Poynting vector $\frac{Re(\mathbf{E}\times\mathbf{H}^*)}{2}$, we find that the fields E_y and H_x are 90° out of phase, giving average power flow zero along z-direction. The fields E_y and H_z are however in phase, and therefore constitute an average power flow along $+x$-direction ($\hat{\mathbf{y}} \times \hat{\mathbf{z}} = \hat{\mathbf{x}}$). The average power flow density is

$$P_x = 2E_{i0}^2\sin\theta_i\sin^2(\beta_1 z\cos\theta_i) \tag{5.171}$$

7. The wavelength along the interface is

$$\lambda_g = \frac{2\pi}{\beta_1\sin\theta_i} = \frac{\lambda_1}{\sin\theta_i} \tag{5.172}$$

8. This case is analogous to a short-circuited transmission line. If we take E_y for voltage V on the transmission line, the variation of E_y is same as the voltage standing wave pattern on the line.
9. For $\theta_i = \pi/2$, i.e. for grazing incidence, $\mathbf{E_1}$ becomes zero. We can therefore conclude that no wave with $\perp$ polarization can be launched at a conducting interface at the grazing angle.

5.10.2 Parallel Polarization

Let us now carry out an analysis for a wave which has its electric field in the plane of incidence as shown in Fig. 5.27. Let the angle of incidence be θ_i.

Since, $\mathbf{E_i}$ and the incident wave vector lie in the plane of the paper, the magnetic field vector has to be perpendicular to the plane of the paper. Without losing

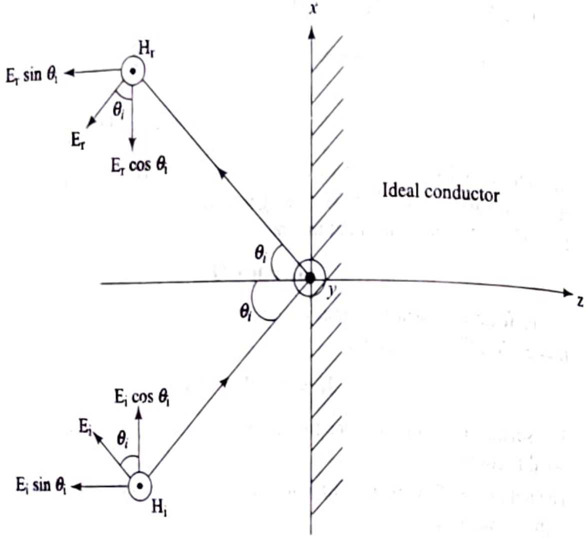


Fig. 5.27 *Reflection from conducting boundary: parallel polarization.*

generality, let us assume that the magnetic fields for both, incident and reflected waves point out of the plane of the paper, i.e. along $+y$-axis. The direction of the electric fields can be appropriately chosen to give correct directions of the Poynting vectors for the two waves as shown in Fig. 5.27.

The vector fields for the incident and the reflected waves can be written as

(a) Incident Wave:

$$\mathbf{H_i} = \frac{E_{i0}}{\eta_1} e^{-j\beta_1(x\sin\theta_i + z\cos\theta_i)}\hat{\mathbf{y}} \qquad (5.173)$$

$$\mathbf{E_i} = E_{i0} e^{-j\beta_1(x\sin\theta_i + z\cos\theta_i)}\{\cos\theta_i\hat{\mathbf{x}} - \sin\theta_i\hat{\mathbf{z}}\} \qquad (5.174)$$

(b) Reflected Wave:

$$\mathbf{H_r} = \frac{E_{r0}}{\eta_1} e^{-j\beta_1(x\sin\theta_i - z\cos\theta_i)}\hat{\mathbf{y}} \qquad (5.175)$$

$$\mathbf{E_i} = E_{r0} e^{-j\beta_1(x\sin\theta_i - z\cos\theta_i)}\{-\cos\theta_i\hat{\mathbf{x}} - \sin\theta_i\hat{\mathbf{z}}\} \qquad (5.176)$$

Continuity of the tangential components of the electric field at the interface ($z = 0$) yields

$$E_{i0}\cos\theta_i e^{-j\beta_1 x\sin\theta_i} - E_{r0}\cos\theta_i e^{-j\beta_1 x\sin\theta_i} = 0 \qquad (5.177)$$

$$\Rightarrow \quad E_{i0} = E_{r0} \qquad (5.178)$$

That is, the reflection coefficient for the field direction assumed in Fig. 5.27 is

$$\Gamma_{\parallel} = \frac{E_{r0}}{E_{i0}} = 1 \qquad (5.179)$$

Substituting for E_{r0} from Eqn (5.178) and carrying out some algebric manipulations we get

$$\mathbf{E_1} = \mathbf{E_i} + \mathbf{E_r}$$
$$= 2E_{i0}e^{-j\beta_1 x\sin\theta_i}\{j\cos\theta_i\sin(\beta_1 z\cos\theta_i)\hat{\mathbf{x}} - \sin\theta_i\cos(\beta_1 z\cos\theta_i)\hat{\mathbf{z}}\} \qquad (5.180)$$

and

$$\mathbf{H_1} = \mathbf{H_i} + \mathbf{H_r} = 2\frac{E_{i0}}{\eta_1}\cos(\beta_1 z\cos\theta_i)e^{-j\beta_1 x\sin\theta_i}\hat{\mathbf{y}} \qquad (5.181)$$

The interpretations of the above equations are similar to that for the perpendicular polarization except that the field components are different in two cases. In this case the electric field has two components E_x and E_z and the magnetic field has only H_y component. The fields have standing wave nature in z-direction and a travelling wave nature in the x-direction. The standing wave patterns for the three components are shown in Fig. 5.28.

Following observations can be made from Eqns (5.180) and (5.181):

1. $|E_z|$ and $|H_y|$ patterns are aligned in space and $|E_x|$ pattern is in quardrature with respect to them.
2. At the interface the magnetic field is enhanced by factor of 2 i.e. the magnetic field at the interface is double of that of the incident field.
3. The surface current at the interface is $|J_s| = 2E_{i0}/\eta_1$ and is independent of the angle of incidence.
4. The wave travels along x-direction with a phase constant of

$$\beta_{1x} = \beta_1\sin\theta_i$$

giving guided wavelength along the interface

$$\lambda_g = \frac{2\pi}{\beta_{1x}} = \frac{\lambda_1}{\sin\theta_i} \qquad (5.182)$$

5. The average power flow is along x-direction and the average power flow density is

$$P_x = \frac{1}{2}Re\{-E_z H_y\} = 2E_{i0}^2\sin\theta_i\cos^2(\beta_1 z\cos\theta_i) \qquad (5.183)$$

The power flow ceases for $\theta_i = 0$, i.e. for the normal incidence.

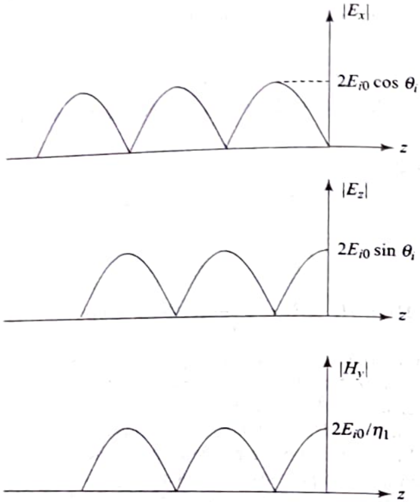


Fig. 5.28 *Amplitude variation of field components as function of distance from the conducting boundary.*

6. The x-component of the electric field goes to zero in planes parallel to the interface and at locations given by

$$z = \frac{-m\pi}{\beta_1 \cos\theta_i} = -\frac{m\pi}{2\cos\theta_i} \tag{5.184}$$

Again, only -ve sign is taken as z is negative in medium 1.

7. For $\theta_i = \pi/2$, i.e. at the grazing incidence,

$$\mathbf{E_1} = -2E_{i0}e^{-j\beta_1 x}\hat{\mathbf{z}} \tag{5.185}$$

$$\mathbf{H_1} = 2\frac{E_{i0}}{\eta_1}e^{-j\beta_1 x}\hat{\mathbf{y}} \tag{5.186}$$

Note that unlike $\perp$ polarization, the reflection at grazing angle is possible for the parallel polarization. This is due to the fact that for the grazing incidence of a parallely polarized wave, the electric field is perpendicular to the conducting boundary, and the magnetic field is tangential to the conducting boundary. In other words, in this case the tangential component of the electric field and normal component of the magnetic field are intrinsically zero satisfying the natural boundary conditions at the interface. The conducting boundary consequently does not have any effect on the wave propagation and the wave keeps propagating as it

would have propagated in the absence of the boundary. There is a little difference, however, and that is, in the absence of the boundary the wave fronts will be over the infinite space ($-\infty \le z \le +\infty$), whereas, in the presence of the boundary the phase fronts is only over the semi-infinite space ($-\infty \le z \le 0$).

5.10.3 Normal Incidence

For normal incidence, $\theta_i = 0$ and $\mathbf{E_i}$ and $\mathbf{H_i}$ become parallel to the conducting boundary. Taking $\mathbf{E_i}$ and $\mathbf{E_r}$ oriented along the x-direction, and $\mathbf{H_i}$ and $\mathbf{H_r}$ along y-direction the fields can be written as

(a) Incident Wave:

$$\mathbf{E_i} = E_{i0}e^{-j\beta_1 z}\hat{\mathbf{x}} \tag{5.187}$$

$$\mathbf{H_i} = \frac{E_{i0}}{\eta_1}e^{-j\beta_1 z}\hat{\mathbf{y}} \tag{5.188}$$

(b) Reflected Wave:

$$\mathbf{E_r} = E_{r0}e^{j\beta_1 z}\hat{\mathbf{x}} \tag{5.189}$$

$$\mathbf{H_r} = \frac{E_{r0}}{\eta_1}e^{j\beta_1 z}\hat{\mathbf{y}} \tag{5.190}$$

At the interface (i.e. $z = 0$), the electric field, which is inherently tangential to the boundary in this case, has to be zero, giving

$$\mathbf{E_1} = (E_{i0} + E_{r0})\hat{\mathbf{x}} = 0 \tag{5.191}$$

$$\Rightarrow \quad E_{i0} = -E_{r0} \tag{5.192}$$

The reflection coefficient therefore is

$$\Gamma \equiv \frac{E_{r0}}{E_{i0}} = -1 \tag{5.193}$$

It can be observed that this case is identical to a transmission line where the wave propagation is in one dimension (along z-direction), and the line is terminated in a short circuit load. Since the magnitude of the reflection coefficient is unity, there is no net average power flow in any direction.

The above analysis clearly indicates that irrespective of the polarization of the incident wave, the electric field component parallel to the boundary goes to zero at a distance $m\lambda_1/2\cos\theta_i$ from the boundary. This means, if we place an infinite conducting plane parallel to the interface at this distance, the fields would inherently satisfy the boundary conditions and hence would not be disturbed. The presence of another conducting plane however, creates a bound structure in the z-direction. In other words, the electromagnetic fields are confined between two parallel conducting planes and there is a net flow of power along the planes (along x-direction).

We, therefore, find that a parallel conducting plane geometry is capable of guiding electromagnetic waves. This structure is hence called the 'parallel plane waveguide'. The parallel plane waveguide, although not a very practical structure, forms the foundation of a more practical structure, the 'rectangular waveguide'. It is interesting to note that even though no energy penetrates a conducting boundary, the boundary controls the flow of power. Of course, there are surface charges and surface currents which are responsible for the guiding of the wave.

5.11 SUMMARY

In this chapter, we have first developed a frame work for the wave propagation in an arbitrary direction with respect to the coordinate axes. The reflection of plane waves from media interfaces was investigated further. The laws of reflection and refraction have been established from the phase matching of the waves at the interface. Normal as well as oblique incidences for both the polarizations have been investigated and the expressions for the reflection and transmission coefficients have been derived. The phenomenon of total internal reflection have been analysed in detail. Towards the end of the chapter, reflection from the conducting boundaries is studied and it is shown that the conducting boundaries guide the electromagnetic energy along them.

This sets the foundation for the wave guiding structures called the waveguides. In the next chapter we will analyse different metallic waveguides and later the dielectric waveguides.

Review Questions

5.1 What is a media interface?

5.2 What is the equation of a phase front of a uniform plane wave travelling in an arbitrary direction?

5.3 What is phase velocity and what is its range?

5.4 What is group velocity?

5.5 In what situation are the phase and group velocities equal?

5.6 What is a plane of incidence?

5.7 Give arguments to show that when a uniform plane wave is incident on a plane media interface, the reflected and refracted waves are also uniform plane waves.

5.8 What are the laws of reflection and refraction?

5.9 Define angles of incidence, reflection and refraction.

5.10 What does perpendicular and parallel polarization mean?

5.11 What are reflection and transmission coefficients?

5.12 If the magnitude of reflection coefficient is one, should the transmission coefficient be zero? Why?

5.13 Can the transmitted electric field be greater than the incident electric field?

5.14 Can the reflected electric field be greater than the incident electric field?

5.15 Under what condition is the reflection coefficient real?

5.16 If the transmitted electric field is greater than the incident electric field should the transmitted magnetic field be necessarily smaller than the incident magnetic field?

5.17 On a media interface can one get reflectionless transmission?

5.18 What is Brewster angle?

5.19 Why is the Brewster angle also called the polarizing angle?

5.20 An infinitely thick dielectric sheet is coated with a dielectric layer. What is the dielectric constant of the coated layer so as to not get any reflection for a normally incident wave?

5.21 Why does a linearly polarized wave remain linearly polarized after reflection from a dielectric interface but the circularly polarised wave become elliptically polarized?

5.22 What is total internal reflection?

5.23 What is the difference between TIR and ordinary reflection?

5.24 Will a linearly polarized wave remain linearly polarized after TIR? Explain.

5.25 What is the propagation constant of the transmitted fields at TIR?

5.26 How does the transmitted fields vary as a function of distance away from the interface?

5.27 Draw the amplitude variation of net travelling wave at TIR.

5.28 For a dielectric interface does the Brewster angle always exist?

5.29 At TIR what is the direction of net power flow?

5.30 At a conducting interface why is the reflection coefficient −1?

5.31 A uniform plane wave is obliquely incident on an ideal conducting interface. What is the direction of the surface current at the interface? See for both perpendicular and parallel polarizations.

Problems

5.1 The electric field of a uniform plane wave in vacuum is given as

$$\mathbf{E} = (2\hat{x} + 4\hat{y} - 3\hat{z})e^{-j[\text{illegible}]}\ \text{V/m}$$

Find the vector magnetic field and the direction of the wave.

5.2 In a dielectric medium a wave has electric and magnetic fields given as

$$\mathbf{E} = (j\hat{x} + 2\hat{y} - j\hat{z})e^{-j[\text{illegible}]}\ \text{V/m}$$

$$\mathbf{H} = \frac{1}{60\pi}(-\hat{x} + j\hat{y} + \hat{z})e^{-j[\text{illegible}]}\ \text{V/m}$$

Show that the wave is a uniform plane wave? Find (i) phase constant of the wave (ii) velocity of the wave (iii) frequency of the wave.

5.3 A plane wave travels in the yz-plane at an angle of 30° from the +y direction. The electric field of the wave is oriented in the x-direction with amplitude 50 V/m. If the medium has the dielectric constant of 2.5, find the vector magnetic field, the wave vector and the phase constant of the wave. Frequency of the wave is 1 GHz.

5.4 The wave vector of a plane wave makes 45° angle with the x-axis and 120° angle with the y-axis. The wave is circularly polarized with peak electric field of 100 V/m. If the relative permeability and relative permittivity of the medium are 1.1 and

4 respectively, find the vector electric and magnetic fields. What are the phase constants and phase velocities of the wave in the x, y and z-directions? Frequency of the wave is 10 MHz.

5.5 Two uniform plane waves of equal strengths, one travelling along the $+z$-direction and other travelling in the xz-plane at an angle of 60° with respect to the $+z$-axis, co-exist in a medium. The dielectric constant of the medium is 4. Find the phase velocity of the composite wave along the z-direction. Also find the constant phase and constant electric field amplitude surfaces of the combined wave.

5.6 A 100 MHz plane wave is launched at an angle θ from the ground. The wavelength along the ground is measured to be 5 m. Find the value of θ and the phase velocity of the wave along the ground surface and in the vertical direction. What is the group velocity of the wave in the vertical direction?

5.7 An elliptically polarized TEM wave travels in vacuum in the xy-plane with direction cosines (0.8, 0.6, 0). The axial ratio of the ellipse is 2.5 and the major axis of the ellipse lies in the xy-plane. Write the expression for the electric and magnetic fields if the peak electric field is 20 V/m.

5.8 A uniform wave is incident from air on an infinitely thick medium at the angle of incidence of 35°. Find the angle of reflection and angle of transmission. The medium has $\mu_r = 49$ and $\epsilon_r = 6$. What is the phase velocity of the wave along the media interface?

5.9 A dielectric interface is along $z = 0$ plane. The medium for $z > 0$ is air and for $z < 0$ is water with refractive index 1.33. If the incident wave vector has direction cosines $(0.4, 0.5, \ell)$. Find the direction cosines of the wave vectors of the reflected and transmitted waves.

5.10 When a light beam enters a dielectric medium from air, its path is deviated by 20° and is slowed down by a factor 1.5. What is the phase velocity of the wave along the dielectric air interface?

5.11 A plane wave having peak electric field of 25 V/m is incident at a air dielectric interface with perpendicular polarization. The dielectric constant of the medium is 2.4 and the angle of incidence is 30°. Find the power density of the incident, reflected and transmitted waves.

5.12 Show that a circularly polarized wave cannot remain circularly polarized after reflecting from a dielectric interface.

5.13 A light beam is incident on a glass slab at an angle of incidence of 50°. The light is linearly polarized and the plane of polarization makes an angle of 30° with the plane of incidence. By what angle does the plane of polarization rotate when the beam gets inside the glass slab?

5.14 A plane wave is incident with parallel polarization at a dielectric interface. The two media have parameters $\mu_1 = \mu_0$, $\epsilon_{r1} = 3$, $\mu_2 = 10\mu_0$, $\epsilon_{r2} = 2$. Find the expressions for the incident, reflected and transmitted waves if the angle between the reflected and transmitted wave vectors is 90°. The peak incident electric field is 10 V/m.

5.15 A 1 GHz electromagnetic wave is normally incident on a 3 cm thick plastic slab of dielectric constant 5. What per cent of the incident power is transmitted through the slab?

CHAPTER 6

Waveguides

dielectric waveguides are used at high frequency transmission

As the name suggests, waveguide is a structure which can guide electromagnetic energy. Transportation of electromagnetic energy is one of the prime needs of a communication system. Communication engineers are in constant search of structures which can transport electromagnetic energy over long distances with minimum possible loss. Metallic waveguides, which include two conductor transmission lines like co-axial cables, can transport the electromagnetic energy efficiently over frequencies ranging between few tens of kHz to few tens of GHz. As the frequency increases further to optical spectrum, the metallic waveguides due to their excessive loss, become inefficient and the dielectric waveguides are to be used in the form of optical fibers.

The key feature of all the wave guiding structures is the 'Modal propagation'. The electromagnetic energy propagates along a waveguide in the form of some definite field patterns called 'modes'. The propagation of waves in the form of modes is a direct consequence of the bounded nature of the structure which a waveguide has. In this chapter, we first develop the concept of Modal propagation for a simple structure like a parallel plane waveguide, and subsequently analyse the rectangular waveguide.

Whenever we have a new problem to analyze in electromagnetics, the general approach is to solve the Maxwell's equations under the constraints of the given problem. Indeed, any arbitrarily complex problem can be solved with this approach. The only drawback of this approach, however, is that one quickly gets lost in the vector manipulations and therefore, does not get a good physical insight into the problem. We will certainly rely on this approach for rectangular and circular waveguides, but, for parallel plane waveguides we will approach the problem with more physical understanding rather than the routine approach. Once we get the physical understanding of the modal propagation, it will be easy to interpret the results obtained with the routine approach for more complex waveguides.

6.1 PARALLEL PLANE WAVEGUIDE

As discussed in Chapter 5, a uniform plane wave when reflected from a perfectly conducting boundary, forms a standing wave in the space. The standing wave pattern travels along the conducting boundary giving a net power flow along the boundary. Let us build up on this concept to get the parallel plane waveguide. We will show in the following sections that, in a parallel plane waveguide, the electromagnetic wave will propagate in the form of Transverse Electric or Transverse Magnetic modes that is, either the electric field or the magnetic field will be perpendicular to the direction of the net wave propagation.

6.1.1 Transverse Electric (TE) mode

Let us re-define the coordinate system for a conducting boundary as shown in Fig. 6.1. The conducting boundary is in the yz-plane and the x-axis is normal to the boundary. The reason for choosing this coordinate system is that generally, the direction of power flow in a waveguide is taken along the z-axis. Later on, it will become clear that this coordinate system gives a net power flow along the z-axis.

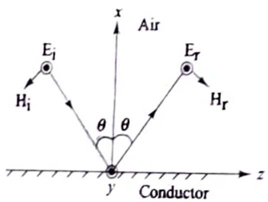


Fig. 6.1 *Reflection of a uniform plane wave with perpendicular polarization giving rise to TE modes.*

Let us consider a plane wave with the perpendicular polarization incident on the conducting boundary at an angle of incidence θ. Then from the laws of reflection, the angle of reflection is also θ. If we compare the coordinate system used here with that used in the previous chapter we note that the new coordinate system has been rotated around y-axis by 90° with respect to the old coordinate system. The old x-axis becomes new z-axis and the old z axis becomes new $-x$-axis. Therefore if we replace x by z and z by $-x$ in the expression derived in the last chapter we get the expressions for the fields in the new coordinate system. The total electric and magnetic field in medium 1 can hence be written from Eqns (5.161), (5.164) as

$$\mathbf{E}_1 = 2jE_{i0}\sin(\beta_1 x\cos\theta)e^{-j\beta_1 z\sin\theta}\hat{\mathbf{y}} \tag{6.1}$$

and

$$\mathbf{H}_1 = 2j\frac{E_{i0}}{\eta_1}\sin(\beta_1 x\cos\theta)e^{-j\beta_1 z\sin\theta}\cos\theta\hat{\mathbf{x}} + 2\frac{E_{i0}}{\eta_1}\cos(\beta_1 x\cos\theta)e^{-j\beta_1 z\sin\theta}\sin\theta\hat{\mathbf{z}} \tag{6.2}$$

The electric field which is parallel to the conducting boundary becomes zero at the planes located at

$$x = d = \frac{m\lambda_1}{2\cos\theta} \qquad m = 0, 1, 2\ldots \tag{6.3}$$

As discussed earlier, this means that the fields are unaffected if we place another conducting boundary at any x given by Eqn (6.3). It is then clear that for a given angle of incidence θ, the height of the second conducting boundary, which does not disturb the fields, is a multiple of $\lambda_1/(2\cos\theta)$.

Let us now reverse Eqn (6.3) to yield

$$\cos\theta = \frac{m\lambda_1}{2d} = \frac{m\pi}{\beta_1 d} \tag{6.4}$$

This implies that if we have two parallel conducting boundaries separated by a distance d, the fields between them will be due to superposition of waves incident at angles θ given by Eqn (6.4). Since, in Eqn (6.4) m is an integer, we obtain discrete values of θ. Note that in case of single boundary, the wave would have reflected and the fields in region 1 would have survived at any angle of incidence. In case of parallel boundaries however, there are definite angles of incidence for which the fields satisfy the boundary conditions and therefore survive. For other angles of incidence, the fields cannot survive as they do not meet the boundary conditions. The angle θ which was continuous in case of single boundary has become discrete in case of parallel conducting boundaries. So, we note that as the second boundary is introduced, some things are radically changed. These changes can be summarized as follows:

(i) We have naturally migrated from the continuous domain of θ to the discrete domain.

(ii) Since $\cos\theta$ is ≤ 1, there are finite number of angles θ for a given boundary separation 'd' and the frequency.

(iii) For $d < \lambda_1/2$, $\cos\theta$ is always greater than 1 even for the smallest value of $m = 1$. This implies that no wave can be launched between the two conducting planes, if the separation between them is less than $\frac{\lambda_1}{2}$.

(iv) As the separation between the boundaries increases or the wavelength decreases, the number of angles at which the wave can be launched, also increases.

Substituting for $\cos\theta$ from Eqn (6.4) into the field expressions we get

$$\mathbf{E}_1 = 2jE_{i0}\sin\left(\beta_1 x\frac{m\lambda_1}{2d}\right)e^{-j\beta_1 z\sqrt{1-\left(\frac{m\lambda_1}{2d}\right)^2}}\,\hat{\mathbf{y}} \tag{6.5}$$

$$\mathbf{H}_1 = 2j\frac{E_{i0}}{\eta_1}\sin\left(\beta_1 x\frac{m\lambda_1}{2d}\right)e^{-j\beta_1 z\sqrt{1-\left(\frac{m\lambda_1}{2d}\right)^2}}\cos\theta\,\hat{\mathbf{x}} + 2\frac{E_{i0}}{\eta_1}\cos\left(\beta_1 x\frac{m\lambda_1}{2d}\right)e^{-j\beta_1 z\sqrt{1-\left(\frac{m\lambda_1}{2d}\right)^2}}\sin\theta\,\hat{\mathbf{z}} \tag{6.6}$$

We can simplify the field expressions by noting that

$$\beta_1 x\frac{m\lambda_1}{2d} = \frac{2\pi}{\lambda_1}x\frac{m\lambda_1}{2d} = \frac{m\pi}{d}x \tag{6.7}$$

Also the phase constant of the wave along the z-direction can be written as

$$\beta = \beta_1\sqrt{1-\left(\frac{m\lambda_1}{2d}\right)^2} = \beta\sin\theta \tag{6.8}$$

$$\Rightarrow \qquad \sin\theta = \frac{\beta}{\beta_1} \tag{6.9}$$

(Instead of calling the phase constant along the z-direction, β_z, we have called it β since the net propagation of wave is along the z direction only and therefore the suffix z is implicit). The fields which can exist between two conducting boundaries can finally be written as

$$\mathbf{E}_1 = 2jE_{i0}\sin\left(\frac{m\pi x}{d}\right)e^{-j\beta z}\hat{\mathbf{y}} \tag{6.10}$$

$$\mathbf{H}_1 = 2j\frac{E_{i0}}{\eta_1}\sin\left(\frac{m\pi x}{d}\right)e^{-j\beta z}\left(\frac{m\pi}{\beta_1 d}\right)\hat{\mathbf{x}} + \frac{2E_{i0}}{\eta_1}\cos\left(\frac{m\pi x}{d}\right)e^{-j\beta z}\frac{\beta}{\beta_1}\hat{\mathbf{z}} \tag{6.11}$$

From Eqns (6.10) and (6.11) we note the interesting fact that, for a given value of m the variation of fields in x direction is fixed, irrespective of the frequency. For example, if we take $m = 1$, the fields have one half cycle variation in the x direction, if we take $m = 2$, the fields have one full cycle variation in the x direction and so on. In other words, for $m = 1$, we get maximum electric field half way between the two planes and the electric field is zero at the two planes as shown in Fig. 6.2. The field pattern therefore is unique for a given value of m. One would also note that there is no gradual change from one pattern to another i.e. either we get one pattern or the next or both. These unique field patterns are called the 'modal field patterns' and the propagation of an electromagnetic wave in the form of these patterns is referred to as the 'modal propagation'. It should be re-emphasized that if an electromagnetic wave has to exist between two conducting boundaries, it has to exist only in the form of either of the modes.

From Eqns (6.10) and (6.11), it is clear that for $m = 0$, the electric field vanishes and the magnetic field has only z-component. However, since no time varying magnetic field can exist without the corresponding electric field, the z-component

of the magnetic field has to be identically zero. Since $m = 0$ corresponds to $\theta = \pi/2$, it can be concluded that no wave can be launched between two conducting boundaries with its electric field parallel to them if $m = 0$. This can also be seen from the following argument.

Since $m = 0$ means no field variation in x-direction, the electric field is constant along the x-direction. However the electric field which is tangential to the boundary has to be zero at $x = 0$ and $x = d$. This is possible only if the field is identically zero everywhere between the boundaries.

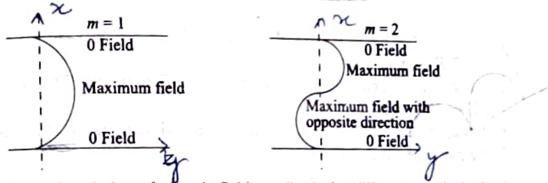


Fig. 6.2 *Spatial variation of electric field amplitude for different mode indices m.*

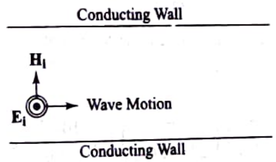


Fig. 6.3 *Electric field between two parallel conducting planes for $m = 0$.*

From Eqns (6.10) and (6.11) we note that the fields travel along $+z$-direction. The electric field which is in y-direction is transverse to the direction of wave propagation. The magnetic field has x as well as z-component, i.e. it has components parallel to as well as perpendicular to the direction of the wave propagation. The electric field therefore has a special nature, that it is transverse to the direction of the wave propagation and consequently, the mode is called the 'Transverse Electric or TE mode'. m is put as a suffix to TE to indicate the order of the mode. A TE_m mode therefore is a transverse electric mode with fields having m half cycle variations in the transverse plane. As discussed above no fields exist for $m = 0$ and consequently m should be 1, 2,...,etc.

EXAMPLE 6.1 An air filled parallel plane waveguide has 10 cm height. The maximum peak electric field measured inside the waveguide is 10 V/m. If the frequency of the wave is 3 GHz and if the waveguide is excited in TE_1 mode, find the expressions for the electric and the magnetic fields inside the waveguide.

Solution:

Since, the waveguide is air filled, we have,

$$\beta_1 = \frac{\omega}{c} = \frac{2\pi \times 3 \times 10^9}{3 \times 10^8} = 20\pi \text{ rad/m}.$$

$$\Rightarrow \qquad \lambda_1 = \frac{2\pi}{\beta_1} = \frac{2\pi}{20\pi} = 0.1 \text{ m}.$$

Also $\quad \eta_1 = 120\pi$

The height of the waveguide d = 10 cm = 0.1 m.

For TE_1 mode, $\quad m = 1$

Since, the peak electric field is 10 V/m, from Eqn (6.10), We get $2E_{i0}$= 10V/m.

Assuming that the mode travels along the z-direction, the phase constant of the mode (from Eqn (6.8)) is

$$\beta = \beta_1\sqrt{1-\left(\frac{\lambda_1}{2d}\right)^2} = 20\pi\sqrt{1-\left(\frac{0.1}{2 \times 0.1}\right)^2}$$

$$= 10\sqrt{3}\pi \text{ rad/m}$$

From Eqns (6.10) and (6.11), we can write the fields as

$$\mathbf{E} = 10j\sin\left(\frac{\pi x}{0.1}\right)e^{-j10\sqrt{3}\pi z}\hat{\mathbf{y}}$$

$$\mathbf{H} = \frac{10}{120\pi}\left\{ j\sin\left(\frac{\pi x}{0.1}\right)\left(\frac{\pi}{20\pi \times 0.1}\right)\hat{\mathbf{x}} + \frac{10\sqrt{3}\pi}{20\pi}\cos\left(\frac{\pi x}{0.1}\right)\hat{\mathbf{z}}\right\} e^{-j10\sqrt{3}\pi}$$

Simplifying the expressions we get

$$\mathbf{E} = 10j\sin(10\pi x)e^{-j10\sqrt{3}\pi z}\hat{\mathbf{y}} \text{ V/m}$$

$$\mathbf{H} = \frac{1}{12\pi}\left\{\frac{j}{2}\sin(10\pi x)\hat{\mathbf{x}} + \frac{\sqrt{3}}{2}\cos(10\pi x)\hat{\mathbf{z}}\right\} e^{-j10\sqrt{3}\pi z} \text{ A/m}$$

6.1.2 Transverse Magnetic (TM) Mode

In the previous section, we visualized the TE modes in terms of the reflection of a plane wave with perpendicular polarization. Instead, if we take parallel polarization, we get the Transverse Magnetic (TM) modes.

Let us consider the coordinate system as shown in Fig. 6.4 and a plane wave with parallel polarization incident at an angle θ. Using the same coordinate transformation as used in the TE case, the fields in medium 1 can be written as

$$\mathbf{E}_1 = 2E_{i0}e^{-j\beta_1 z\sin\theta}\left[\sin\theta\cos(\beta_1 x\cos\theta)\hat{\mathbf{x}} + j\cos\theta_i\sin(\beta_1 x\cos\theta)\hat{\mathbf{z}}\right] \quad (6.12)$$

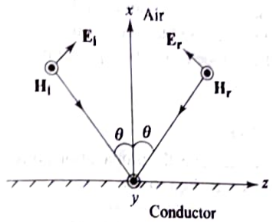


Fig. 6.4 *Reflection of a uniform plane wave with parallel polarization giving rise to TM modes.*

$$\mathbf{H}_1 = 2\frac{E_{i0}}{\eta_1}\cos(\beta_1 x\cos\theta)e^{-j\beta_1 z\sin\theta}\hat{\mathbf{y}} \quad (6.13)$$

The component of the electric field which is parallel to the conducting boundary is the z component. The z component of the electric field goes to zero when

$$\beta_1 x\cos\theta = m\pi \qquad m = 0, 1, 2, \ldots \quad (6.14)$$

This condition is identical to Eqn (6.3), and therefore all the arguments presented in the previous section are applicable in this case also i.e. we can place another conducting boundary at a distance x given by Eqn (6.14) without affecting the fields. Alternatively, for a given separation 'd' between the two boundaries, there are definite angles of incidence only for which the fields can exist. These angles are given by

$$\cos\theta = \frac{m\pi}{\beta_1 d} = \frac{m\pi}{\frac{2\pi d}{\lambda_1}} = \frac{m\lambda_1}{2d} \quad (6.15)$$

$$\Rightarrow \qquad \sin\theta = \frac{\beta}{\beta_1} \quad (6.16)$$

Equation (6.15) is identical to Eqn (6.4), i.e. the angles of incidence for which the fields can exist between two conducting boundaries are same for parallel and perpendicular polarizations.

Substituting for $\cos\theta$ from Eqn (6.15) and appropriately for $\sin\theta$ from Eqn (6.16) the fields between two conducting boundaries with parallel polarization can be written as

$$\mathbf{E}_1 = 2jE_{i0}\sin\left(\frac{m\pi x}{d}\right)e^{-j\beta z}\left(\frac{m\pi}{\beta_1 d}\right)\hat{\mathbf{z}} + 2E_{i0}\frac{\beta}{\beta_1}\cos\left(\frac{m\pi x}{d}\right)e^{-j\beta z}\hat{\mathbf{x}} \quad (6.17)$$

$$\mathbf{H}_1 = \frac{2E_{i0}}{\eta_1}\cos\left(\frac{m\pi x}{d}\right)e^{-j\beta z}\hat{\mathbf{y}} \quad (6.18)$$

From Eqns (6.17) to (6.18), we note that, in this case also the fields travel in $+z$ direction, however, there is no component of magnetic field along the direction of the wave propagation. This mode is hence called the Transverse Magnetic (TM)

mode. Again, m is used as suffix to TM and it indicates the order of the mode. It is worth noting here that in this case the fields do not vanish for $m = 0$, therefore, the TM_0 mode exists. This mode is a rather special mode and will be discussed in detail in the later sections.

6.1.3 Cut-off Frequency

The modal propagation constant, i.e. the propagation constant of the modal fields along the z-direction is given as (Eqn (6.8)).

$$\beta = \beta_1 \sqrt{1 - \left(\frac{m\lambda_1}{2d}\right)^2} \tag{6.19}$$

$$= \sqrt{\beta_1^2 - \left(\frac{m\pi}{d}\right)^2} \tag{6.20}$$

From the field expressions it is clear that for the fields to be traveling, the phase constant β should be real. If β becomes imaginary, the term $-j\beta z$ becomes real and the function does not represent wave propagation. The condition for the wave propagation is therefore

$$\beta = \text{real} \tag{6.21}$$

$$\Rightarrow \quad \beta_1 \geq \frac{m\pi}{d} \tag{6.22}$$

Since

$$\beta_1 = \frac{2\pi}{\lambda_1} = \frac{2\pi f}{v_1},$$

where, v_1 is the velocity of the uniform plane wave in medium 1, we get

$$f \geq \frac{mv_1}{2d} \tag{6.23}$$

$$\Rightarrow \quad \lambda_1 \leq \frac{2d}{m} \tag{6.24}$$

Two things can be noted from Eqns (6.23) and (6.24):

1. For a given waveguide height d, the frequency has to be higher than certain threshold for propagation of a particular mode. The threshold frequency is called the cut-off frequency of the mode and is given by

$$f_{cm} = \frac{mv_1}{2d} \tag{6.25}$$

The corresponding cut-off wavelength is

$$\lambda_{cm} = \frac{2d}{m} \tag{6.26}$$

2. For a given waveguide height, d, and frequency, f, only those modes propagate for which $m \leq \frac{2df}{v_1} = \frac{2d}{\lambda}$. In other words, at any given frequency, there is a possibility of propagation of finite number of modes. We say here

'possibility', because even though the waveguide can support a particular mode, the excitation source may not excite that mode.

The cut-off frequencies for different modes are shown in Fig. 6.5. All the modes which have cut off frequencies less than the frequency of the excitation field, can propagate inside the waveguide. Note that, TM_0 mode has no cut-off frequency, i.e. it can propagate at any arbitrarily small frequency. All other modes have their respective cut-off frequencies.

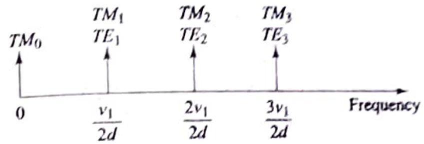


Fig. 6.5 *Cut-off frequencies for different TE and TM modes.*

6.1.4 Phase Velocity and Dispersion

It is clear from earlier sections that, the fields which exist between two parallel conducting boundaries are the same as one would get from superposition of the incident and the reflected waves at the boundary. It should be noted however, that in a parallel plane geometry there is nothing like a reflected or incident wave. Let us consider reflection of a plane wave between two planes as shown in Fig. 6.6. Wave 2 is a reflected wave for plane P_1 but is an incident wave for plane P_2 and wave 3 is incident wave for plane P_1 and reflected wave for plane P_2.

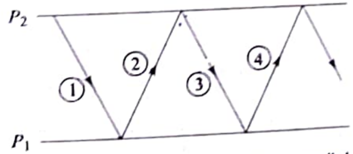


Fig. 6.6 *Multiple reflection of plane waves between parallel conducting planes.*

There is no sanctity for the incident or reflected wave and what we have is just a set of two waves one going downwards like 1, 3, etc, and other going upwards like 2, 4, etc. The propagation of electromagnetic waves between two conducting planes can therefore either be visualized as multiple reflections of a plane wave at the two planes as shown in Fig. 6.6 or superposition of two plane waves as shown in Fig. 6.7. The two waves make the same angle with respect to $+z$ axis, the direction of net wave propagation. We can now find the phase velocity of the modal fields either from the first principles or from the knowledge of the phase constant in the z-direction.

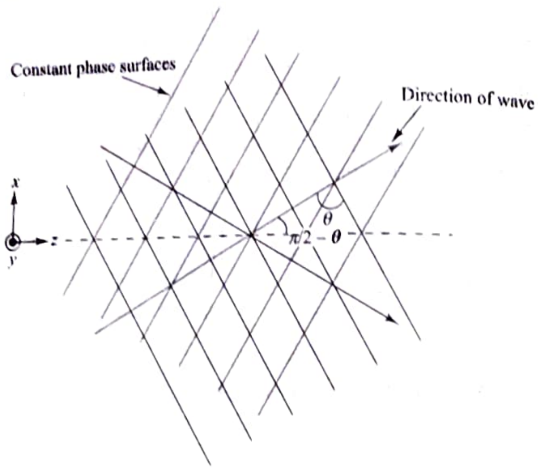


Fig. 6.7 *Superposition of two plane waves travelling at an angle.*

The phase velocity of the mode is

$$v_p = \frac{\omega}{\beta} = \frac{\omega}{\beta_1\sqrt{1-\left(\frac{m\lambda_1}{2d}\right)^2}} = \frac{v_1}{\sqrt{1-\left(\frac{m\lambda_1}{2d}\right)^2}} \tag{6.27}$$

where v_1 is the velocity of the plane wave in the medium filling the region between the two planes. Noting from Eqn (6.26) that $\frac{2d}{m} = \lambda_{cm}$, the phase velocity can be written as

$$v_p = \frac{v_1}{\sqrt{1-\left(\frac{\lambda_1}{\lambda_{cm}}\right)^2}} = \frac{v_1}{\sqrt{1-\left(\frac{f_{cm}}{f}\right)^2}} \tag{6.28}$$

The phase velocity of a mode is, therefore, a function of frequency. This phenomenon is called the 'wave dispersion'. It is interesting to note that, even though the medium filling the waveguide is non-dispersive, the waveguide is a dispersive structure. Referring to Fig. 6.7 we can see that the group velocity of the mode along z direction is

$$v_g = v_1 \sin\theta = v_1\sqrt{1-\left(\frac{f_{cm}}{f}\right)^2} \tag{6.29}$$

A plot of phase velocity and group velocity as a function of frequency is shown in Fig. 6.8. At the cut-off of a mode, the phase velocity is infinite and the group velocity is zero. As the frequency increases, both phase and group velocities asymptotically approach the velocity of the plane wave in the medium.

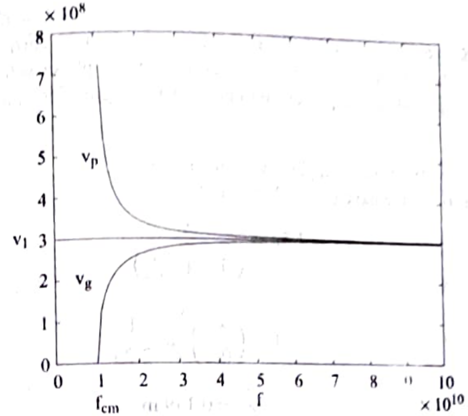


Fig. 6.8 *Group and phase velocities of a mode as a function of frequency.*

For multimode propagation, the phase/group velocity diagram is as shown in Fig. 6.9. It is clear from the Fig. 6.9 that at a given frequency, different modes travel with different phase/group velocities. Also, they travel with different velocities at different frequencies, the only exception being the TM_0 mode which is non-dispersive. For every mode, since the group velocity approaches zero at the cut-off, the energy propagation ceases at the cut-off frequency.

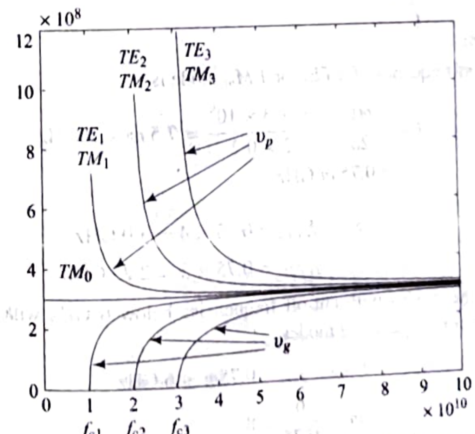


Fig. 6.9 *Variation of v_p and v_g for different modes inside a parallel plane waveguide.*

EXAMPLE 6.2 An air-filled parallel plane waveguide carries the TM_2 mode. The height of the waveguide is 20 cm. If the phase velocity of the mode is 1.5 c, find the frequency and guided wavelength of the mode.

Solution:

Intrinsic velocity of uniform planewave in air is $v_1 = c$
From Eqn (6.28) we have ($m = 2, d = 0.2$ m)

$$v_p = 1.5c = \frac{c}{\sqrt{1-\left(\frac{2\lambda_1}{2\times 0.2}\right)^2}}$$

$$\Rightarrow \quad 1-\left(\frac{\lambda_1}{0.2}\right)^2 = \frac{1}{2.25}$$

$$\Rightarrow \quad \lambda_1 = 0.149 \text{ m}$$

The frequency of the wave is $\quad f = \dfrac{c}{\lambda_1} = 2.012$ GHz

The guided wavelength $\quad \lambda_g = \dfrac{v_p}{f} = \dfrac{1.5c}{2.012\times 10^9} = 0.298$ m

EXAMPLE 6.3 In Example 6.2, find the cut off frequencies of TM_3 and TE_4 modes. How many TE and TM modes will propagate in the waveguide at 6 GHz?

Solution:

The cut-off frequency of a TE_m or TM_m mode is

$$f_{cm} = \frac{mv}{2d} = \frac{m\times 3\times 10^8}{2\times 0.2} = 7.5\, m\times 10^8 \text{ Hz}$$
$$= 0.75\, m \text{ GHz}$$

$$\Rightarrow \quad f_{cTE_4} = 0.75\times 4 = 3.0 \text{ GHz}$$

and $\quad f_{cTM_3} = 0.75\times 3 = 2.25$ GHz

Now, all the modes with cut-off frequencies below 6 GHz will propagate. Therefore, for propagating modes

$$f_{cm} = 0.75m \le 6 \text{ GHz}$$

$$m \le \frac{6}{0.75} = 8$$

Since, cutoff frequencies for TM_m and TE_m are equal we have $TE_1, \ldots, TE_7$, and $TM_1, \ldots, TM_7$ propagating modes. Of course this does not include the modes corresponding to $m = 0$ which will be discussed later.

EXAMPLE 6.4 In a parallel plate waveguide, the phase velocity of TE_3 mode is 1.4 c. Find the guided wavelength of TM_2 mode inside the waveguide. The waveguide has been filled with a material having dielectric constant 9 and frequency of the wave is 1 GHz.

Solution:

The intrinsic velocity of a wave in the material filling the waveguide is

$$v_1 = \frac{c}{\sqrt{\epsilon_r}} = \frac{3\times 10^8}{\sqrt{9}} = 10^8 \text{ m/s} = c/3$$

From Eqn (6.27) we get for TE_3 mode ($m = 3$)

$$\left(\frac{m\lambda_1}{2d}\right) = \sqrt{1-\left(\frac{v_1}{v_p}\right)^2} = \sqrt{1-\left(\frac{c/3}{1.4c}\right)^2}$$

$v_p\sqrt{1-\left(\frac{m\lambda_1}{2d}\right)^2} = \frac{v_1}{\Phi}$

$$\frac{3\lambda_1}{2d} = \sqrt{1-\left(\frac{1}{4.2}\right)^2} = 0.58$$

$$\Rightarrow \quad \frac{\lambda_1}{d} = 2\times 0.58/3 = 0.387$$

The phase velocity of TM_2 mode will be

$$v_{pTM_2} = \frac{v_1}{\sqrt{1-\left(\frac{2\lambda_1}{2d}\right)^2}}$$
$$= \frac{10^8}{\sqrt{1-\left(\frac{\lambda_1}{d}\right)^2}} = 1.229\times 10^8 \text{ m/s}$$

The guided wavelength $\lambda_g = \frac{v_{pTM_2}}{1GHz} = 0.122$ m

6.2 TRANSVERSE ELECTROMAGNETIC (TEM) MODE

We have seen in the previous section that the TE_0 mode does not exist but the TM_0 mode can exist as both electric and magnetic fields are non-zero for the mode. Substituting $m = 0$ in Eqns (6.17), (6.18) and (6.19), we get the phase constant

$$\beta = \beta_1$$

and the modal fields as (6.30)

$$\mathbf{E}_1 = 2E_{i0}e^{-j\beta_1 z}\hat{\mathbf{x}}$$

$$\mathbf{H}_1 = \frac{2E_{i0}}{\eta_1} e^{-j\beta_1 z}\hat{\mathbf{y}} \tag{6.31}$$

It can be noted that for $m = 0$, neither electric nor magnetic field has any component in the direction of the wave propagation. Also, from Eqns (6.30) and (6.31) we can see that the electric field is along the x- direction, and the magnetic field is along y-direction, i.e. the electric field, the magnetic field and the direction of modal propagation are perpendicular to each other. This mode is then similar to the TEM wave in an unbound medium. Like a plane wave in unbound medium, for this mode also the ratio of $|\mathbf{E}|$ and $|\mathbf{H}|$ is

$$\frac{|\mathbf{E}|}{|\mathbf{H}|} = \eta_1 = \text{Intrinsic impedance of the medium} \tag{6.32}$$

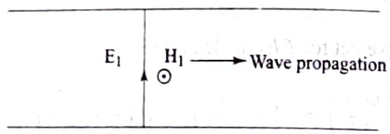


Fig. 6.10 *Electric and magnetic fields for the TEM mode inside a parallel plane waveguide.*

One would then wonder, why this mode of the parallel plane waveguide has identical characteristics to that of the wave propagation in the unbound medium. "Are the conducting boundaries not playing any role!" In fact the boundaries are ineffective in propagation of this mode. The E-field is intrinsically normal to the conducting planes and the magnetic field is intrinsically tangential to the conducting planes. That is, inherently the tangential component of the electric field and the normal component of the magnetic field are zero at the waveguide planes. In other words, the fields inherently satisfy the boundary conditions and hence the propagation of the wave is unaltered by the planes. The conducting planes therefore are transparent to the wave and the wave propagates as if it is propagating in an unbound medium.

Most of the two conductor systems like the parallel wire line, coaxial cable, the propagation of electromagnetic energy takes place through the TEM mode.

EXAMPLE 6.5 A parallel plane waveguide has 0.05 m height and is filled with material having relative permittivity 2.4. What is the frequency range over which there is single mode propagation? If the magnetic field for the mode is 0.1 A/m. Find the elctric field inside the waveguide.

Solution:

Intrinsic velocity in the material filling the waveguide

$$v_1 = \frac{c}{\sqrt{\epsilon_r}} = \frac{3 \times 10^8}{\sqrt{2.4}} = 1.936 \times 10^8 \text{ m/s}$$

The intrinsic impedance of the medium filling the waveguide is

$$\eta_1 = \frac{\eta_0}{\sqrt{\epsilon_r}} = \frac{120\pi}{\sqrt{2.4}} = 243.34\ \Omega$$

The first mode which propagates on the parallel plane waveguide is the TEM mode which has zero cut off frequency. The frequency range over which single mode operation exists is from 0 to f_{c_1}

$$f_{c_1} = \frac{v_1}{2d} = \frac{1.936 \times 10^8}{2 \times 0.05} = 1.936 \text{ GHz.}$$

For TEM mode,

$$\frac{|\mathbf{E}|}{|\mathbf{H}|} = \eta_1$$

$$\Rightarrow \quad |\mathbf{E}| = \eta_1 |H| = 243.34 \times 0.1 \text{ V/m} = 24.33 \text{ V/m}$$

EXAMPLE 6.6 In the following waveguide find the frequency at which the TE_1 mode does not have any reflection at the waveguide junction at $z = 0$. Here, $\epsilon_{r1} = 4$, $\mu_{r1} = 10$ and $\epsilon_{r2} = 1$

Solution:

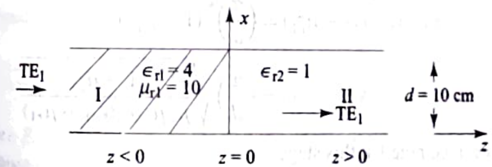


Fig. 6.11 *Parallel plane waveguide partly filled with a dielectric material.*

We can write the fields for $z < 0$ and $z > 0$ separately. Since there is no reflections at $z = 0$, the modal fields in regions I and II must satisfy the boundary conditions, i.e. the tangential component of E and normal component of H at the dielectric junction must be continuous, (Fig. 6.11)

That is to say that at $z = 0$

$$Ey_1 = Ey_2$$
$$Hz_1 = Hz_2$$

The modal fields will have different phase constants in two regions. Let they be denoted by βz_1 and βz_2 respectively.

From Eqns (6.10) and (6.11) we get

$$E_{i0_1}\sin\left(\frac{\pi x}{d}\right) = E_{i0_2}\sin\left(\frac{\pi x}{d}\right)$$

$$\Rightarrow E_{i0_1} = E_{i0_2}$$

and
$$\frac{E_{i0_1}}{\eta_1}\left(\frac{\beta_{z1}}{\beta_1}\right)\sin\left(\frac{\pi x}{d}\right) = \frac{E_{i0_2}}{\eta_2}\left(\frac{\beta_{z2}}{\beta_2}\right)\sin\left(\frac{\pi x}{d}\right)$$

$$\Rightarrow \quad \frac{\beta_{z1}}{\eta_1\beta_1} = \frac{\beta_{z2}}{\beta_2\eta_2}$$

Now, $\beta_1\eta_1 = \omega\sqrt{\mu_0\epsilon_0}\sqrt{\epsilon_{r1}\mu_{r1}}\sqrt{\dfrac{\mu_0\mu_{r1}}{\epsilon_0\epsilon_{r1}}} = \omega\mu_0\mu_{r1}$

and $\beta_2\eta_2 = \omega\mu_0$

Therefore we get $\omega\mu_0\beta_{z1} = \beta_{z2}\omega\mu_0\mu_{r1}$

$$\sqrt{\beta_1^2 - \left(\frac{\pi}{d}\right)^2} = \mu_{r1}\sqrt{\beta_2^2 - \left(\frac{\pi}{d}\right)^2}$$

$$\Rightarrow \quad \beta_1^2 - \mu_{r1}^2\beta_2^2 = \left(\frac{\pi}{d}\right)^2(1-\mu_{r1}^2)$$

Now, noting that $\beta_1^2 = \omega^2\mu_1\epsilon_1 = \omega^2\mu_0\epsilon_0\mu_{r1}\epsilon_{r1}$

and $\beta_2^2 = \omega^2\mu_0\epsilon_0$

we get $\omega^2\mu_0\epsilon_0(\mu_{r1}\epsilon_{r1} - \mu_{r1}^2) = \left(\dfrac{\pi}{d}\right)^2(1-\mu_{r1}^2)$

$$\Rightarrow \quad \omega = \left(\frac{\pi}{d}\right)\sqrt{\frac{1-\mu_{r1}^2}{\mu_0\mu_{r1}\epsilon_0(\epsilon_{r1}-\mu_{r1})}}$$

Two points can be noted at this stage:

(i) For reflection-less condition for a TE mode, one of the media has to be magnetic.

(ii) Since μ_{r1} is greater than or equal to unity, for ω to be real, μ_{r1} has to be $> \epsilon_{r1}$

Substituting the waveguide parameters we get

$$2\pi f = \frac{\pi}{0.1}\sqrt{\frac{(1-100)}{10(4-10)\mu_0\epsilon_0}}$$

$$f = \frac{1}{2}\sqrt{\frac{99}{60}} \times 3 \times 10^9 \text{ Hz}$$
$$= 1.927 \text{ GHz}$$

6.3 ANALYSIS OF WAVEGUIDE—GENERAL APPROACH

In the previous section, we developed the concept of modal propagation in a bound structure like the parallel plane waveguide. The modal propagation has been visualized as superposition of multiple times reflected plane waves. This visualization helps in understanding, how modal fields are generated in a waveguide, and provides a physical insight to the modal propagation. This approach though elegant, becomes algebraically unmanageable for complex wave guiding structures. In these situations one uses the general solution of the wave equation in a given guiding structure. In the following sections first we develop the general frame work for propagating fields in a bound structure and show that the solution obtained by this approach is indeed the same as that obtained by the superposition of the plane waves.

We note that, since the various components of the electric and magnetic fields are related through the Maxwell's equations, all the field components cannot be independent. We select some field components as independent components and express the remaining field components as functions of the chosen independent components. Without losing generality, if we assume that the wave is moving in z-direction, the field components which are along the z-direction, have special significance as they represent the longitudinal field components. The other four field components, lie in a plane transverse to the direction of the wave propagation and hence can be called the transverse components. Since, the longitudinal components have special significance, let us assume E_z and H_z to be independent and let us express the transverse components in terms of E_z and H_z.

Let us first separate the longitudinal and transverse field components and write the total fields as

$$\mathbf{E} = \mathbf{E}_\perp + E_z\hat{\mathbf{z}} \tag{6.33}$$

$$\mathbf{H} = \mathbf{H}_\perp + H_z\hat{\mathbf{z}} \tag{6.34}$$

where $\mathbf{E}_\perp$ and $\mathbf{H}_\perp$ are vectors lying in the transverse plane (xy-plane in the Cartesian coordinate system). E_z and H_z are scalar quantities as their direction $\hat{\mathbf{z}}$ has been taken out explicitly.

Let the differential vector operator ∇ also be split into its transverse and longitudinal parts as

$$\nabla = \nabla_\perp + \frac{\partial}{\partial z}\hat{\mathbf{z}} \tag{6.35}$$

where $\nabla_\perp$ is derivative vector operator in the transverse plane. In Cartesians coordinate system

$$\nabla_\perp \equiv \frac{\partial}{\partial x}\hat{\mathbf{x}} + \frac{\partial}{\partial y}\hat{\mathbf{y}} \tag{6.36}$$

Now, substituting for ∇, $\mathbf{E}$ and $\mathbf{H}$ in the Maxwell's equation for the source free homogenous medium (Eqn (4.3)),

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{6.37}$$

we get

$$\left[\nabla_\perp + \frac{\partial}{\partial z}\hat{\mathbf{z}}\right] \times (\mathbf{E}_\perp + E_z\hat{\mathbf{z}}) = -j\omega\mu(\mathbf{H}_\perp + H_z\hat{\mathbf{z}}) \tag{6.38}$$

$$\Rightarrow \nabla_\perp \times \mathbf{E}_\perp + \nabla_\perp \times (E_z\hat{\mathbf{z}}) + \frac{\partial}{\partial z}\hat{\mathbf{z}} \times \mathbf{E}_\perp + \frac{\partial}{\partial z}\hat{\mathbf{z}} \times (E_z\hat{\mathbf{z}}) = -j\omega\mu(\mathbf{H}_\perp + H_z\hat{\mathbf{z}}) \tag{6.39}$$

In Eqn (6.39) we note the following points

1. Any cross product with $\hat{\mathbf{z}}$ lies in the transverse plane.
2. Cross product of the two vectors in the transverse plane is along $\hat{\mathbf{z}}$ direction and hence represents a longitudinal vector.
3. Cross product of $\hat{\mathbf{z}} \times \hat{\mathbf{z}}$ is identically zero.

The first term on the LHS of Eqn (6.39) is therefore a longitudinal vector and the last term on the LHS is zero.

The equality of the transverse fields on two sides of the equation yields

$$-j\omega\mu\mathbf{H}_\perp = \nabla_\perp \times (E_z\hat{\mathbf{z}}) + \frac{\partial}{\partial z}\hat{\mathbf{z}} \times \mathbf{E}_\perp \tag{6.40}$$

$$\Rightarrow \quad \mathbf{H}_\perp = \frac{-1}{j\omega\mu}\left\{\nabla_\perp \times (E_z\hat{\mathbf{z}}) + \frac{\partial}{\partial z}\hat{\mathbf{z}} \times \mathbf{E}_\perp\right\} \tag{6.41}$$

Similarly from the Maxwell's Eqn (4.4), we get

$$\mathbf{E}_\perp = \frac{1}{j\omega\epsilon}\left[\nabla_\perp \times (H_z\hat{\mathbf{z}}) + \frac{\partial}{\partial z}\hat{\mathbf{z}} \times \mathbf{H}_\perp\right] \tag{6.42}$$

Substituting for $\mathbf{H}_\perp$ from Eqn (6.41) into Eqn (6.42), we get

$$\mathbf{E}_\perp = \frac{1}{j\omega\epsilon}\nabla_\perp \times (H_z\hat{\mathbf{z}}) + \frac{1}{j\omega\epsilon}\frac{\partial}{\partial z}\hat{\mathbf{z}} \times \frac{-1}{j\omega\mu}\left[\nabla_\perp \times (E_z\hat{\mathbf{z}}) + \frac{\partial}{\partial z}\hat{\mathbf{z}} \times \mathbf{E}_\perp\right] \tag{6.43}$$

Bringing the term of $\mathbf{E}_\perp$ on the left side of the equation and rearranging the terms we get

$$\left\{\omega^2\mu\epsilon\mathbf{E}_\perp - \frac{\partial}{\partial z}\hat{\mathbf{z}} \times \frac{\partial}{\partial z}\hat{\mathbf{z}} \times \mathbf{E}_\perp\right\} = -j\omega\mu\nabla_\perp \times (H_z\hat{\mathbf{z}}) + \left(\frac{\partial}{\partial z}\hat{\mathbf{z}}\right) \times \nabla_\perp \times (E_z\hat{\mathbf{z}}) \tag{6.44}$$

Using vector triple product identity

$$\mathbf{A} \times \mathbf{B} \times \mathbf{C} = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}$$

we can simplify the Eqn (6.44) noting the following points:

(i) $$\frac{\partial}{\partial z}\hat{\mathbf{z}} \times \frac{\partial}{\partial z}\hat{\mathbf{z}} \times \mathbf{E}_\perp = \frac{\partial}{\partial z}\left(\frac{\partial}{\partial z}\hat{\mathbf{z}} \cdot \mathbf{E}_\perp\right)\hat{\mathbf{z}} - \left(\frac{\partial}{\partial z}\hat{\mathbf{z}} \cdot \frac{\partial}{\partial z}\hat{\mathbf{z}}\right)\mathbf{E}_\perp$$

$$= 0 - \frac{\partial^2}{\partial z^2}\mathbf{E}_\perp \tag{6.45}$$

(ii) $$\frac{\partial}{\partial z}\hat{\mathbf{z}} \times \nabla_\perp \times (E_z\hat{\mathbf{z}}) = \nabla_\perp\left(\frac{\partial}{\partial z}\hat{\mathbf{z}} \cdot E_z\hat{\mathbf{z}}\right) - \left(\frac{\partial}{\partial z}\hat{\mathbf{z}} \cdot \nabla_\perp\right)E_z\hat{\mathbf{z}}$$

$$= \nabla_\perp\left(\frac{\partial E_z}{\partial z}\right) - 0 \tag{6.46}$$

Equation (6.44), therefore, finally becomes

$$\left[\omega^2\mu\epsilon\mathbf{E}_\perp + \frac{\partial^2}{\partial z^2}\mathbf{E}_\perp\right] = -j\omega\mu\nabla_\perp \times (H_z\hat{\mathbf{z}}) + \nabla_\perp\left(\frac{\partial E_z}{\partial z}\right) \tag{6.47}$$

If we now assume that the wave has the propagation constant γ, all field components will have z-variation as $e^{-\gamma z}$. The derivatives with respect to z can then be written as

$$\frac{\partial}{\partial z} \equiv -\gamma \tag{6.48}$$

and

$$\frac{\partial^2}{\partial z^2} \equiv \gamma^2 \tag{6.49}$$

Substituting for derivatives with respect to z in Eqn (6.47) we get

$$(\omega^2\mu\epsilon + \gamma^2)\,\mathbf{E}_\perp = -j\omega\mu\nabla_\perp \times (H_z\hat{\mathbf{z}}) - \gamma\nabla_\perp E_z \tag{6.50}$$

For brevity sake, defining

$$\omega^2\mu\epsilon + \gamma^2 = h^2 \tag{6.51}$$

the transverse electric field can be written as

$$\mathbf{E}_\perp = \frac{-j\omega\mu}{h^2}\nabla_\perp \times (H_z\hat{\mathbf{z}}) - \frac{\gamma}{h^2}\nabla_\perp(E_z) \tag{6.52}$$

On similar lines, the transverse magnetic field can be written as

$$\mathbf{H}_\perp = \frac{j\omega\epsilon}{h^2}\nabla_\perp \times (E_z\hat{\mathbf{z}}) - \frac{\gamma}{h^2}\nabla_\perp(H_z) \tag{6.53}$$

For a loss-less medium, γ is purely imaginary,

$$\gamma = j\beta \tag{6.54}$$

and we have

$$h^2 = \omega^2\mu\epsilon - \beta^2 \tag{6.55}$$

Note that in Eqn (6.55), $\omega^2\mu\epsilon$ is the square of the phase constant in the medium, and β is the propagation constant in the z direction, from the law of direction cosines, h represents the phase constant of the wave in the transverse plane.

From Eqns (6.52) and (6.53) we can make following interesting observations:

(i) Transverse fields can exist provided at least one of the longitudinal components (E_z or H_z) is non-zero, except when $h = 0$. That is, in general there is no transverse electromagnetic wave propagation except when $h = 0$.

(ii) For $h = 0$, we get $\beta = \omega\sqrt{\mu\epsilon}$ implying that the transverse electromagnetic (TEM) wave can exist in a waveguide if its propagation constant is same as that of the unbound medium filling the waveguide. In other words, the TEM wave is essentially non-dispersive.

(iii) The fields corresponding to $E_z = 0$, have electric field transverse to the direction of wave propagation and hence represent the Transverse electric (TE) wave.

(iv) The fields corresponding to $H_z = 0$ have magnetic field transverse to the direction of wave propagation and hence represent Transverse magnetic (TM) wave.

(v) For TM or TE case, since E_z or H_z respectively is to be non zero, h essentially has to be non zero. Otherwise the transverse fields would become infinite. In other words, the TE and TM modes cannot have the phase constant same as that of the unbound medium. The TE and TM modes then essentially have to be dispersive modes, i.e. their phase velocity should vary as a function of frequency.

Before we proceed further, let us write the expressions for the transverse field components in the Cartesian system. The transverse curl and gradient can be written as

$$\nabla_\perp \times (\psi\hat{\mathbf{z}}) \equiv \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & 0 \\ 0 & 0 & \psi \end{vmatrix} \tag{6.56}$$

$$= \frac{\partial \psi}{\partial y}\hat{\mathbf{x}} - \frac{\partial \psi}{\partial x}\hat{\mathbf{y}} \tag{6.57}$$

and

$$\nabla_\perp \psi = \frac{\partial \psi}{\partial x}\hat{\mathbf{x}} + \frac{\partial \psi}{\partial y}\hat{\mathbf{y}} \tag{6.58}$$

Here, ψ is a scalar function of (x, y, z). ψ can either be H_z or E_z.

Separation of x and y components in Eqns (6.52) and (6.53) yields

$$E_x = -\frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial y} - \frac{j\beta}{h^2}\frac{\partial E_z}{\partial x} \tag{6.59}$$

$$E_y = \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x} - \frac{j\beta}{h^2}\frac{\partial E_z}{\partial y} \tag{6.60}$$

$$H_x = \frac{j\omega\epsilon}{h^2}\frac{\partial E_z}{\partial y} - \frac{j\beta}{h^2}\frac{\partial H_z}{\partial x} \tag{6.61}$$

$$H_y = -\frac{j\omega\epsilon}{h^2}\frac{\partial E_z}{\partial x} - \frac{j\beta}{h^2}\frac{\partial H_z}{\partial y} \tag{6.62}$$

It is clear from Eqn (6.59) to Eqn (6.62) that if we know E_z and/or H_z, the transverse fields can be derived. Indeed E_z and H_z are to be obtained independently from the wave equation.

For source-free homogenous medium, E_z and H_z satisfy the wave equation

$$\nabla^2(E_z, H_z) + \omega^2\mu\epsilon(E_z, H_z) = 0 \tag{6.63}$$

The analysis of waveguide in general has to be carried out in two steps:

(i) First solve the wave equation to get the general solution of E_z and/or H_z.

(ii) Substitute E_z and/or H_z in Eqn (6.59) to Eqn (6.62), find transverse components, and apply appropriate boundary conditions to get the solution to a particular waveguide geometry.

In the following sections, we analyse the modal fields for a rectangular waveguide and subsequently obtain the results for the parallel plane waveguide as a limiting case of the rectangular waveguide.

6.4 RECTANGULAR WAVEGUIDES

A rectangular waveguide is a hollow metallic pipe with rectangular cross-section as shown in Fig. 6.12. The walls of the waveguide have infinite conductivity and the medium filling the waveguide is an ideal dielectric with permeability μ, permittivity ϵ and conductivity $\sigma = 0$. The waveguide has a cross-section $(a \times b)$ and is assumed to be infinitely long. Without losing generality we follow a convention that $a \geq b$, and the x-axis is oriented along the broader cross sectional dimension, and the y-axis is along the shorter cross-sectional dimension of the waveguide. The z-axis is oriented along the length of the waveguide.

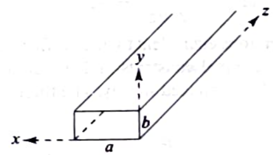


Fig. 6.12 *Rectangular waveguide.*

As discussed above, the fields have to be either TE or TM i.e. either E_z or H_z has to be non zero. The TEM fields do not exist in a rectangular waveguide and the reason for that will be mentioned a little later. In the following section, therefore, we exclusively analyze the TE and TM modes of a rectangular waveguide.

6.4.1 Transverse Magnetic (TM) Mode

For a transverse magnetic mode, we have $H_z = 0$ and $E_z \neq 0$. Writing the wave Eqn (6.63) for E_z in the Cartesian coordinate system we have

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} + \omega^2\mu\epsilon E_z = 0 \tag{6.64}$$

The equation can be solved by separation of variables. Let E_z be a product of functions of each variable x, y, z as

$$E_z(x, y, z, t) = X(x)Y(y)Z(z) \tag{6.65}$$

The time variation of $e^{j\omega t}$ is implicitly assumed for all field components. Substituting Eqn (6.65) in Eqn (6.64) yields

$$YZ\frac{d^2X}{dx^2} + XZ\frac{d^2Y}{dy^2} + XY\frac{d^2Z}{dz^2} + \omega^2\mu\epsilon XYZ = 0 \tag{6.66}$$

$$\Rightarrow \quad \frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} + \omega^2\mu\epsilon = 0 \tag{6.67}$$

Note : Partial derivatives in Eqn (6.64) have been changed to full derivatives in Eqn (6.66) as X is a function of x only, Y is a function of y only, and Z is a function of z only.

In Eqn (6.67), first three terms are functions of x, y, z only respectively and the last term is a constant. The Eqn (6.67) is to be satisfied for every value of (x, y, z). This is possible provided each term is constant. That is,

$$\frac{1}{X}\frac{d^2X}{dx^2} = -A^2 \tag{6.68}$$

$$\frac{1}{Y}\frac{d^Y}{dy^2} = -B^2 \tag{6.69}$$

and

$$\frac{1}{Z}\frac{d^2Z}{dz^2} = -\beta^2 \tag{6.70}$$

Equation (6.67) is therefore, equivalent to three differential Eqns (6.68), (6.69) and (6.70). A, B and β are real constants and the choice of negative sign in Eqns (6.68) to (6.70) will be explained shortly. The three Eqns (6.68), (6.69) and (6.70) can be written as

$$\frac{d^2X}{dx^2} + A^2X = 0 \tag{6.71}$$

$$\frac{d^2Y}{dy^2} + B^2Y = 0 \tag{6.72}$$

and

$$\frac{d^2z}{dz^2} + \beta^2Z = 0 \tag{6.73}$$

The three Eqns (6.71), (6.72) and (6.73) are identical to the transmission line equations (see Chapter 2) and their solutions are wave type of solutions. If instead of -ve sign, the positive sign was used in Eqns (6.68) to (6.70), the solutions would be exponential functions which would represent evanescent fields and not the waves.

From Fig. 6.12, it is clear that the rectangular waveguide becomes a parallel plane waveguide by making a or $b \to \infty$. The rectangular waveguide should, therefore, have fields which will reduce to the fields of parallel plane waveguide in the limit a or $b \to \infty$. In a parallel plane waveguide we have seen that the field variation in the transverse plane (xy-plane) is due to superposition of plane waves bouncing between the two planes. That is to say that the fields in transverse plane are of standing wave nature, whereas the fields have traveling wave nature in z-direction. We, therefore, expect from the understanding of the parallel plane waveguide, that the magnitude of the field in the transverse plane undergoes through maxima and minima, whereas the fields travel in the z-direction. The solutions to the Eqns (6.71), (6.72), (6.73) can be appropriately chosen as

$$X = C_1 \cos Ax + C_2 \sin Ax \text{.............. Standing wave} \tag{6.74}$$

$$Y = C_3 \cos By + C_4 \sin By \text{........... Standing wave} \tag{6.75}$$

and

$$Z = C_5 e^{-j\beta z} + C_6 e^{+j\beta z} \text{.......... Traveling waves} \tag{6.76}$$

where $C_1, C_2, \ldots, C_6$ are the arbitrary constants which are to be evaluated from the boundary conditions.

In general, there are two traveling modes, one travelling in $+z$-direction and other travelling in $-z$ direction. If we, however, assume that the waveguide is infinitely long, and there is only one mode travelling in $+z$-direction, we can choose

$$C_6 \equiv 0 \tag{6.77}$$

Substituting for X, Y and Z from Eqns (6.74) to (6.76) in Eqn (6.65) the field can be written as

$$E_z = C_5(C_1 \cos Ax + C_2 \sin Ax)(C_3 \cos By + C_4 \sin By)e^{-j\beta z} \tag{6.78}$$

Applying boundary condition that the tangential component of the electric field should be zero at the conducting boundary, we get

$$E_z = 0 \quad \text{at} \quad x = 0, \quad x = a$$
$$y = 0, \quad y = b \tag{6.79}$$

Note that the field component E_z is tangential to all the four walls of the waveguide.

Applying Boundary condition at $x = 0$, we get

$$E_z|_{x=0} = C_5(C_1 \cos Ax)(C_3 \cos By + C_4 \sin By)e^{-j\beta z} = 0$$
$$\Rightarrow \quad C_1 = 0 \tag{6.80}$$

Substituting $C_1 = 0$ in Eqn (6.78) and applying boundary condition at $y = 0$, we get

$$E_z|_{y=0} = C_5(C_2 \sin Ax)(C_3 \cos By)e^{-j\beta z} = 0$$
$$\Rightarrow \quad C_3 = 0 \tag{6.81}$$

Substituting $C_1 = 0$ in and $C_3 = 0$ in Eqn (6.78) we get

$$E_z = C_3C_2C_4 \sin(Ax)\sin(By)e^{-j\beta z} \tag{6.82}$$

We can combine C_2, C_3, C_4, into one constant say C, to give

$$E_z = C\sin(Ax)\sin(By)e^{-j\beta z} \tag{6.83}$$

Applying now the boundary conditions at $x = a$, and $y = b$, we have

$$E_z = 0 = C\sin(Aa)\sin(Bb)e^{-j\beta z} \tag{6.84}$$

$$\Rightarrow \quad Aa = m\pi \quad \Rightarrow \quad A = \frac{m\pi}{a} \tag{6.85}$$

$$\text{and} \quad Bb = n\pi \quad \Rightarrow \quad B = \frac{n\pi}{b} \tag{6.86}$$

where m and n are integers.

Substitution of Eqns (6.85), (6.86) in Eqn (6.83) yields

$$E_z = C\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \tag{6.87}$$

This is the complete solution of the wave equation for the TM mode. Constant C remains unknown as it governs the amplitude of the wave and has nothing to do with the boundary conditions.

The other field components can now be obtained by substituting $H_z = 0$ and E_z from Eqn (6.87) into Eqns (6.59) to (6.62) as

$$E_x = \frac{-j\beta}{h^2}\frac{\partial E_z}{\partial x} = \frac{-j\beta}{h^2}\left(\frac{m\pi}{a}\right)C\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \tag{6.88}$$

$$E_y = -\frac{j\beta}{h^2}\frac{\partial E_z}{\partial y} = -\frac{j\beta}{h^2}\left(\frac{n\pi}{b}\right)C\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \tag{6.89}$$

$$H_x = \frac{j\omega\epsilon}{h^2}\frac{\partial E_z}{\partial y} = \frac{j\omega\epsilon}{h^2}\left(\frac{n\pi}{b}\right)C\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \tag{6.90}$$

$$H_y = \frac{-j\omega\epsilon}{h^2}\frac{\partial E_z}{\partial x} = \frac{-j\omega\epsilon}{h^2}\left(\frac{m\pi}{a}\right)C\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \tag{6.91}$$

Substituting Eqn (6.87) in Eqn (6.64) and comparing with Eqn (6.55), we get

$$h^2 = \omega^2\mu\epsilon - \beta^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 \tag{6.92}$$

The integers m and n together now define the order of the mode, and the mode is designated as TM_{mn} mode. Looking at the field expressions, we can make following observations regarding the TM_{mn} mode:

1. Fields exist in discrete patterns called the modes of the waveguide.
2. All field components have sinusoidal amplitude variation in x and y directions.

3. Subscripts m and n represent number of half sinusoidal cycles of the field amplitude in x and y directions respectively.
4. All transverse fields go to zero if, either m or n is zero. In other words, both the indices m and n have to be non-zero for existence of the TM mode. That is, TM_{m0} and TM_{0n} modes cannot exist. Consequently, the lowest order mode which can exist is TM_{11} mode.
5. On the waveguide walls the tangential component of the electric field is zero and the tangential component of the magnetic field is maximum. In other words, the sinusoidal function for tangential electric field is spatially shifted by quarter cycle with respect to the sinusoidal function for the tangential magnetic field.

6.4.2 Transverse Electric (TE) Mode

The analysis of the TE modes can be carried out on the lines similar to that of the TM modes. However, TE modes have little different properties. For TE modes we have

$$E_z = 0 \quad \text{and} \quad H_z \neq 0 \tag{6.93}$$

The wave equation is to be solved for H_z in this case. That is, we have to find a solution to

$$\nabla^2 H_z + \omega^2\mu\epsilon H_z = 0 \tag{6.94}$$

for the rectangular waveguide. The initial steps needed to find the general solution by separation of variables are identical to that given in the previous section. We, therefore, get a general solution of H_z same as that given by Eqn (6.78).

$$H_z = D_5(D_1\cos Ax + D_2\sin Ax)(D_3\cos By + D_4\sin By)e^{-j\beta z} \tag{6.95}$$

Just to differentiate the above expression from Eqn (6.78) we have however used different arbitrary constants D's. Till this point the analysis of TE and TM Propogation is identical. However, beyond this point the approach differs since the boundary conditions for E_z and H_z are different. In fact, since H_z is the tangential component at the waveguide walls, there is no boundary condition for H_z. (Remember, there is boundary condition only for the normal component of the magnetic field and not for the tangential component). The arbitrary constants, therefore, cannot be evaluated directly from H_z. Instead, first we have to derive the transverse components and then apply the boundary conditions.

Substituting $E_z = 0$ and H_z from Eqn (6.95) into Eqns (6.59) to (6.62) we get the transverse electric field components as

$$E_x = \frac{-j\omega\mu}{h^2}\frac{\partial H_z}{\partial y}$$

$$= \frac{-j\omega\mu}{h^2}D_5(D_1\cos Ax + D_2\sin Ax)(-D_3B\sin By + D_4B\cos By)e^{-j\beta z} \tag{6.96}$$

$$E_y = \frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x}$$
$$= \frac{j\omega\mu}{h^2} D_5(-D_1 A \sin Ax + D_2 A \cos Ax)(D_3 \cos By + D_4 \sin By)e^{-j\beta z} \quad (6.97)$$

We can apply the boundary conditions to E_x and E_y as follows:

(i) since E_x is tangential to the horizontal walls of the waveguide, we have $E_x = 0$ at $y = 0$ and $y = b$.

(ii) since E_y is tangential to the vertical walls of the waveguide, we have $E_y = 0$ at $x = 0$ and $x = a$.

The boundary condition (i) yields

$$D_4 = 0 \quad (6.98)$$

and $$B = \frac{n\pi}{b} \quad (6.99)$$

The boundary condition (ii) yields

$$D_2 = 0 \quad (6.100)$$

and $$A = \frac{m\pi}{a} \quad (6.101)$$

Substituting for D_2, D_4, A and B in Eqn (6.95) and combining D_1, D_5, D_3 into one constant D we get

$$H_z = D\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \quad (6.102)$$

The corresponding transverse field components can be obtained by substituting H_z in Eqns (6.59) to (6.62) as

$$E_x = \frac{+j\omega\mu}{h^2} D\left(\frac{n\pi}{b}\right)\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \quad (6.103)$$

$$E_y = \frac{-j\omega\mu}{h^2}\left(\frac{m\pi}{a}\right) D\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \quad (6.104)$$

$$H_x = \frac{+j\beta}{h^2}\left(\frac{m\pi}{a}\right) D\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \quad (6.105)$$

$$H_y = \frac{j\beta}{h^2}\left(\frac{n\pi}{b}\right) D\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \quad (6.106)$$

Again, in this case also we have

$$h^2 = A^2 + B^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 \quad (6.107)$$

and integers m and n together define the order of the mode. The mode is designated by TE_{mn} mode.

From the field expressions, following observations can be made regarding the TE modes:

1. TE fields exist in discrete field patterns called the Transverse Electric modes of the waveguide.
2. All field components have sinusoidal amplitude variation in x and y directions.
3. Indices m and n represent number of half cycles of the field amplitudes in x and y directions respectively.
4. Sinusoidal functions for electric and magnetic fields are shifted with respect to each other in space by quarter cycle.
5. Either m or n has to be non-zero for atleast some of the transverse fields to be non-zero. If both m and n are zero, although H_z appears to be non-zero, it cannot exist without any electric field. That is to say, the TE_{00} mode can not exist but TE_{m0} and TE_{0n} modes can exist. This is unlike the TM modes which needed both the indices to be non-zero. The lowest order modes for the transverse electric case would then be TE_{10} and TE_{01}.

6.4.3 Does TEM Mode Exist?

We have observed that for a parallel plane waveguide, the TM_{00} mode which is same as the TEM mode, exists. Let us now see whether the TEM mode exists inside a rectangular waveguide!

Let us suppose the TEM mode exists inside a rectangular waveguide. That means the electric and magnetic fields vectors lie in the transverse plane (the xy-plane). Since, the magnetic field lines close on themselves, the magnetic field form loops in the xy-plane. However, according to the Ampere's law, the magnetic fields must enclose a current. Since, the waveguide is hollow, there is no conduction current enclosed by the magnetic field lines. So, there should be displacement current flowing along the waveguide. But, the displacement current along the waveguide will need a longitudinal electric field which is not present for a TEM mode. Hence, we find that the magnetic field loops in the transverse plane can not exist without longitudinal electric field and consequently, the TEM does not exist inside a rectangular waveguide.

6.4.4 Cut-off Frequency of a Mode

For both TE and TM modes the phase constant is given by

$$\beta^2 = \omega^2\mu\epsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2 \quad (6.108)$$

$$\Rightarrow \quad \beta = \sqrt{\omega^2\mu\epsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} \quad (6.109)$$

For the mode to be a travelling mode, β has to be real. The frequency at which β changes from real to imaginary is called the cut-off frequency of the mode. At the cut-off frequency, therefore, $\beta = 0$, giving

$$\omega_c = \frac{1}{\sqrt{\mu\epsilon}}\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]^{1/2} \quad (6.110)$$

$$\Rightarrow \qquad f_c = \frac{1}{2\pi\sqrt{\mu\epsilon}}\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \qquad (6.111)$$

The cut-off frequency of the lowest possible TM mode (TM)is

$$f_{cTM_{11}} = \frac{1}{2\pi\sqrt{\mu\epsilon}}\sqrt{\left(\frac{\pi}{a}\right)^2 + \left(\frac{\pi}{b}\right)^2} \qquad (6.112)$$

whereas the cut-off frequencies of the lowest order TE modes are

$$f_{cTE_{10}} = \frac{1}{2\pi\sqrt{\mu\epsilon}}\frac{\pi}{a} = \frac{1}{2a\sqrt{\mu\epsilon}} \qquad (6.113)$$

$$\text{and} \qquad f_{cTE_{01}} = \frac{1}{2\pi\sqrt{\mu\epsilon}}\frac{\pi}{b} = \frac{1}{2b\sqrt{\mu\epsilon}} \qquad (6.114)$$

Since by convention $a > b$, we get

$$f_{cTE_{10}} < f_{cTE_{01}} < f_{cTM_{11}} \qquad (6.115)$$

Hence, the lowest frequency which can propagate on the waveguide is $f_{cTE_{10}}$. No energy can propagate in a rectangular waveguide at a frequency below $f_{cTE_{10}}$. This is the absolute cut-off frequency of the waveguide.

It is clear from Eqn (6.110) that as the order of the mode increases the cut off frequency also increases. It is difficult to give general order of cut off frequencies because depending upon the values of a and b, $f_{cTE_{20}}$ and/or $f_{cTE_{30}}$, could be higher or lower compared to $f_{cTE_{01}}$. Similarly, $f_{cTE_{02}}$ or f_{cTE03} could be higher or lower than $f_{cTM_{11}}$. One thing however is clear that irrespective of the dimensions of the waveguide the TE_{10} mode always has the lowest cut-off frequency. TE_{10}, therefore, is called the dominant mode of the rectangular waveguide.

The cut-off wavelength for a mode can be obtained from the cut-off frequency as

$$\lambda_c = \frac{2}{\sqrt{(\frac{m}{a})^2 + (\frac{n}{b})^2}} \qquad (6.116)$$

(Note: the intrinsic velocity of an electromagnetic wave in the medium filling the waveguide is $v = \frac{1}{\sqrt{\mu\epsilon}}$, and $\lambda = \frac{v}{f}$).

The cut-off wavelength of the dominant mode (TE_{10} mode, i.e. $m = 1, n = 0$) is

$$\lambda_{cTE_{10}} = 2a$$

$$\Rightarrow \qquad a = \frac{\lambda_{cTE_{10}}}{2} \qquad (6.117)$$

Equation (6.117) indicates that the longest wavelength which can propagate on a rectangular waveguide is equal to twice the width of the waveguide, (the broader cross-sectional dimension of the waveguide). In other words, for propagation of an electromagnetic wave of wavelength λ, the width of the waveguide has to be greater than $\frac{\lambda}{2}$.

EXAMPLE 6.7 A rectangular waveguide of 4 cm × 3 cm cross-section carries the dominant mode at 6 GHz. The maximum peak electric field measured inside the waveguide is 50 V/m. Find the expression for the electric and magnetic fields inside the waveguide and the power carried by the waveguide.

Solution:

For the waveguide $a = 0.04$ m and $b = 0.03$ m

From Eqn (6.107) we have,

$$h = \frac{\pi}{a} = \frac{\pi}{0.04}$$

$$\Rightarrow \qquad \beta = \sqrt{\omega^2\mu\epsilon - \left(\frac{\pi}{a}\right)^2} = 98.09 \text{ rad/m}$$

The fields for TE_{10} mode using Eqns (6.102) to (6.106) we get,

$$H_z = D\cos\left(\frac{\pi x}{a}\right)e^{-j\beta z}$$

$$E_x = 0$$

$$E_y = \frac{-j\omega\mu}{(\pi/a)}D(\pi/a)\sin\left(\frac{\pi x}{a}\right)e^{-j\beta z}$$

$$E_z = 0$$

$$H_x = \frac{j\beta}{\pi/a}(\pi/a)D\sin\left(\frac{\pi x}{a}\right)e^{-j\beta z}$$

$$H_y = 0$$

Now it is given that $|E| = |E_y| = \omega\mu D = 50$ V/m.

$$\Rightarrow \qquad D = \frac{50}{\omega\mu} = \frac{50}{2\pi \times 6 \times 10^9 \times 4\pi \times 10^{-7}} = 1.055 \times 10^{-3}$$

Substituting for D we get the fields inside the waveguide as

$$E_y = -j50\sin\left(\frac{\pi x}{0.04}\right)e^{-j\beta z} \text{ V/m}$$

$$H_x = j0.1035\sin\left(\frac{\pi x}{0.04}\right)e^{-j\beta z} \text{ A/m}$$

$$H_z = 1.055 \times 10^{-3}\cos\left(\frac{\pi x}{0.04}\right)e^{-j\beta z} \text{ A/m}$$

Now the power density of the mode is

$$\mathbf{P} = \frac{1}{2}Re\{\mathbf{E} \times \mathbf{H}^*\} = \frac{1}{2} \times 50 \times 0.1035 \times \sin^2\left(\frac{\pi x}{a}\right)$$

The total power carried by the mode can be obtained by integrating **P** over the waveguide cross section

The total power in the mode is

$$W = \int_{x=0}^{a}\int_{y=0}^{b} P dxdy$$

$$= 2.588b \int_0^a \sin^2\left(\frac{\pi x}{a}\right) dx$$

$$W = 2.588b \times \frac{a}{2} = 1.55 \times 10^{-3} \text{ Watts}$$

EXAMPLE 6.8 The longitudinal electric field for TM_{11} mode is given by

$$E_z = \sin(5x)\sin(8y)e^{-j\beta z}\ \text{V/m}$$

Derive remaining field components and the dispersion relation of the mode. Find the cut-off frequency of the mode and the phase and the group velocity of the mode at a frequency twice the cut-off frequency of the mode.

Solution:

The dispersion relation can be obtained by substituting E_z in the wave equation, i.e .

$$\nabla^2 E_z + \omega^2\mu\epsilon E_z = 0$$

$$\Rightarrow \quad \frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} + \omega^2\mu\epsilon = 0$$

$$-25 - 64 - \beta^2 + \omega^2\mu\epsilon = 0$$

The dispersion relation is $\quad \beta = \sqrt{\omega^2\mu\epsilon - 89}$

Also comparing the expression for β with Eqn (6.107), we get $h = \sqrt{89} = 9.434$

Since, the mode is TM, $H_z = 0$. Using Eqns (6.59) to (6.62) the other field components can be written as

$$E_x = \frac{-j\beta}{h^2}\frac{\partial E_z}{\partial x} = \frac{-j\beta}{89} 5\cos(5x)\sin(8y)e^{-j\beta z}$$

$$E_y = \frac{-j\beta}{h^2}\frac{\partial E_z}{\partial y} = \frac{-j\beta}{89} 8\sin(5x)\cos(8y)e^{-j\beta z}$$

$$H_x = \frac{j\omega\epsilon}{h^2}\frac{\partial E_z}{\partial y} = \frac{j\omega\epsilon}{89} 8\sin(5x)\cos(8y)e^{-j\beta z}$$

$$H_y = \frac{-j\omega\epsilon}{h^2}\frac{\partial E_z}{\partial x} = \frac{j\omega\epsilon}{89} 5\cos(5x)\sin(8y)e^{-j\beta z}$$

The cut off frequency of the mode is

$$\omega_c = \frac{1}{\sqrt{\mu\epsilon}}\sqrt{89}$$

$$\text{Noting that } \frac{1}{\sqrt{\mu\epsilon}} = 3 \times 10^8 \text{m/s}$$

$$\omega_c = 2.83 \times 10^9 \text{rad/s}$$

From the dispersion relation, we get

$$\text{Phase velocity} \quad v_p = \frac{\omega}{\beta} = \frac{\omega}{\sqrt{\omega^2\mu\epsilon - 89}}$$

$$\text{Group velocity} \quad v_g = \frac{\partial\omega}{\partial\beta} = \frac{1}{\mu\epsilon}\frac{\beta}{\omega}$$

At $\omega = 2\omega_c$, we get

$$v_p = 3.46 \times 10^8 \text{ m/s}$$

$$v_g = 2.598 \times 10^8 \text{ m/s}$$

EXAMPLE 6.9 The cross section of a rectangular waveguide is 20 cm × 5 cm. Find 6 lowest order modes which will propagate on the waveguide and their cut-off frequencies.

Solution:

For the waveguide $a = 0.20$ m and $b = 0.05$ m

The cut-off frequency of a mode is given by

$$\omega_c = \frac{1}{\sqrt{\mu\epsilon}}\left\{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right\}^{1/2}$$

$$\Rightarrow \quad 2\pi f_c = 3 \times 10^8 \left\{\left(\frac{m}{0.2}\right)^2 + \left(\frac{n}{0.05}\right)^2\right\}^{1/2} \pi$$

$$\Rightarrow \quad f_c = 1.5 \times 10^{10}\left\{\left(\frac{m}{20}\right)^2 + \left(\frac{n}{5}\right)^2\right\}^{1/2}$$

The cut-off frequency in ascending order will correspond to

$m = 1$	$n = 0$	$m = 4$	$n = 0$
$m = 2$	$n = 0$	$m = 0$	$n = 1$
$m = 3$	$n = 0$	$m = 1$	$n = 1$

Since, for TM modes m and n both have to be non zero we get the 6 lowest order modes as TE_{10}, TE_{20}, TE_{30}, TE_{40}, TE_{01}, TE_{11}, TM_{11}

In fact TE_{11} and TM_{11} have same cut off frequency. The cut-off frequencies of the modes are:

Mode	Cut-off frequency		
TE_{10}	0.75 GHz	TE_{20}	1.5 GHz
TE_{30}	2.25 GHz	TE_{40}	3.0 GHz
TE_{01}	3.0 GHz	TE_{11}	3.092 GHz
TM_{11}	3.092 GHz		

6.5 PARALLEL PLANE WAVEGUIDE (as a limiting case of Rectangular Waveguide)

In the previous sections, we derived the fields for the parallel plane waveguide and the rectangular waveguide using two different approaches. For the parallel plane waveguide the fields are derived using interference of plane waves, whereas for the rectangular waveguide the analysis is carried out using general mathematical approach. It is then worthwhile to compare the results obtained by the two approaches and to make sure that indeed the two results are same. For comparison of the two results, let us treat the parallel plane waveguide as a limiting case of the rectangular waveguide. For a rectangular waveguide if we make a or b equal to ∞, the waveguide becomes a parallel plane waveguide. To be consistent with the parallel plane geometry used in Section 6.1, (that is the waveguide is infinitely extended in y-direction) we appropriately choose $b \to \infty$ in the following analysis. Figure 6.13 shows a rectangular waveguide and its limiting geometry for $b \to \infty$.

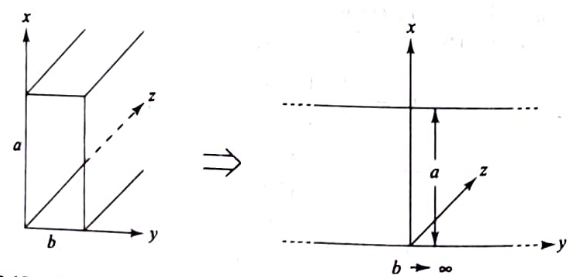


Fig. 6.13 *Parallel plane wave guide as a limiting case of rectangular waveguide with b→ ∞.*

Substituting $b = \infty$ in Eqn (6.108), we get the dispersion relation for the parallel plane waveguide as

$$\beta^2 = \omega^2 \mu\epsilon - \left(\frac{m\pi}{a}\right)^2 \tag{6.118}$$

and from Eqn (6.92)

$$h = \left(\frac{m\pi}{a}\right) \tag{6.119}$$

Equation (6.119) is identical to that derived in Section 6.1 (Eqn (6.14)).

6.5.1 TE Mode

Substituting for $b = \infty$ in Eqns (6.102) to (6.106) we get the TE fields of the rectangular waveguide in the limit $b \to \infty$ as

$$H_z = D\cos\left(\frac{m\pi x}{a}\right)e^{-j\beta z} \tag{6.120}$$

$$E_x = 0 \tag{6.121}$$

$$E_y = \frac{-j\omega\mu}{h^2}\left(\frac{m\pi}{a}\right)D\sin\left(\frac{m\pi x}{a}\right)e^{-j\beta z} \tag{6.122}$$

$$= \frac{-j\omega\mu}{\left(\frac{m\pi}{a}\right)}D\sin\left(\frac{m\pi x}{a}\right)e^{-j\beta z} \tag{6.123}$$

$$H_x = \frac{j\beta}{h^2}\left(\frac{m\pi}{a}\right)D\sin\left(\frac{m\pi x}{a}\right)e^{-j\beta z} \tag{6.124}$$

$$= \frac{j\beta}{\left(\frac{m\pi}{a}\right)}D\sin\left(\frac{m\pi x}{a}\right)e^{-j\beta z} \tag{6.125}$$

$$H_y = 0 \tag{6.126}$$

If we compare Eqns (6.10) and (6.11) with Eqns (6.120) to (6.126) we note that the two sets of field components are identical in their spatial behavior but their absolute coefficients are different. Since, there is nothing special about constant E_{i0} or D, the two can be related through a relation,

$$2jE_{i0} = \frac{-j\omega\mu}{\left(\frac{m\pi}{a}\right)}D \tag{6.127}$$

$$\Rightarrow \quad D = \frac{-2}{\omega\mu}\left(\frac{m\pi}{a}\right)E_{i0} \tag{6.128}$$

Substituting for D in Eqns (6.120) to (6.126) yields the field components identical to that in Eqns (6.10) and (6.11).

6.5.2 TM Mode

The TM fields for the parallel plane waveguide can be deduced by substituting b→ ∞ in the TM fields of the rectangular waveguide. However, there is small trick envolved in doing so. We note that if we substitute $b \to \infty$ in Eqns (6.88) to (6.91) all the field components identically go to zero since each expression either has a term $(n\pi/b)$ or $\sin(n\pi/b)$. Does this mean that the TM modes do not exist or we cannot deduce the TM fields as a limiting case for $b \to \infty$? Before we jump to any such conclusion, let us reinvestigate the fields in Eqns (6.88) to

(6.91). For $b \to \infty$, all fields go to zero provided C is a finite constant. Since, C is just a mathematical constant, it could be even infinite. The product C $(n\pi/b)$ or C $\sin(n\pi y/b)$ then does not become zero but a finite quantity. Therefore, in this case it is more appropriate to define a new constant

$$C' = C \sin\left(\frac{n\pi y}{b}\right) \tag{6.129}$$

which is finite for $b \to \infty$.

The longitudinal electric field from Eqn (6.88) therefore is

$$E_z = C' \sin\left(\frac{m\pi x}{a}\right) e^{-j\beta z} \tag{6.130}$$

and consequently the other field components from (6.89) to (6.91) (note $h = m\pi/a$ from Eqn (6.119)) are

$$E_x = \frac{-j\beta}{\left(\frac{m\pi}{a}\right)} C' \cos\left(\frac{m\pi x}{a}\right) e^{-j\beta z} \tag{6.131}$$

$$E_y = 0 \tag{6.132}$$

$$H_x = 0 \tag{6.133}$$

$$H_y = \frac{-j\omega\epsilon}{\left(\frac{m\pi}{a}\right)} C' \cos\left(\frac{m\pi x}{a}\right) e^{-j\beta z} \tag{6.134}$$

Again defining constant C' as

$$C' = \frac{-j2}{\omega\epsilon\eta_1}\left(\frac{m\pi}{a}\right) E_{i0} \tag{6.135}$$

we get the field Eqns (6.131) to (6.134) identical to Eqns (6.17) and (6.18).

The above comparison not only establishes confirmity of the two approaches but also indicates that there is nothing special about the amplitude coefficients of the field components. As long as the ratio of the amplitudes of the various field components is maintained, any field amplitude can be chosen independently and others can be expressed in terms of it. The choice of independent amplitude is decided by the data given for a particular problem. This aspect will be clear from the numerical examples given later.

6.6 VISUALIZATION OF FIELDS INSIDE A WAVEGUIDE

The three dimensional visualization of the fields inside a waveguide is not only interesting from academic viewpoint but is important for identifying regions of high fields so that the fields can be efficiently tapped by field probes. The field probes are devices which can either extract energy from a waveguide or induce fields inside a waveguide. The visualization of the modal fields appear complex as they are complex functions of space and time. However, an easier way to visualize the fields is to first freeze the time, i.e. get the fields at some instant of time. Without losing generality we can always take the instant as $t = 0$. Now

plot the fields vectorially at different locations, generally plane by plane, inside a waveguide. A good approach would be to follow a field vector till it either closes on itself or ends up on the waveguide walls. Since, the fields are repetitive over one guided wavelength λ_g $(= 2\pi/\beta)$, the field tracing needs to be carried out over one guided wavelength of the waveguide only. The total field distribution can be obtained by repeating the block of one guided wavelength over the entire length of the waveguide. Having obtained the fields at some instant, the time behavior can be incorporated by letting the field pattern move along the direction of wave propagation with a velocity equal to the modal phase velocity.

To demonstrate the approach let us visualize few specific modal patterns inside a parallel plane or rectangular waveguide.

6.6.1 Fields for TM₁ Mode in a Parallel Plane Waveguide

Substituting $m = 1$ in Eqns (6.17) and (6.18) and noting that $j = e^{\frac{j\pi}{2}}$, and $\beta = 2\pi/\lambda_g$, the fields for TM_1 mode can be written as

$$E_z = Re\left[2E_{i0}\left(\frac{\pi}{\beta_1 a}\right)\sin\left(\frac{\pi x}{a}\right)e^{(-j\beta z + j\frac{\pi}{2})}\right]$$
$$= 2\frac{E_{i0}\pi}{\beta_1 a}\sin\left(\frac{\pi x}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g} - \frac{\pi}{2}\right) \tag{6.136}$$

$$E_x = Re\left[2E_{i0}\frac{\beta}{\beta_1}\cos\left(\frac{\pi x}{a}\right)e^{-j\beta z}\right]$$
$$= 2\frac{E_{i0}\beta}{\beta_1}\cos\left(\frac{\pi x}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.137}$$

$$H_y = Re\left[2\frac{E_{i0}}{\eta_1}\cos\left(\frac{\pi x}{a}\right)e^{-j\beta z}\right]$$
$$= \frac{2E_{i0}}{\eta_1}\cos\left(\frac{\pi x}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.138}$$

Other field components E_y, H_x, H_z are zero.

The first thing which can be noted from Eqns (6.136) to (6.138) is that the fields are constant as a function of y, i.e. there is no field variation along the y-direction. The fields therefore have to be traced in the xz plane only. Also as mentioned above, the fields need to be traced only over a length λ_g in the z direction, say from $z = 0$ to $z = \lambda_g$. Tracing of magnetic field in this case is rather straight forward as it has only one component.

The magnetic field has maximum +ve amplitude at $x = 0$, $z = 0$, and $x = a$, $z = \lambda_g/2$, and maximum negative amplitude at $x = a$, $z = 0$ and $x = 0$, $z = \lambda_g/2$. For no specific reason if we take +ve field means field oriented along $+y$ direction, the field at $x = 0$, $z = 0$ will point outwards from the plane of the paper and the field at $x = a$, $z = 0$ will point inwards the plane of the paper. The fields are, therefore, appropriately marked by ⊗ and ⊙ in Fig. 6.14. The radii of the

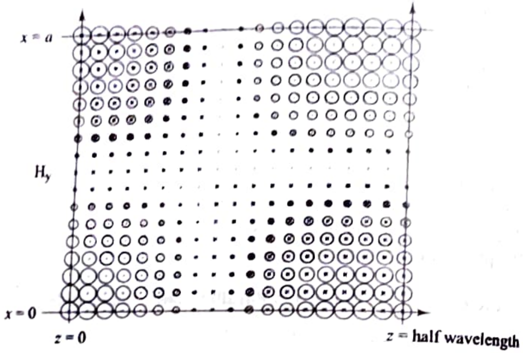


Fig. 6.14 *Magnetic field for a TM_1 mode inside a parallel plane waveguide in the xz-plane.*

circles indicate the strength of the field. The larger the circle higher is the field amplitude. The magnetic field can, therefore, be imagined like rods of various diameters placed perpendicular to the plane of the paper (the y-axis).

For tracing the electric field we can make following observations:

(i) E_x and E_z components being cos and sin functions respectively are in space quardrature along the x-direction, i.e. where E_x is maximum, E_z is zero and vice versa. E_x is maximum at $x = 0$ and $x = a$, and it is 0 at $x = a/2$, whereas, E_z is 0 at $x = 0$ and $x = a$, and is maximum at $x = a/2$.

(ii) E_x and E_z are staggered in z-direction by a distance $\lambda_g/4$. E_x amplitude goes to maximum at $z = 0$ and $z = \lambda_g/2$, whereas, E_z amplitude is maximum at $z = \lambda_g/4$, and $z = 3\lambda_g/4$.

(iii) In general the electric field amplitude and its direction is given by

$$|\mathbf{E}| = |E_x|^2 + |E_z|^2 \tag{6.139}$$

$$= \left|\frac{2E_{i0}\beta}{\beta_1}\cos\left(\frac{\pi x}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g}\right)\right|^2 + \left|\frac{2E_{i0}\pi}{\beta_1 a}\sin\left(\frac{\pi x}{a}\right)\sin\left(\frac{2\pi z}{\lambda_g}\right)\right|^2 \tag{6.140}$$

and

$$\tan\theta = \frac{E_x}{E_z} = \frac{\beta a}{\pi}\cot\left(\frac{\pi x}{a}\right)\cot\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.141}$$

where θ is the angle measured from $+z$ axis towards $+x$ direction. After plotting the field vectors at few locations, and stretching our imagination we get the electric field lines as shown in Fig. 6.15. Superposition of Fig. 6.14 and Fig. 6.15 gives the complete picture of the fields inside a parallel plane waveguide for TM_1 mode. The fields as a block drift along $+z$ direction with the phase velocity

$$v_p = \frac{\omega}{\beta} = \frac{2\pi f}{\frac{2\pi}{\lambda_g}} = f\lambda_g \tag{6.142}$$

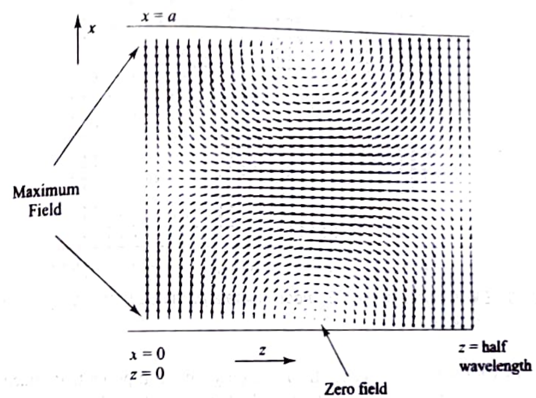


Fig. 6.15 *Electric field for the TM_1 mode inside a parallel plane waveguide in the xz-plane.*

6.6.2 Fields for TE_2 Mode in a Parallel Plane Waveguide

Substituting $m = 2$ in Eqns (6.10) and (6.11) the field components for the TE_2 mode can be obtained as

$$E_y = 2jE_{i0}\sin\left(\frac{2\pi x}{a}\right)e^{-j\beta z} \tag{6.143}$$

$$H_x = 2j\frac{E_{i0}}{\eta_1}\left(\frac{2\pi}{\beta_1 a}\right)\sin\left(\frac{2\pi x}{a}\right)e^{-j\beta z} \tag{6.144}$$

$$H_z = \frac{2E_{i0}\beta}{\eta_1\beta_1}\cos\left(\frac{2\pi x}{a}\right)e^{-j\beta z} \tag{6.145}$$

Taking real part of the field components yields

$$E_y = 2E_{i0}\sin\left(\frac{2\pi x}{a}\right)Re(e^{-j\beta z + j\frac{\pi}{2}})$$

$$= 2E_{i0}\sin\left(\frac{2\pi x}{a}\right)\cos\left(\frac{2\pi}{\lambda_g}(z - \frac{\lambda_g}{4})\right) \tag{6.146}$$

$$H_x = \frac{2E_{i0}}{\eta_1}\left(\frac{2\pi}{\beta_1 a}\right)\sin\left(\frac{2\pi x}{a}\right)\cos\left(\frac{2\pi}{\lambda_g}(z - \frac{\lambda_g}{4})\right) \tag{6.147}$$

$$H_z = \frac{2E_{i0}}{\eta_1}\frac{\beta}{\beta_1}\cos\left(\frac{2\pi x}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.148}$$

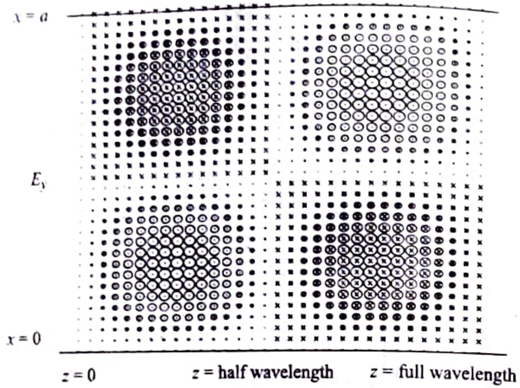


Fig. 6.16 *Electric field for the TE_2 mode inside a parallel plane waveguide in the xz-plane.*

In this case, since the electric field has only one component, its tracing is straight forward. Assuming E_{i0} to be +ve, the E-field will be along $+y$ direction between $z = 0$ and $z = \lambda_g/2$, and along $-y$ direction between $z = \lambda_g/2$ and $z = \lambda_g$. The field lines are like infinitely long rods perpendicular to the plane of the paper as shown in Fig 6.16. Since the field component is parallel to the waveguide planes, as expected, it is zero on the two planes.

The magnetic field has magnitude and direction as

$$|\mathbf{H}| = |H_x|^2 + |H_z|^2 \tag{6.149}$$

$$\tan\theta = \frac{H_x}{H_z} = \frac{2\pi}{a\beta}\tan\left(\frac{2\pi x}{a}\right)\frac{\cos\left(\frac{2\pi}{\lambda_g}(z - \frac{\lambda_g}{4})\right)}{\cos(\frac{2\pi}{\lambda_g}z)} \tag{6.150}$$

Since, there is no y-component of the H-field, the field lines are in the xz-plane only. The field lines are horizontal (along z direction) for $x = 0$, $x = a/2$ and $x = a$, and $z = 0$, $z = \lambda_g/2$ and $z = \lambda_g$. The field lines become vertical at $x = a/4$, $x = 3a/4$ and $z = \lambda_g/4$ and $z = 3\lambda_g/4$.

The field can be roughly drawn at these locations. Stretching the imagination a bit we can form the field loops as shown in Fig. 6.17. The magnetic field lines are like infinitely long rolled carpets stacked next to each other. Since, the mode is TM_2 mode, there are two rolls one above the other.

The tangential magnetic field induces surface current on the waveguide walls and the current flows in the y-direction ($\hat{\mathbf{z}} \times \hat{\mathbf{x}}$). It is interesting then to note that there is no current flow in the direction of the wave propagation, i.e. direction of

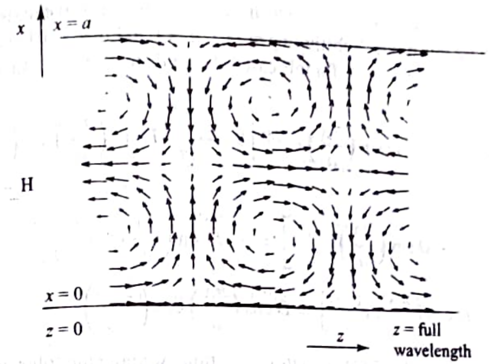


Fig. 6.17 *Magnetic field for the TE_2 mode inside a parallel plane waveguide in the xz-plane.*

the energy flow (the z-direction). The total fields (electric and magnetic together) can be visualized as shown in Fig. 6.18.

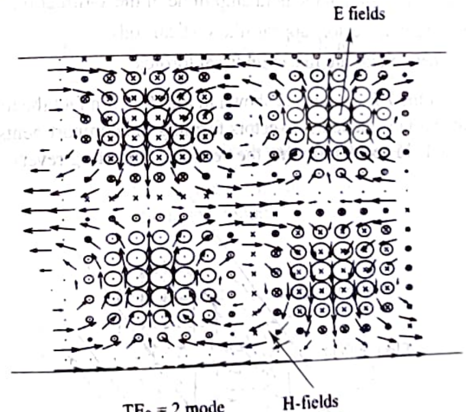


Fig. 6.18 *Electric and magnetic fields for TE_2 mode inside a parallel plane waveguide.*

6.6.3 Fields for TE_{10} Mode in a Rectangular Waveguide

The visualization of three dimensional fields in a rectangular waveguide is more complex compared to that in a parallel plane waveguide as the fields are confined in a closed cross-sectional area. However, if we systematically follow the steps outlined above, the visualisation of the fields is quite straight forward. TE_{10} mode has been chosen here as an example because this is the dominant mode of a rectangular waveguide and any one who handles rectangular waveguide has to be conversant with this mode. Substituting $m = 1, n = 0$ in Eqns (6.102) to (6.106) (and also noting that $h = \pi/a$ for this case), the fields for the TE_{10} mode can be obtained as

$$E_y = Re\left[-j\frac{\omega\mu a}{\pi}D\sin\left(\frac{\pi x}{a}\right)e^{-j\beta z}\right] = -\frac{\omega\mu a}{\pi}D\sin\left(\frac{\pi x}{a}\right)\sin\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.151}$$

$$H_x = Re\left[\frac{j\beta a}{\pi}D\sin\left(\frac{\pi x}{a}\right)e^{-j\beta z}\right] = \frac{\beta a}{\pi}D\sin\left(\frac{\pi x}{a}\right)\sin\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.152}$$

$$H_z = Re\left[D\cos\left(\frac{\pi x}{a}\right)e^{-j\beta z}\right] = D\cos\left(\frac{\pi x}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.153}$$

Freezing the fields in space at an instant of time, we note the following:

1. The electric field has only y component. The amplitude of the electric field has one half cycle variation in the x-direction, and a sinusoidal variation in the z-direction.
2. The magnetic field forms loops in the xz-plane as shown in Fig. 6.19.
3. There is no variation in the field amplitude in the y-direction.
4. The electric field vectors appear like vertical rods.
5. The magnetic fields are like cylindrical toroids.

The fields for the TE_{10} mode are shown in Fig. 6.19. In fact the total field will be like a train with magnetic field vectors forming the compartments of the train and the electric field vectors forming the vertical connecting revets.

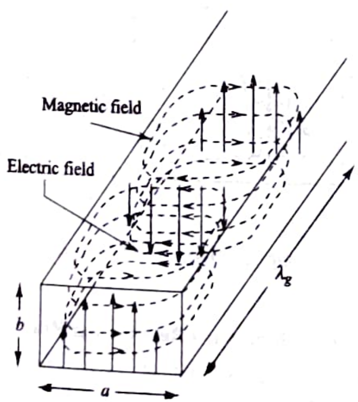


Fig. 6.19 *Modal fields for TE_{10} mode. Thick lines indicate the electric field and the dotted lines indicate the magnetic field.*

6.7 SURFACE CURRENT ON THE WAVEGUIDE WALLS

Let us assume that the waveguide walls are made of highly conducting material. The modal fields inside a waveguide then induce surface charges and surface currents on the waveguide walls. In other words, we may say that the modal fields are supported by the surface charges and currents on the inner walls of the waveguide. Since, the fields are time varying, the surface charges and currents also vary as a function of time.

The power flow inside the waveguide is normally related to these times varying charges and currents. The students, therefore, get a wrong impression that the power is carried by the charges and the currents and hence the charges and currents should flow in the direction of the power flow. It should be emphasized that the power inside the waveguide or for that matter along any guided structure, is carried by the fields and not by the charges. It is then possible that the currents in waveguide walls may flow in any direction along the surface of the walls whereas the power flows along the length of the waveguide only. The surface current distributions for TE_1 mode (Section 6.7.1) and TE_{10} mode (Section 6.7.2) clearly demonstrate this point.

The approach to finding surface current distribution is similar to that used in visualising the modal fields inside a waveguide. First we freeze time to get instantaneous surface current distribution along the waveguide walls, then we allow the distribution to drift along the waveguide with the phase velocity. Without losing generality, we, therefore, first obtain spatial surface current distribution at $t = 0$.

6.7.1 Surface Currents in a Parallel Plane Waveguide for TE_1 Mode

The surface current on a conducting boundary is related to the tangential component of the magnetic field at the conducting surface.

$$\mathbf{J_S} = \hat{\mathbf{n}} \times \mathbf{H} \tag{6.154}$$

where $\mathbf{H}$ is the magnetic field at the conducting surface and $\hat{\mathbf{n}}$ is the outward normal at the surface.

From Eqn (6.11) the magnetic field at the waveguide plates is

$$\mathbf{H} = \frac{2E_{i0}}{\eta_1}\frac{\beta}{\beta_1}e^{-j\beta z}\hat{\mathbf{z}} \qquad \text{at} \qquad x = 0 \text{ wall} \tag{6.155}$$

$$= \frac{-2E_{i0}}{\eta_1}\frac{\beta}{\beta_1}e^{-j\beta z}\hat{\mathbf{z}} \qquad \text{at} \qquad x = d \text{ wall} \tag{6.156}$$

At $x = 0$ wall, $\mathbf{n}$ is oriented in $+x$ direction and at $x = d$ wall, $\hat{\mathbf{n}}$ is oriented in $-x$ direction (Fig. 6.20). The surface current along $x = 0$ wall flows along $\hat{\mathbf{n}} \times \mathbf{H} \rightarrow \hat{\mathbf{x}} \times \hat{\mathbf{z}} = -\hat{\mathbf{y}}$ direction, and along $x = d$ wall it flows along $-\hat{\mathbf{x}} \times (-\hat{\mathbf{z}}) = -\hat{\mathbf{y}}$ direction.

Now, taking the real part of $\mathbf{H}$ we get at $t = 0$ time.

$$\mathbf{H} = \frac{2E_{i0}}{\eta_1}\frac{\beta}{\beta_1}\cos(\beta z)\hat{\mathbf{z}} \qquad x = 0 \text{ wall} \tag{6.157}$$

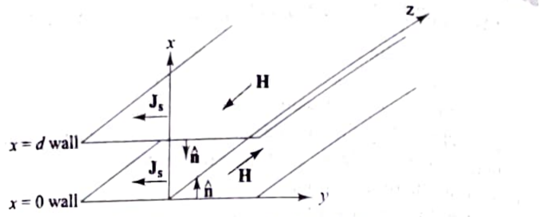


Fig. 6.20 *Surface current direction on the walls of a parallel plane waveguide for TE_1 mode.*

$$= \frac{-2E_{i0}}{\eta}\frac{\beta}{\beta_1}\cos(\beta z)\hat{\mathbf{z}} \qquad x = d \text{ wall} \tag{6.158}$$

Since, $\beta = 2\pi/\lambda_g$, where λ_g is the guided wavelength of the mode, the magnetic field and consequently the surface current is periodic in the z-direction. The current is maximum at $z = 0, \pi, 2\pi, 3\pi,$ and zero at $z = \lambda_g/2, 3\lambda_g/2, 5\lambda_g/2\ldots$ and so on. The current therefore is y-oriented with no amplitude variation along y-direction. The current sinusoidally varies in amplitude along the z-direction. The surface current distribution then looks like as shown in Fig. 6.21

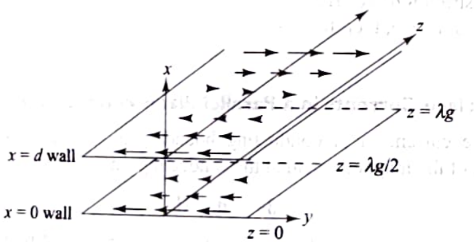


Fig. 6.21 *Surface current on the walls of a parallel plane waveguide for TE_1 mode. Length of the arrow indicates the field amplitude.*

It can be noted that the currrent is oriented in y-direction at all times. That is, the current is transverse to the direction of power flow (z-direction). There is no component of the current in the direction of the power flow.

6.7.2 Surface Currents on the Walls of a Rectangular Waveguide for TE_{10} Mode

Let us now obtain the surface current distribution for the domimnant mode of a rectangular waveguide (TE_{10} mode).

The magnetic field for TE_{10} mode from Eqns (6.152) and (6.153) can be written as

$$H_x = \frac{\beta a}{\pi} D \sin\left(\frac{\pi x}{a}\right)\sin\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.159}$$

$$H_z = D\cos\left(\frac{\pi x}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.160}$$

Since the surface current is related to the tangential component of the magnetic field at the wall, we should look at only that magnetic field component which is tangential to a particular wall.

For vertical walls the tangential components is only H_z, whereas for the horizontal walls the both H_x and H_z components are tangential.

For vertical walls we have

$$H_z = D\cos\left(\frac{2\pi z}{\lambda_g}\right) \qquad x = 0 \tag{6.161}$$

$$= -D\cos\left(\frac{2\pi z}{\lambda_g}\right) \qquad x = a \tag{6.162}$$

The outward unit normal $\hat{\mathbf{n}}$ is along $\hat{\mathbf{x}}$ on $x = 0$ wall, wheras it is along $-\hat{\mathbf{x}}$ on $x = a$ wall. The current direction on $x = 0$ wall, therefore, is $\hat{\mathbf{n}} \times \mathbf{H} \to \hat{\mathbf{x}} \times \hat{\mathbf{z}} = -\hat{\mathbf{y}}$. Similarly for $x = a$ wall, the current direction is $\hat{\mathbf{n}} \times \mathbf{H} \to \hat{\mathbf{x}} \times \hat{\mathbf{z}} = -\hat{\mathbf{y}}$. Now, since H_z is independent of y, the current magnitude is constant along the height of the vertical walls.

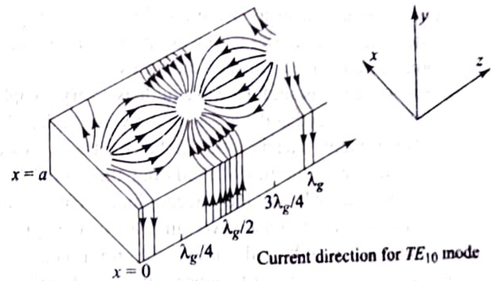


Fig. 6.22 *Surface current on the inner walls of a rectangular waveguide for the TE_{10} mode.*

On the horizontal walls the surface current is due to both H_x and H_z. At $y = 0$ wall, the outward unit normal $\hat{\mathbf{n}} = \hat{\mathbf{y}}$ and at $y = b$, $\hat{\mathbf{n}} = -\hat{\mathbf{y}}$. The vectorial current distribution at lower wall $y = 0$, can be written as

$$\mathbf{J}_s = \hat{\mathbf{n}} \times \mathbf{H} = \mathbf{y} \times (H_x\hat{\mathbf{x}} + H_z\hat{\mathbf{z}} = -H_x\hat{\mathbf{z}} + H_z\hat{\mathbf{x}} \tag{6.163}$$

$$\Rightarrow \quad J_{sx} = H_z = D\cos\left(\frac{\pi x}{a}\right)\cos\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.164}$$

$$J_{sz} = -H_x = -\frac{\beta a}{\pi}\sin\left(\frac{\pi x}{a}\right)\sin\left(\frac{2\pi z}{\lambda_g}\right) \tag{6.165}$$

On $y = b$ wall the direction of $\hat{\mathbf{n}}$ is reversed, i.e. along $-\hat{\mathbf{y}}$ direction but the magnetic field directions is identical to that on the lower wall. The surface current distribution on the upper ($y = b$) wall therefore is negative of the current distribution on the lower ($y = 0$) wall.

The current distribution is shown in Fig. 6.22. The current flows from the center of the upper wall to the center of lower wall and vice versa. The current also reverses sign every $\lambda_g/2$ distance along the waveguide. The current distribution in Fig. 6.22 drifts along the z-direction with the phase velocity.

From the above two examples, it is clear that as such the direction of surface current has nothing to do with the power flow inside the waveguide. In the first case the current distribution is very simple and transverse, whereas, in the second case the distribution is quite complex.

6.8 ATTENUATION IN A WAVEGUIDE

In the analysis of waveguides, it is assumed that the waveguide is ideal, that is, the walls of the waveguide are made of ideal conductor and the medium filling the waveguide is an ideal dielectric. Since, neither ideal conductor nor ideal dielectric absorb any power, the energy flow inside the waveguide is without any attenuation. However, this picture is an idealistic picture and in practice we neither find perfect conductors nor perfect dielectrics. Consequently, the electromagnetic wave attenuates as it travels along the waveguide and the field amplitude systematically decreases with wave propagation. The attenuation effects get more pronounced as the frequency increases.

Calculation of attenuation constant for a waveguide is a very complex problem in general. This is due to the fact that the losses in the waveguide, which depend upon the modal field distribution, alter the modal fields. In other words, calculation of losses require knowledge of the modal fields but the modal fields depend upon the losses. For an arbitrary loss in the waveguide, the calculation has to be carried out in some iterative fashion. However, in most of the applications it is not necessary since in practice, to keep the attenuation minimal, waveguides are made of highly conducting walls (made of copper, silver, gold) and are filled with low-loss dielectric material. In this situation one can assume that the distortion of the modal fields due to losses, is negligibly small and the modal fields are almost same as that of the loss-less waveguide (ideal conductor and ideal dielectric). With the knowledge of the modal fields we can then calculate the losses due to finite conductivity of the waveguide.

The attenuation constant α of a waveguide has two components, one due to losses in dielectric, α_d, and other due to losses in the conducting walls, α_c.

$$\alpha = \alpha_d + \alpha_c \tag{6.166}$$

Since, α_d and α_c are small, we calculate one at a time assuming the other to be zero. That is, while calculating dielectric loss we assume that the waveguide walls are made of perfect conductor, and while calculating the conductor loss we assume that the dielectric is lossless.

6.8.1 Attenuation due to Dielectric Loss

The dielectric loss is calculated using the concept of complex dielectric constant for the lossy material. From the modal dispersion relation we get

$$\beta_l^2 = \omega^2\mu\epsilon_l - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2 = \omega^2\mu\epsilon_l - h^2 \tag{6.167}$$

$$= \omega^2\mu\epsilon_0\epsilon_{rl} - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2 = \omega^2\mu\epsilon_0\epsilon_{rl} - h^2 \tag{6.168}$$

Suffix 'l' is used to indicated that the dielectric is lossy. Now, if the dielectric filling the waveguide has the loss tangent $\tan\delta$, the dielectric constant for the lossy medium is

$$\epsilon_{rl} = \epsilon_r(1 - j\tan\delta) \tag{6.169}$$

where ϵ_r is the dielectric constant of the loss-less medium. (Recall that the loss tangent is defined as $\tan\delta = \frac{\sigma}{\omega\epsilon_0\epsilon_r}$, σ is the conductivity of the medium)
Substituting Eqn (6.169) in Eqn (6.168) we get

$$\beta_l = \sqrt{\omega^2\mu\epsilon_0\epsilon_r(1 - j\tan\delta) - h^2}$$

$$= \left[(\omega^2\mu\epsilon_0\epsilon_r - h^2) - j\omega^2\mu\epsilon_0\epsilon_r\tan\delta\right]^{\frac{1}{2}} \tag{6.170}$$

Noting that, $\omega^2\mu\epsilon_0\epsilon_r - h^2 = \beta^2$, where β is the phase constant of the mode in the loss-less waveguide, we get

$$\beta_l = (\beta^2 - j\omega^2\mu\epsilon_0\epsilon_r\tan\delta)^{\frac{1}{2}} \tag{6.171}$$

Since, $\tan\delta << 1$, we can approximate Eqn (6.171) (except for very close to the mode cut-off, when $\beta \to 0$) to get

$$\beta_l \simeq \beta\left(1 - \frac{j\omega^2\mu_0\epsilon_0\tan\delta}{2\beta^2}\right)$$

$$\simeq \beta - \frac{j\omega^2\mu\epsilon_0\tan\delta}{2\beta} \tag{6.172}$$

The imaginary part of β_l represents the attenuation constant of the mode. The attenuation constant due to dielectric loss, therefore, is given as

$$\alpha_d = \frac{\omega^2\mu\epsilon_0\epsilon_r\tan\delta}{2\beta}$$

$$= \frac{\omega\mu\sigma}{2\beta} \qquad \text{since } \tan\delta = \frac{\sigma}{\omega\epsilon_0\epsilon_r} \tag{6.173}$$

Using expression for β, we get

$$\alpha_d = \frac{\sigma\eta_d}{2\sqrt{1-(f_c/f)^2}} \tag{6.174}$$

where η_d ($\equiv \sqrt{\mu/\epsilon_0\epsilon_r}$) is the intrinsic impedance of the dielectric medium, and f_c is the cut off frequency of the mode. For the modes operating far from cut off, $f_c/f << 1$ and $\alpha_d \rightarrow \sigma\eta_d/2$.

6.8.2 Attenuation due to Imperfect Conducting Walls

The attenuation constant α_c is determined using first principles, i.e. using the basic definition of the attenuation constant. In the presence of loss, the **E** and **H** fields attenuate exponentially as

$$(\mathbf{E}, \mathbf{H}) = (\mathbf{E_0}, \mathbf{H_0})e^{-\alpha_c z} \tag{6.175}$$

where suffix '0' indicates quantities at $z = 0$. The power flow in the waveguide, W, which is proportional to the Poynting vector $\mathbf{P}(= 1/2Re(\mathbf{E} \times \mathbf{H}^*))$, varies as

$$W = \frac{1}{2}Re(\mathbf{E_0}e^{-\alpha_c z} \times \mathbf{H_0^*}e^{-\alpha_c z}) = \frac{1}{2}Re(\mathbf{E_0} \times \mathbf{H_0^*})e^{-2\alpha_c z} \tag{6.176}$$

$$= W_0 e^{-2\alpha_c z} \tag{6.177}$$

Differentiating Eqn (6.177) with respect to z yields

$$\frac{dW}{dz} = -2\alpha_c W_0 e^{-2\alpha_c z} = -2\alpha_c W \tag{6.178}$$

$$\Rightarrow \qquad \alpha_c = \frac{(\frac{-dW}{dz})}{2W} \tag{6.179}$$

The quantity $-(\frac{dW}{dz})$ represents the loss per unit length of the waveguide and W represents the total power flow inside the waveguide. Therefore, for determining α_c we need two quantities:

(i) Power loss in the walls of the waveguide per unit length of the waveguide.
(ii) Total power flow inside the waveguide.

Of course, in all these calculations we are assuming that the losses in the walls are small and hence the modal fields are same as that of the lossless waveguide.

Determination of α_c for general TM or TE mode is algebraically cumbersome and therefore to bring out the concepts clearly we derive α_c for two commonly encountered modes namely TM_0 (i.e. TEM) mode in a parallel plane waveguide, and TE_{10} mode in a rectangular waveguide. Once the approach is clear, interested readers may derive the expression for α_c for any desired mode.

1. α_c for TEM Mode in a Parallel Plane Waveguide

From Section 6.2 we know that the fields for the TEM mode are

$$E_x = 2E_{i0}e^{-j\beta_1 z} \equiv E_0 e^{-j\beta_1 z} \tag{6.180}$$

$$H_y = \frac{2E_{i0}}{\eta_1}e^{-j\beta_1 z} \equiv \frac{E_0}{\eta_1}e^{-j\beta_1 z} \tag{6.181}$$

where E_0 is the peak amplitude of the modal electric field. Since, the waveguide is infinitely wide, it is appropriate to calculate the attenuation constant for unit width of the waveguide (say $y = 0$ to 1).

The power carried per unit width of the waveguide can be obtained by integrating the Poynting vector over the waveguide cross section as

$$W = \int_{x=0}^{a}\int_{y=0}^{1}\frac{1}{2}Re(\mathbf{E} \times \mathbf{H}^*)_z dx dy \tag{6.182}$$

$$= \int_{x=0}^{a}\int_{y=0}^{1}\frac{1}{2}Re(E_x H_y^*) dx dy \tag{6.183}$$

$$= \int_{x=0}^{a}\int_{y=0}^{1}\frac{1}{2}\frac{|E_0|^2}{\eta_1} dx dy = \frac{1}{2}\frac{|E_0|^2}{\eta_1}a \tag{6.184}$$

The surface current in the waveguide walls

$$\mathbf{J_s} = \hat{\mathbf{n}} \times \mathbf{H} = \hat{\mathbf{x}} \times H_y\hat{\mathbf{y}} \tag{6.185}$$

$$\Rightarrow \qquad |\mathbf{J_s}| = |H_y| \tag{6.186}$$

The power loss per unit length per unit width, of the waveguide is given by

$$-\left(\frac{dW}{dz}\right) = 2\int_{y=0}^{1}\int_{z=0}^{1}\left(\frac{1}{2}R_s|J_s|^2\right)dy dz \tag{6.187}$$

where factor '2' is used since there are two walls for the waveguide.

Substituting for $|J_s|$ from Eqn (6.186) yields

$$-\frac{dW}{dz} = \int_{y=0}^{1}\int_{z=0}^{1}R_s\frac{|E_0|^2}{\eta_1^2}dy dz = R_s\frac{|E_0|^2}{\eta_1^2} \tag{6.188}$$

The attenuation constant $|\alpha_c|$ for the TEM mode is

$$\alpha_c = \frac{1}{2}\frac{R_s\frac{|E_0|^2}{\eta_1^2}}{\frac{1}{2}\frac{|E_0|^2}{\eta_1}a} = \frac{R_s}{\eta_1 a} \tag{6.189}$$

Since, $R_s = \sqrt{\omega\mu/2\sigma}$ and $\eta_1 = \sqrt{\mu/\epsilon}$, we get

$$\alpha_c = \frac{\sqrt{\omega\mu/2\sigma}}{\sqrt{\mu/\epsilon}a} = \frac{1}{a}\sqrt{\frac{\omega\epsilon}{2\sigma}} \text{ nepers/m.} \tag{6.190}$$

Note that the attenuation constant α_c, increases with frequency. That is, the waveguide becomes more lossy as the frequency increases.

2. α_c for TE_{10} in a Rectangular Waveguide

From Eqns (6.151) to (6.153) the fields for the TE_{10} mode are

$$E_y = \frac{-j\omega\mu a}{\pi}D\sin\left(\frac{\pi x}{a}\right)e^{-j\beta z} \tag{6.191}$$

$$H_x = \frac{j\beta a}{\pi}D\sin\left(\frac{\pi x}{a}\right)e^{-j\beta z} \tag{6.192}$$

$$H_z = D\cos\left(\frac{\pi x}{a}\right)e^{-j\beta z} \tag{6.193}$$

The total power flow through the waveguide is

$$W = \int_{x=0}^{a}\int_{y=0}^{b}\frac{1}{2}Re(\mathbf{E}\times\mathbf{H}^*)_z dxdy$$
$$= -\int_{x=0}^{a}\int_{y=0}^{b}\frac{1}{2}Re\left\{\left[\frac{j\omega\mu a}{\pi}D\sin\left(\frac{\pi x}{a}\right)e^{-j\beta z}\right]\left[\frac{j\beta a}{\pi}D\sin\left(\frac{\pi x}{a}\right)e^{-j\beta z}\right]^*\right\} dxdy \quad (6.194)$$

(the -ve sign is used since $E_y\hat{\mathbf{y}}\times H_x\hat{\mathbf{x}} = -\hat{\mathbf{z}}E_yH_x$)

$$W = \int_{x=0}^{a}\int_{y=0}^{b}\frac{\omega\mu\beta a^2}{2\pi^2}|D|^2\sin^2\left(\frac{\pi x}{a}\right)dxdy$$
$$= \frac{\omega\mu\beta a^2 b|D|^2}{2\pi^2}\int_{x=0}^{a}\sin^2\left(\frac{\pi x}{a}\right)dx$$
$$W = \frac{\omega\mu\beta a^3 b|D|^2}{4\pi^2} \quad (6.195)$$

To determine the power loss in the waveguide walls we need surface current on the four walls of the waveguide. One can easily note that the surface current on the vertical wall has only y component whereas on the horizontal walls (*i.e.* $y = 0$, $y = b$) the surface current has x and z components. The magnitude of the surface current on the vertical walls is

$$|\mathbf{J}_s(x=0)| = |\mathbf{J}_s(x=a)| = |H_z(x=0)| = |D| \quad (6.196)$$

and consequently the loss per unit length of vertical walls is

$$W_{L_{ver}} = 2\int_{y=0}^{b}\int_{z=0}^{1}\frac{1}{2}R_s|D|^2dydz$$
$$= R_s|D|^2b \quad (6.197)$$

The magnitude of the surface current on horizontal walls

$$|\mathbf{J}_s(y=0)| = |\mathbf{J}_s(y=b)| = |\mathbf{H}(y=0)| \quad (6.198)$$
$$\Rightarrow \quad |\mathbf{J}_s(y=0)|^2 = |\mathbf{J}_s(y=b)|^2 = |\mathbf{H}(y=0)|^2$$
$$= H_x^2(y=0) + H_z^2(y=0) \quad (6.199)$$

The loss per unit length of the horizontal walls is

$$W_{L_{hor}} = 2\int_{x=0}^{a}\int_{z=0}^{1}\frac{1}{2}R_s\left(|H_x(y=0)|^2 + |H_z(y=0)|^2\right)dxdz$$
$$= R_s\int_{x=0}^{a}\left[\left(\frac{\beta a}{\pi}\right)^2|D|^2\sin^2\left(\frac{\pi x}{a}\right) + |D|^2\cos^2\left(\frac{\pi x}{a}\right)\right]dx \quad (6.200)$$
$$= \frac{R_s|D|^2a}{2}\left[\left(\frac{\beta a}{\pi}\right)^2 + 1\right] \quad (6.201)$$

which can be further simplified by noting that $(\frac{\beta a}{\pi})^2 = (\frac{f}{f_c})^2 - 1$.
The total loss per unit length of the waveguide is

$$W_L = W_{L_{ver}} + W_{L_{hor}} \quad (6.202)$$
$$= R_s|D|^2b + \frac{R_s|D|^2a}{2}\left[\left(\frac{\beta a}{\pi}\right)^2 + 1\right]$$
$$W_L = R_s|D|^2\left[b + \frac{a}{2}\left(\frac{f}{f_c}\right)^2\right] \quad (6.203)$$

The attenuation constant for the TE_{10} mode can finally be written as

$$\alpha_c = \frac{R_s|D|^2\left[b + \frac{a}{2}(\frac{f}{f_c})^2\right]}{\frac{2\omega\mu\beta a^3 b|D|^2}{4\pi^2}} = \frac{R_s\left[1 + \frac{2b}{a}(\frac{f_c}{f})^2\right]}{\eta b\sqrt{1-(\frac{f_c}{f})^2}} \quad (6.204)$$

As can be seen, the variation of α_c as a function of frequency is quite complicated. However, it can be shown that for a given waveguide, there is a frequency f_{min} above the cut-off frequency where α_c goes minimum. α_c is infinite at the cut off frequency and $\alpha_c \to R_s/\eta b$ as $f \to \infty$. One may then argue that since $\alpha_{c_{min}}$ has to be less than $R_s/\eta b$, $\alpha_{c_{min}}$ can be decreased by increasing the height b of the waveguide. This, however cannot be done arbitrarily since for TE_{10} mode to be dominant, a has to be greater than b. Also, as b increases the cut off frequencies of TE_{01} and TM_{11} decrease and the frequency range over which the waveguide has only dominant mode propagation, decreases. In practice, therefore, the waveguides have b close to $a/2$.

EXAMPLE 6.10 There is a rectangular hole of 5 cm diameter through which a rectangular waveguide carrying dominant mode at 10 GHz is to be passed. What should be the waveguide dimension so that maximum power is transmitted through the waveguide and the peak electric field does not exceed 1 kV/m?

Solution:

For the waveguide of dimension $a \times b$, we have

$$a^2 + b^2 = d^2 = (0.05)^2 = 25\times 10^{-4}$$

where 'd' is the diameter of the hole. (see Fig. 6.23)
For TE_{10} mode the electric field is given by Eqn (6.191). The amplitude of E should be less than 1 KV/m, i.e.

$$\frac{\omega\mu a}{\pi}D \le 1\ \text{kV/m}$$
$$\Rightarrow \quad D \le \frac{10^3\pi}{\omega\mu a}\ \text{V/m}$$

The power carried by the waveguide (from Eqn (6.195)) is

$$W = \frac{\omega\mu\beta a^3 b|D|^2}{4\pi^2}$$

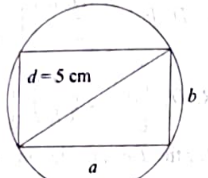


Fig. 6.23 *Waveguide fitted in a circular hole.*

Susbstituting for β for TE_{10} from Eqn (6.109) and b and D from previous equations we get,

$$W = \frac{\omega\mu\sqrt{\omega^2\mu\epsilon - (\pi/a)^2}\cdot a^3\sqrt{d^2-a^2}}{4\pi^2}\,\frac{10^6\pi^2}{\omega^2\mu^2a^2}$$

$$= \frac{10^6}{4\omega\mu}a\sqrt{\omega^2\mu\epsilon-(\pi/a)^2}\sqrt{d^2-a^2}$$

For maximum power $\partial W/\partial a = 0$, i.e.

$$\partial(W^2)/\partial a = 0$$

$$\Rightarrow \quad \frac{\partial}{\partial a}\left\{(\omega^2\mu\epsilon a^2-\pi^2)(d^2-a^2)\right\} = 0$$

$$\Rightarrow \quad a = 0 \quad \text{or} \quad a = \sqrt{\frac{d^2}{2}+\frac{\pi^2}{2\omega^2\mu\epsilon}}$$

For $a = 0$ the area of cross section of the waveguide goes to 0 and therefore the power carried goes to zero. The maximum power corresponds to

$$a = \sqrt{\frac{d^2}{2}+\frac{\pi^2}{2\omega^2\mu\epsilon}} = \sqrt{\frac{d^2}{2}+\frac{\lambda^2}{8}}$$

At 10 GHz, the wavelength $\lambda = 3$ cm. Hence, we get

$$a = \sqrt{\frac{25}{2}+\frac{9}{8}} = 3.69 \text{ cm}$$

$$b = \sqrt{25-a^2} = 3.37 \text{ cm}$$

The waveguide dimensions for maximum power transfer therefore are 3.69 × 3.37 cm

EXAMPLE 6.11 A rectangular waveguide made of copper has 7 cm × 4 cm cross section. The waveguide is filled with a material with dielectric constant 3.0 and loss tangent 10^{-5}. Determine all the modes which will propagate at 3 GHz. What is the total attenuation of the waveguide in dB/m for the dominant mode. The conductivity of copper is 5.88×10^7 ℧/m.

Solution:

The cut-off frequency of a mode is given by

$$f_c = \frac{1}{2\pi\sqrt{\mu_0\epsilon_0\epsilon_r}}\left\{\left(\frac{m\pi}{a}\right)^2+\left(\frac{n\pi}{b}\right)^2\right\}^{1/2}$$

Since,

$$\frac{1}{\sqrt{\mu_0\epsilon_0}} = 3\times10^8 \text{ m/sec,}$$

$$f_c = \frac{1.5\times10^8}{\sqrt{3.0}}\left\{\left(\frac{m}{0.07}\right)^2+\left(\frac{n}{0.04}\right)^2\right\}^{1/2}$$

$$= \frac{1.5\times10^{10}}{\sqrt{3.0}}\left\{\left(\frac{m}{7}\right)^2+\left(\frac{n}{4}\right)^2\right\}^{1/2} \text{ Hz}$$

For a mode to propagate $f_c < 3\times10^9$ Hz

$$\Rightarrow \quad \frac{m^2}{49}+\frac{n^2}{16} < \frac{3}{25}$$

Following comnbinations of m and n are possible.

$m = 1, 0, 1, 2$

$n = 0, 1, 1, 0$

$\Rightarrow \quad TE_{10}, TE_{01}, TM_{11}, TE_{11}, TE_{20}.$

Ignoring the loss, the phase constant of TE_{10} mode is

$$\beta = \sqrt{\omega^2\mu_0\epsilon_0\epsilon_r-(\pi/a)^2}$$

$$= \sqrt{(2\pi\times3\times10^9)^2\mu_0\epsilon_0\times3.0-(\pi/0.07)^2} = 99.14 \text{ rad/m}$$

The attenuation constant due to dielectric loss is

$$\alpha_d = \frac{\omega^2\mu_0\epsilon_0\tan\delta}{2\beta} = 0.2\times10^{-3} \text{ nepers/m}$$

The attenuation constant due to conductor loss is

$$\alpha_c = \frac{R_s\left[1+\frac{2b}{a}\left(\frac{f_c}{f}\right)^2\right]}{\eta b\sqrt{1-\left(\frac{f_c}{f}\right)^2}}$$

where $$f_c = \frac{1}{2a\sqrt{\mu\epsilon}} = 1.237 \text{ GHz}$$

and $$R_s = \sqrt{\frac{\omega\mu}{2\sigma}} = 0.014$$

$$\Rightarrow \quad \alpha_c = 0.438 \times 10^{-3} \text{ nepers/m}$$

Hence,

$$\alpha = \alpha_d + \alpha_c \simeq 0.638 \times 10^{-3}$$

$$\Rightarrow \quad \alpha = 5.54 \times 10^{-3} \text{ dB/m}$$

EXAMPLE 6.12 Find three lowest order modes which can propagete in a waveguide having conducting fins extending along the length of the guide as shown in Fig. 6.24. Also find the cut-off frequencies of these modes.

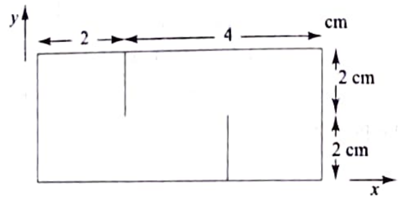


Fig. 6.24 *Rectangular waveguide with conducting fins.*

Solution:

The conducting fins impose additional boundary conditions on the modal fields. That is, vertical component of the electric field must go to zero at the fins.
For TE_{m0} mode the electric field has only y-component which is parallel to the fins. In this case then TE_{10} mode cannot propagate as its E-fields cannot be zero at the fins.
For TE_{m0} to propagate, therefore, we must have m an integer multiple of 3 i.e. $m = 3, 6, 9 \ldots$, etc.
For TE_{0n} mode the electric field is x-oriented and therefore is perpendicular to the fins. Since, normal component of electric field does not have a boundary condition, there is no constraint on 'n'. We can hence have modes like

$TE_{30}, TE_{60}, TE_{90} \ldots$
$TE_{01}, TE_{02}, TE_{03} \ldots$
$TE_{31}, TE_{32}, TE_{33} \ldots$
$TM_{31}, TM_{32}, TM_{33} \ldots$ etc

The three lowest order modes are TE_{01}, TE_{30} and TE_{02}

The cutoff frequencies of the modes are

$$f_{cTE_{01}} = \frac{3 \times 10^8}{2\pi}\left(\frac{\pi}{0.04}\right) = 3.75 \text{ GHz}$$

$$f_{cTE_{02}} = \frac{3 \times 10^8}{2\pi}\left(\frac{2\pi}{0.04}\right) = 7.5 \text{ GHz}$$

$$f_{cTE_{10}} = \frac{3 \times 10^8}{2\pi}\left(\frac{3\pi}{0.06}\right) = 7.5 \text{ GHz}$$

6.9 CAVITY RESONATOR

In the previous sections, we discussed propagation of modes in a rectangular waveguide. In the direction of the modal propagation the waveguide is assumed to be of infinite length and consequently we assume only one travelling wave along $+z$ direction along the waveguide. Let us now investigate what will happen if we take a section of a waveguide and close it from both sides with metal plates?

By placing a metal plate at the end of the waveguide the propagation of the electromagnetic wave is blocked and the wave is reflected. Obviously, in this situation then we have two waves travelling in opposite directions. The two waves together have to satisfy the boundary conditions at the two ends of the waveguide. For the sake of clarity let us investigate the fields for the TE_{mn} modes inside this structure.

For the TE_{mn} mode the longitudinal magnetic fields for the travelling wave is $+z$ direction is given as (see Eqn (6.102)).

$$H_z = D\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \tag{6.205}$$

For the reflected TE_{mn} wave let the amplitude be D'. The longitudinal magnetic field for the reflected wave is

$$H_z' = D'\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-j\beta z} \tag{6.206}$$

The total longitudinal magnetic field inside the waveguide then is

$$H_Z = H_z + H_z' \tag{6.207}$$

Now, since, the magnetic field component H_z is perpendicular to the closing plates at $z = 0$ and $z = d$, its value should be zero.
Making $H_z = 0$ at $z = 0$ we get

$$D = -D' \tag{6.208}$$

Equation (6.207) can be written as

$$H_z = D\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)\left(e^{-j\beta z} - e^{j\beta z}\right)$$

$$= -2jD\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)\sin(\beta z) \tag{6.209}$$

Now, substituting $H_z = 0$ at $z = d$, we get

$$\beta d = p\pi; \qquad p = 1, 2, 3, \ldots \tag{6.210}$$

$$\Rightarrow \qquad \beta = \frac{p\pi}{d} \tag{6.211}$$

Substituting for β from Eqn (6.211) in Eqn (6.92) we get,

$$\omega^2\mu\epsilon = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2 \tag{6.212}$$

$$\Rightarrow \qquad \omega = \frac{1}{\sqrt{\mu\epsilon}}\left\{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2\right\}^{1/2} \tag{6.213}$$

Since, m, n and p are integers, Eqn (6.213) suggests that there are discrete frequencies at which the electromagnetic fields can be excited inside the closed waveguide. These are the resonant frequencies of the closed waveguide and the closed waveguide is called the 'Cavity resonator'. Since the fields are characterized by three indices, the mode is designated as TE_{mnp}.

Although the above analysis was carried out for the TE_{mn} mode, same arguments are applicable to TM_{mn} modes also. Both TE_{mnp} and TM_{mnp} fields at discrete frequencies can exist inside a cavity resonator.

The cavity resonator which is a closed hollow metallic box can be excited by a voltage or current probes mounted at suitable location on the box. In the following section we obtain the field distribution of the TE_{101} mode of the cavity resonator.

6.9.1. Fields for TE_{101} Mode

For TE_{101} mode ($m = 1, n = 0$), the longitudinal magnitude field from Eqn (6.209) is given as

$$H_z = -2jD\cos\left(\frac{\pi x}{a}\right)\sin\left(\frac{\pi z}{d}\right) \tag{6.214}$$

Substituting for H_z and H_z' in Eqns (6.88) to (6.91), we get the other field components as

$$H_x = 2j\frac{a}{d}D\sin\left(\frac{\pi x}{a}\right)\cos\left(\frac{\pi z}{d}\right) \tag{6.215}$$

$$E_y = -2\omega_0\mu D\left(\frac{a}{\pi}\right)\sin\left(\frac{\pi x}{a}\right)\sin\left(\frac{\pi z}{d}\right) \tag{6.216}$$

Here, ω_0 is the resonant frequency of the TE_{101} mode and is given as

$$\omega_0 = \frac{1}{\sqrt{\mu\epsilon}}\left\{\left(\frac{\pi}{a}\right)^2 + \left(\frac{\pi}{d}\right)^2\right\}^{1/2} \tag{6.217}$$

The electric and magnetic fields inside the resonator for TE_{101} are shown in Fig. 6.25.

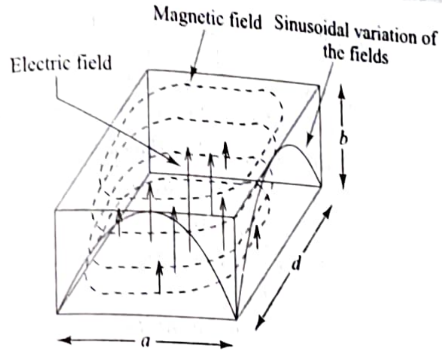


Fig. 6.25 Electric and magnetic fields inside a rectangular cavity resonator for the TE_{101} mode.

6.9.2 Energy Stored and Quality Factor

Every resonant circuit stores electromagnetic energy and has the characteristic parameter, the quality factor Q.

The average electrical energy stored inside the cavity is

$$W_e = \frac{\epsilon}{2}\int_{x=0}^{a}\int_{y=0}^{b}\int_{z=0}^{d}\frac{1}{2}Re(\mathbf{E}\cdot\mathbf{E}^*)dxdydz$$

$$= \epsilon|D|^2\left(\frac{\omega\mu a}{\pi}\right)^2\int\int\int\sin^2\left(\frac{\pi x}{a}\right)\sin^2\left(\frac{\pi z}{d}\right)dxdydz$$

$$\Rightarrow \qquad W_e = \frac{abd\epsilon}{4}|D|^2\left(\frac{\omega_0\mu a}{\pi}\right)^2 \tag{6.218}$$

Similarly, the average magnetic energy stored inside the cavity is

$$W_m = \frac{\mu}{2}\int_{x=0}^{a}\int_{y=0}^{b}\int_{z=0}^{d}\frac{1}{2}Re(\mathbf{H}\cdot\mathbf{H}^*)dxdydz$$

$$= \mu|D|^2\int_{x=0}^{a}\int_{y=0}^{b}\int_{z=0}^{d}\left\{\left(\frac{a}{d}\right)^2\sin^2\left(\frac{\pi x}{a}\right)\cos^2\left(\frac{\pi z}{d}\right)+\cos^2\left(\frac{\pi x}{a}\right)\sin^2\left(\frac{\pi z}{d}\right)\right\}$$
$$dxdydz$$

$$= \frac{abd\mu}{4}|D|^2\left\{\left(\frac{a}{d}\right)^2 + 1\right\}$$

$$= W_e \tag{6.219}$$

(Note: $(a/d)^2 + 1 = \omega_0^2\mu\epsilon a\pi$ from Eqn (6.213).)

The average electrical and magnetic energy is same for the cavity resonator. This is indentical to what happens in the LC circuit at the resonance.

The quality factor, Q for a resonator is a measure of loss in the resonator.

Assuming the loss in the cavity to be small, the field distribution of a lossy cavity is almost the same as that for a lossless cavity. From the knowledge of the magnetic field, we can obtain the surface current on the walls of the waveguide and the conductor loss. The quality factor of the cavity resonator is given as

$$Q = \omega_0 \frac{\textit{Energy stored in the cavity}}{\textit{Power loss in the cavity}} \tag{6.220}$$

$$= \omega_0 \frac{W_e + W_m}{\textit{Power loss in the cavity}}$$

$$= \frac{2\omega_0 W_e}{\textit{Power loss in the cavity}} \tag{6.221}$$

The quality factor for TE_{101} mode is

$$Q_{101} = \frac{\mu^2\epsilon(\omega_0 a d)^3}{2\pi^2 R_s(2a^3b + 2d^3b + a^3d + d^3a)} \tag{6.222}$$

Where R_s is the surface resistance of the cavity walls ($R_s = \sqrt{\omega_0\mu/2\sigma}$; σ is the conductivity of the cavity walls).

The cavity resonators can acheive a Q of few thousand very easily and therefore are popular at microwave frequncies for realizing highly tuned systems.

EXAMPLE 6.13 Show that for a rectangular waveguide, the dominant mode exhibits minimum attenuation due to conductor loss at certain frequency, f_{min}. Find the frequency with respect to the cut-off frequency of the mode. If, for the waveguide $a = 3b$, find f_{min}.

Solution:

The attenuation constant due to the conductor loss is given by Eqn (6.204) as

$$\alpha_c = \frac{R_s\left[1 + \frac{2b}{a}(\frac{f_c}{f})^2\right]}{\eta b\sqrt{1-(\frac{f_c}{f})^2}}$$

where surface resistance $R_s = \sqrt{\pi\mu f/\sigma}$

Making $\partial\alpha_c/\partial f = 0$, we get the frequency at which α_c is minimum, as

$$f_{min} = f_c\left\{\frac{3(a+b)}{2a} + \sqrt{\left(\frac{3(a+b)}{2a}\right)^2 - \frac{2b}{a}}\right\}^{1/2}$$

For $a = 3b$, we get

$$f_{min} = f_c\left\{\frac{3(3b+b)}{6b} + \sqrt{\left(\frac{3(3b+b)}{6b}\right)^2 - \frac{2b}{3b}}\right\}^{1/2}$$

$$= 2.2\ f_c$$

EXAMPLE 6.14 A section of a rectangular waveguide of cross-section 2 cm × 1.5 cm is to be used as a delay line in a radar at 10 GHz. What should be the length of the section to realize a delay of 10 nsec?

Solution:

Normally the waveguide is operated in the TE_{10} mode. The cut-off frequency of the TE_{10} mode is (given $a = 2$ cm $= 0.02$ m)

$$\omega_c = \frac{1}{\sqrt{\mu\epsilon}}\frac{\pi}{a} = \frac{3\times10^8\pi}{0.02}$$

$$= 1.5\pi \times 10^{10} \text{ rad/s}$$

The phase constant of the mode is

$$\beta = \frac{\omega}{c}\sqrt{1-\left(\frac{\omega_c}{\omega}\right)^2}$$

The group velocity of the mode is

$$v_g = \frac{\partial\omega}{\partial\beta} = \frac{c}{\sqrt{1-(\frac{\omega_c}{\omega})^2}}$$

At 10 GHz, $\omega = 2\pi \times 10^{10}$ rad/sec, giving $\omega_c/\omega = 0.75$. The group velocity is

$$v_g = \frac{c}{\sqrt{1-(0.75)^2}} = 0.66c = 1.98\times10^8 \text{ m/s}$$

For a delay of 10 nsec ($= 10^{-8}$ sec), the length of the waveguide is

$$L = v_g t = 1.98\times10^8 \times 10^{-8} = 1.98 \text{ m}$$

EXAMPLE 6.15 For a rectangular cavity of dimensions 3 cm × 2 cm × 4 cm, find the three lowest resonant modes and their resonant frequencies.

Solution:

For the TE_{mnp} modes, any two indices have to be non-zero. Whereas, for the TM_{mnp} modes, all the indices have to be non-zero.

The three lowest order modes would be TE_{101}, TE_{011}, TE_{110}, and their frequencies would be (from Eqn (6.213))

$$f_{101} = \frac{3\times10^8}{2}\left\{\left(\frac{1}{0.03}\right)^2 + \left(\frac{1}{0.04}\right)^2\right\}^{1/2}$$

$$= 6.25 \text{ GHz}$$

$$f_{101} = \frac{3 \times 10^8}{2}\left\{\left(\frac{1}{0.02}\right)^2 + \left(\frac{1}{0.04}\right)^2\right\}^{1/2}$$

$$= 8.38 \text{ GHz}$$

$$f_{101} = \frac{3 \times 10^8}{2}\left\{\left(\frac{1}{0.03}\right)^2 + \left(\frac{1}{0.02}\right)^2\right\}^{1/2}$$

$$= 9.01 \text{ GHz}$$

EXAMPLE 6.16 A cubical cavity resonator made of copper is to be designed to operate at 10 GHz. Find the dimensions of the cavity, its quality factor and its bandwidth.

Solution:

The lowest mode of the cavity is TE_{101}. For a cubical cavity $a = b = d$, and the resonant frequency is

$$f_0 = \frac{3 \times 10^8}{2}\left\{\left(\frac{1}{a}\right)^2 + \left(\frac{1}{b}\right)^2 + \left(\frac{1}{d}\right)^2\right\}^{1/2}$$

$$= \frac{3 \times 10^8}{\sqrt{2}a} = 10^{10} \quad (given)$$

$$\Rightarrow \qquad a = 2.12 \text{ cm}$$

The dimensions of the cavity are 2.12 cm × 2.12 cm × 2.12 cm.

For copper walls the conductivity is 5.8×10^7 ℧/m. The surface resistance at 10 GHz is

$$R_s = \sqrt{\frac{\pi f \mu_0}{\sigma}} = 0.026\ \Omega$$

The quality factor from Eqn (6.222) is

$$Q = \frac{2\pi \mu_0^2 \epsilon f_0^3 a^3}{3R_s} = 10716$$

The Bandwith of the resonator is

$$BW = f_0/Q = 0.933 \text{ MHz}$$

EXAMPLE 6.17 A rectangular cavity of size 5 cm × 3 cm × 4 cm is filled with a liquid of dielectric constant 4 and loss-tangent 10^{-3} at 6 GHz. Assuming that the liquid parameters are independent of the frequency, find the quality factor of the resonator for the lowest mode. Assume the losses in the cavity walls to be negligible.

Solution:

The resonant frequency for the lowest (TE_{101}) mode is

$$f_0 = \frac{3 \times 10^8}{2\sqrt{\epsilon_r}}\left\{\left(\frac{1}{a}\right)^2 + \left(\frac{1}{d}\right)^2\right\}^{1/2}$$

$$= \frac{3 \times 10^8}{2\sqrt{4}}\left\{\left(\frac{1}{0.05}\right)^2 + \left(\frac{1}{0.04}\right)^2\right\}^{1/2}$$

$$= 2.4 \text{ GHz}$$

The loss-tangent of the liquid at 6 GHz is 10^{-3}, i.e.

$$\tan\delta = 10^{-3} = \frac{\sigma}{\omega\epsilon} = \frac{\sigma}{2\pi \times 6 \times 10^9 \times \epsilon_0 \times 4}$$

$$\Rightarrow \qquad \sigma = \frac{4}{3} \times 10^{-3} \text{ ℧/m}$$

For calculating the quality factor we need the total power dessipation in the cavity and the total energy stored in the cavity. The electrical energy stored in the cavity is

$$W_e = \frac{\epsilon}{2}\int_{vol} \mathbf{E} \cdot \mathbf{E}^* dv$$

$$= \frac{\epsilon}{2}\int_{vol} |\mathbf{E}|^2 dv$$

We have seen earlier that, inside a resonator the electric energy and magnetic energy are equal. The total energy stored in the cavity therefore is $2W_e$.

The power dessipated in the cavity is

$$P_L = \int_{vol} \mathbf{E} \cdot \mathbf{J} dv = \int_{vol} \mathbf{E} \cdot (\sigma \mathbf{E}) dv$$

$$= \sigma \int_{vol} |\mathbf{E}|^2 dv$$

The quality factor of the cavity is

$$Q = 2\pi f_0 \frac{2W_e}{P_L} = \frac{\omega_0 \epsilon}{\sigma}$$

$$= \frac{1}{\text{Loss-tangent at 2.4 GHz}}$$

$$= \left(\frac{2.4}{6.0}\right) \times 10^{-3} = 400$$

6.10 SUMMARY

In this Chapter a very important structure called the 'waveguide' has been discussed. A variety of waveguides are found in practice. However, the rectangular metallic waveguides are the most common ones. To physically understand the wave propagation inside a rectangular waveguide, first the parallel plane waveguides have been investigated. A general approach for analysing waveguides has been developed subsequently. The concept of modal propagation has been introduced and some typical modes of parallel plane and rectangular waveguides are discussed. A three dimensional visualization of the modal fields and the surface currents on the waveguide walls have been developed. The power loss calculations for a practical waveguide have been presented towards the end of the chapter. In the next chapter, we investigate another important waveguiding structures, the dielectric waveguides.

Review Questions

6.1 What is a waveguide?

6.2 Can we have modal propagation in an unbound medium?

6.3 What is a parallel plane waveguide?

6.4 What are TE and TM modes?

6.5 At what angle does a uniform plane wave be launched inside a parallel plane waveguide so as to have sustained propagation?

6.6 As the height of a parallel plane waveguide increases, what happens to the number of modes?

6.7 What does the modal index '*m*' signify?

6.8 How does the phase velocity of a modal field varies as a function of frequency?

6.9 For a mode how does the phase velocity varies as a function of the waveguide height?

6.10 What is the cut-off frequency?

6.11 If frequency is less than the cut-off frequency, what kind of fields are excited inside the waveguide?

6.12 What is the phase velocity if the frequency is equal to cut-off frequency?

6.13 Can TEM mode exist inside a parallel plane waveguide? Explain why.

6.14 Can TEM mode be dispersive? Why?

6.15 What are the boundary conditions on the surface of a waveguide?

6.16 Can TE_{mn} and TM_{mn} modes exist inside a rectangular waveguide for any value of *m* and *n*? Explain your answer.

6.17 Why does TEM mode not exist inside a rectangular waveguide?

6.18 What is the dominant mode of a parallel plane waveguide?

6.19 What is the dominant mode of a rectangular waveguide?

6.20 What is the minimum dimension of a waveguide for propagation of a mode?

6.21 What is a guided wavelength? What are the bounds on guided wavelength?

6.22 What is group velocity of a mode?

6.23 In a rectangular waveguide the lowest order mode is TE_{10} mode. Why can we not tell about the next higher order mode?

6.24 Why is it not necessary to have current flow in the waveguide walls in the direction of net power flow inside the waveguide?

6.25 What are the factors contributing to the loss inside a waveguide?

6.26 What is a cavity resonator?

6.27 What is the special feature of a cavity resonator?

6.28 How do we excite a cavity resonator?

6.29 What is the typical range of the quality factor for the cavity resonator made of copper at microwave frequencies?

Problems

6.1 A uniform plane wave at 150 MHz frequency is incident at an ideal conducting boundary. The angle of incidence is 30°. At what distance from the conducting boundary, can an ideal conducting plane be introduced without affecting the field distribution?

6.2 In Problem 6.1, if the amplitude of the incident field is 10 V/m and it has perpendicular polarization find the magnitude and direction of the total electric and magnetic field at some point. What is the power density of the wave?

6.3 In a parallel plane waveguide the electromagnetic energy travels through multiple reflections of a plane wave. At what angles does the plane wave have a sustained propagation. The width of the waveguide is 50 cm and the frequency of the wave is 1 GHz.

6.4 In Problem 6.3 what is the effective phase constant of the fields propagating in the waveguide.

6.5 For a parallel plane waveguide filled with a material with $\mu_r = 10$ and $\epsilon = 4$, the phase constant of a mode at 1 GHz is 50 rad/m. What is the cut off frequency of the mode?

6.6 Inside an air-filled waveguide the total magnetic field is given as

$$\mathbf{H} = 10\cos(\pi x)e^{-j\beta z}\hat{y}\ \mathrm{A/m}$$

Find the vector electric field, the phase constant β and the cut off frequency of the wave. The frequency of the wave is 2 GHz.

6.7 Show that if $\omega - \beta$ relation for a waveguide is linear, an electromagnetic pulse travels undistorted on the waveguide.

6.8 The cut-off frequency of the TM_2 mode inside a parallel plane waveguide is 1.5 GHz. Find the group and phase velocities of the TE_1 mode at 1.5 GHz.

6.9 A parallel plane waveguide is filled with a material with dielectric constant 9. The height of the waveguide is 50 cm. At 1 GHz how many modes can be propagated inside the waveguide and what are their cut-off frequencies?

6.10 A parallel plane waveguide is excited in a TE mode such that the electric field is zero mid way between the two planes. What should be the height of the waveguide so as to get a guided wavelength of 50 cm at 1 GHz? What would be the cut off frequency of the mode for that waveguide?

6.11 Inside a waveguide for a particular mode, the phase velocity is three times the group velocity at 5 GHz. Find the cut-off frequency and cut-off wavelength of the mode.

6.12 A parallel plane waveguide is filled with a material of dielectric constant 3 for $z < 0$. The height of the waveguide is 10 cm. A TE_1 mode at 1 GHz is propagating in $+z$ direction over $z < 0$ region. Show that the modal power will be completely reflected from $z = 0$ interface. If the amplitude of electric field for the mode is 10 V/m find the amplitude of the electric field at $z = 1$ m.

6.13 For a particular TM mode the cut off frequency for the waveguide for $z < 0$ is 10 GHz. The waveguide is filled with a material with dielectric constant 4 for $z > 0$. Show that the mode will be completely transmitted across $z = 0$ junction if the frequency of the mode is $5\sqrt{5}$ GHz.

6.14 Derive the expressions for the transverse field components in terms of the logitudinal field components E_z and H_z, in Cartesian and cylindrical coordinate systems.

6.15 A rectangular waveguide has 4 cm × 3 cm cross section. What is the frequency range over which there is single mode propagtion?

6.16 For a rectangular waveguide filled with a dielectric material, the cut off frequency of the dominant mode is 3 GHz. If the waveguide has the cross section 3 cm × 2 cm, find the cut-off frequencies of TE_{20}, TE_{01} and TM_{21} modes.

6.17 For a 5 cm × 3 cm rectangular waveguide, the maximum peak electric field of the dominant mode at 5 GHz is 10 V/m. Find the maximum peak magnetic field inside the waveguide. Also find the total power carried by the waveguide.

6.18 For a mode inside a rectangular waveguide the longitudinal magnetic field is given as

$$H_z = 20\cos(10y)e^{-j\beta z}$$

Amp/m. Find the cut-off frequency of the mode. Also, find the frequency at which the group velocity is 1/3 of the phase velocity.

6.19 Inside a square rectangular waveguide the cut off frequency of the dominant mode is 2.5 GHz. What is the frequency range over which there will be four modes propagating inside the waveguide?

6.20 The size of a rectangular waveguide is 10 cm× 6 cm. Find the three modes with lowest cut off frequencies. What are those cut-off frequencies?

6.21 For a 5 cm× 3 cm rectangular waveguide the cut off frequency of TM_{11} mode is 1.5 GHz. What will be the group velocity of the dominant mode at 1.5 GHz ?

6.22 A rectangular waveguide has conducting fins running along its length as shown in the Fig. 6.26. Find the lowest frequency which will propagate on the waveguides.

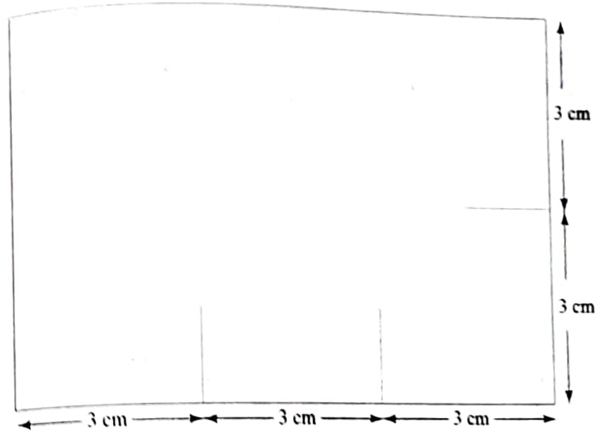


Fig. 6.26

6.23 Inside a parallel plane waveguide of height 10 cm, the TEM mode propagate with a power density of 50 W/m^2. What is the rms value of the surface current on the walls of the waveguide? If the conductivity of the walls is 4×10^6 ℧/m, what is the power loss per unit length per unit width of the waveguide at the frequency of 3 GHz.

6.24 A rectangular waveguide carrying the dominant mode is closed with a conducting plate at the far end, and excited at 4 GHz at the other end. The electric field is measured to be zero at a distance of 5 cm from the far end inside the waveguide. Find the guided wavelength, cut off frequency and the group velocity of the mode.

6.25 A rectangular waveguide has 1.5 GHz cut-off frequency. The waveguide is filled with a material having dielectric constant 2.5 and loss tangent of 10^{-3} at 1 GHz. Find the phase and attenuation constant for the mode at 3 GHz. Assume the walls of the waveguide to be made of ideal conductor.

6.26 In a 5 cm×3 cm waveguide the TE_{01} mode is propagating at 10 GHz. The total power carried by the waveguide is 10 W. If the conductivity of the waveguide walls is 2×10^7 ℧/m, find the attenuation constant of the waveguide in dB/m.

6.27 A TE_{21} mode inside a 10 cm × 7 cm rectangular waveguide carries 100 W power. What are the maximum electric and magnetic fields inside the waveguide and at what location do they occur? Draw the surface currents on the inner walls of the waveguide. Frequency is 10 GHz.

6.28 A resonant cavity of 6 cm × 3 cm × 4 cm is excited in the lowest mode. The peak electric field inside the cavity is 100 V/m. Find the resonant frequency of the cavity and the total energy stored inside the cavity.

6.29 A cubical cavity is excited in the TM_{111} mode. Find the expressions for the electric and magnetic fields inside the cavity, its resonant frequency and its total energy content.

6.30 In Problem 6.29 if the cavity has a linear dimension of 4 cm and is made of copper, find its quality factor and the 3 dB bandwidth.

CHAPTER 7

Dielectric Waveguide

Dielectric waveguide is a dielectric structure which can guide electromagnetic waves. At frequencies beyond millimeter wavelengths, the metallic waveguides have excessive conductor loss. The dielectric waveguides are, therefore, more promising in sub-millimeter and optical wavelength range. The dielectric waveguides have played a prominent role in advancements in semiconductor lasers and optical communication. In fact, almost all the guided wave and integrated optics devices work on the principle of dielectric waveguide.

There are two types of dielectric waveguides namely, slab waveguides and cylindrical waveguides. Slab waveguides find applications in thin film and integrated optical devices, whereas, optical fibers are cylindrical dielectric waveguides. Due to the impetus received in the last few decades, the dielectric waveguides have become a rather important subject in electromagnetics.

In this chapter, we investigate the slab as well as cylindrical dielectric waveguides. Since, the light is an electromagnetic wave, the analysis presented in the following sections is applicable to optical propagation as well as to the millimeter or microwave propagation.

7.1 DIELECTRIC SLAB WAVEGUIDE

The slab waveguides are generally seen in thin film technology. A thin dielectric layer is deposited on another dielectric slab called 'substrate'. The deposited dielectric layer has a higher dielectric constant compared to that of the substrate and has much less propagation loss at the frequency of operation. Therefore, in general, the slab dielectric waveguide is an asymmetric structure since it has substrate on one side and air on the other. A dielectric slab waveguide is schematically shown in Fig. 7.1.

To make the analysis simple, it is assumed that the substrate is infinitely thick. It will become clear later that this assumption is justified since the fields decay rapidly away from the waveguide-substrate interface.

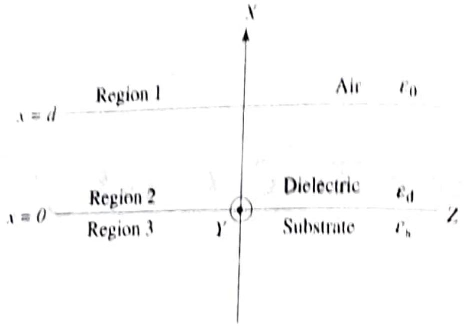


Fig. 7.1 *Dielectric slab waveguide.*

Let the waveguiding layer be having a permittivity ϵ_d and thickness d, and let the substrate be having a permittivity ϵ_s. Let us also assume that, the structure has infinite extension in y and z directions.

As discussed in the previous chapter, we can assume that the wave travels in the $+z$ direction and the longitudinal field components E_z and H_z are the independent components. On the lines similar to that of the metallic waveguides, we can assume either E_z or H_z zero giving rise to modal fields for TE and TM modes respectively. The independent components E_z and H_Z have to satisfy the wave equation

$$\nabla^2\{E_z, H_z\} + \omega^2\mu\epsilon\{E_z, H_z\} = 0 \tag{7.1}$$

$$\left\{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + \omega^2\mu\epsilon\right\}\{E_z, H_z\} = 0 \tag{7.2}$$

However, note that in case of metallic waveguide there were no fields inside the conductor, therefore, the fields existed only in the region between the conductors. In case of dielectric waveguide, the fields exist in all three regions, air, dielectric and the substrate. We, therefore, follow the following approach to the analysis of the dielectric waveguides:

1. Solve wave equation to get E_z and/or H_z in all the three media.
2. Derive transverse field components.
3. Apply boundary conditions at the two sides of the dielectric slab and obtain the unknown constants.
4. Derive modal dispersion relation and the propagation constant.

For the coordinate axes shown in Fig. 7.1 since the wave travels in $+z$ direction, the fields have a z-variation $\sim e^{-j\beta z}$, where, β is the modal phase constant, which is to be determined from the boundary conditions. Now, since, the waveguide is infinitely extended in the y-direction, there are no boundary conditions on the fields along the y-direction, and consequently the fields are a constant function of y i.e. independent of y. We, therefore, have

$$\{E_z, H_z\} = X(x)e^{-j\beta z} \tag{7.3}$$

Let us now take the cases of TE and TM propagation by making E_z and H_z zero respectively.

7.1.1 Transverse Electric (TE) Mode

For a TE mode, $E_z = 0$ and all transverse fields are represented in terms of H_z only.

$$H_z = X(x)e^{-j\beta z} \tag{7.4}$$

Substituting Eqn (7.3) in Eqn (7.2) we get

$$\frac{1}{X}\frac{d^2X}{dx^2} - \beta^2 + \omega^2\mu\epsilon = 0 \tag{7.5}$$

Since Eqn (7.5) has to be satisfied for every value of x, we must have

$$\frac{1}{X}\frac{d^2X}{dx^2} = \text{constant} = \pm k^2 \text{ (say)} \tag{7.6}$$

$$\Rightarrow \quad \frac{d^2X}{dx^2} = \pm k^2 X \tag{7.7}$$

k is a real quantity.

An appropriate sign for k has to be chosen, depending upon the nature of the fields in a particular region. In the present case, we are investigating the wave propagation in the dielectric slab. So, ideally the fields should be confined to the slab only. However, this is not possible because due to the continuity of fields at the dielectric boundaries there will be some fields in region 1 (air) and in region 3 (substrate). The best we can then do is to ensure that the fields are confined as close to the slab as possible. That is, we choose the fields in air and in the substrate such that they decay rapidly away from the slab. This behavior can be obtained by taking +sign for k. On the other hand, within the slab the modal fields are expected to be generated due to superposition of plane waves (as discussed in the case of parallel plane waveguides) giving oscillatory fields. These fields can be obtained by choosing -sign for k. We therefore find that for Eqn (7.7) +sign should be chosen in region 1 and 3, and -sign should be chosen in region 2.

Since, the modal propagation constant, β, is same in all three media, k has different values in the three media. Let k be denoted by k_1, k_2, k_3 in three media respectively. The Eqns (7.7) and (7.5) for the three media can be written as

Region 1

$$\frac{d^2X_1}{dx^2} = k_1^2 X_1 \tag{7.8}$$

$$\text{and} \quad k_1^2 - \beta^2 + \omega^2\mu\epsilon_0 = 0 \tag{7.9}$$

Region 2

$$\frac{d^2X_2}{dx^2} = -k_2^2 X_2 \tag{7.10}$$

and $$-k_2^2 - \beta^2 + \omega^2\mu\epsilon_d = 0 \tag{7.11}$$

Region 3

$$\frac{d^2X_3}{dx^2} = k_3^2 X_3 \tag{7.12}$$

and $$k_3^2 - \beta^2 + \omega^2\mu\epsilon_s = 0 \tag{7.13}$$

The solutions of Eqns (7.8), (7.10) and (7.12) can be written as

$$X_1 = A_1 e^{-k_1 x} + A_2 e^{k_1 x} \tag{7.14}$$

$$X_2 = A_3 \cos k_2 x + A_4 \sin k_2 x \tag{7.15}$$

$$X_3 = A_5 e^{-k_3 x} + A_6 e^{k_3 x} \tag{7.16}$$

Equation (7.14) represents field in region 1, and the two terms on the RHS correspond to exponentially decaying and exponentially growing fields as a function of distance x from the slab waveguide boundary. Since, the objective here is to confine energy to the slab, the growing field has to be identically zero, making $A_2 \equiv 0$.

Similarly, for having only decaying fields in region 3, $A_5 \equiv 0$. Note that, in region 3 since x is negative, $e^{k_3 x}$ represents decaying field and not $e^{-k_3 x}$.

Substituting for X_1, X_2, X_3 with $A_2 = 0$ and $A_5 = 0$ in Eqn (7.4) we get the longitudinal component of magnetic field in the three regions as

$$H_{z1} = A_1 e^{-k_1 x} e^{-j\beta z} \tag{7.17}$$

$$H_{z2} = \{A_3 \cos k_2 x + A_4 \sin k_2 x\}\, e^{-j\beta z} \tag{7.18}$$

$$H_{z3} = A_6 e^{k_3 x} e^{-j\beta z} \tag{7.19}$$

Substituting for H_z in Eqns (6.59) to (6.62) we can obtain the transverse field components as (note that h^2 in Eqns (6.59) to (6.52) is equal to $-k_1^2$, and $-k_3^2$ respectively in region 1 and 3, and is equal to k_2^2 in region 2).

Region 1:

$$E_{x1} = -\frac{j\omega\mu}{-k_1^2}\frac{\partial H_z}{\partial y} = 0 \tag{7.20}$$

$$E_{y1} = -\frac{j\omega\mu}{-k_1^2}\frac{\partial H_z}{\partial x} = -\frac{j\omega\mu}{k_1} A_1 e^{-k_1 x} e^{-j\beta z} \tag{7.21}$$

$$H_{x1} = \frac{-j\beta}{-k_1^2}\frac{\partial H_z}{\partial x} = \frac{-j\beta}{k_1} A_1 e^{-k_1 x} e^{-j\beta z} \tag{7.22}$$

$$H_{y1} = \frac{-j\beta}{-k_1^2}\frac{\partial H_z}{\partial y} = 0 \tag{7.23}$$

Region 2:

$$E_{x2} = 0 \tag{7.24}$$

$$E_{y2} = \frac{j\omega\mu}{k_2}\{A_3 \sin k_2 x - A_4 \cos k_2 x\}\, e^{-j\beta z} \tag{7.25}$$

$$H_{x2} = \frac{j\beta}{k_2}\{A_3 \sin k_2 x - A_4 \cos k_2 x\}\, e^{-j\beta z} \tag{7.26}$$

$$H_{y2} = 0 \tag{7.27}$$

Region 3:

$$E_{x3} = 0 \tag{7.28}$$

$$E_{y3} = \frac{j\omega\mu}{k_3} A_6 e^{k_3 x} e^{-j\beta z} \tag{7.29}$$

$$H_{x3} = \frac{j\beta}{k_3} A_6 e^{k_3 x} e^{-j\beta z} \tag{7.30}$$

$$H_{y3} = 0 \tag{7.31}$$

Since, there are no surface currents on the dielectric interfaces, continuity of tangential components of electric and magnetic fields at $x = 0$ and $x = d$ gives,

At $x = 0$,

$$H_{z2} = H_{z3} \tag{7.32}$$

$$\Rightarrow \quad A_3 = A_6 \tag{7.33}$$

and

$$E_{y2} = E_{y3} \tag{7.34}$$

$$\Rightarrow \quad \frac{-A_4}{k_2} = \frac{A_6}{k_3} \tag{7.35}$$

At $x = d$,

$$H_{z1} = H_{z2} \tag{7.36}$$

$$\Rightarrow \quad A_1 e^{-k_1 d} = A_3 \cos k_2 d + A_4 \sin k_2 d \tag{7.37}$$

and

$$E_{y1} = E_{y2} \tag{7.38}$$

$$\Rightarrow \quad \frac{-A_1}{k_1} e^{-k_1 d} = \frac{1}{k_2}\{A_3 \sin k_2 d - A_4 \cos k_2 d\} \tag{7.39}$$

Equations (7.33) (7.35) (7.37) and (7.39) can be written in a matrix form as

$$\begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & \frac{1}{k_2} & \frac{1}{k_3} \\ e^{-k_1 d} & -\cos k_2 d & -\sin k_2 d & 0 \\ \frac{-e^{-k_1 d}}{k_1} & \frac{-\sin k_2 d}{k_2} & \frac{\cos k_2 d}{k_2} & 0 \end{bmatrix} \begin{bmatrix} A_1 \\ A_3 \\ A_4 \\ A_6 \end{bmatrix} = 0 \tag{7.40}$$

Since, the coefficients A_1, A_3, A_4, A_6 cannot be zero, the solution of Eqn (7.40) exists provided the determinant of the (4×4) matrix is identically zero. Making the determinant equal to zero, we get the so called 'characteristic equation' of a mode as

$$\tan k_2 d = \frac{(k_1 + k_3)k_2}{k_2^2 - k_1 k_3} \tag{7.41}$$

From Eqn (7.9) and (7.11) we get a relation between k_1 and k_2 as

$$k_1^2 + k_2^2 + \omega^2 \mu \epsilon_0 - \omega^2 \mu \epsilon_d = 0 \tag{7.42}$$

$$\Rightarrow \qquad k_1 = \sqrt{\omega^2 \mu\, (\epsilon_d - \epsilon_0) - k_2^2} \tag{7.43}$$

and from Eqns (7.11) and (7.13) we get relation between k_3 and k_2 as

$$k_3^2 + k_2^2 + \omega^2 \mu \epsilon_s - \omega^2 \mu \epsilon_d = 0 \tag{7.44}$$

$$\Rightarrow \qquad k_3 = \sqrt{\omega^2 \mu (\epsilon_d - \epsilon_s) - k_2^2} \tag{7.45}$$

Equations (7.41), (7.43) and (7.45) are to be solved simultaneously for k_2 using some numerical method. Once k_2 is known, using Eqn (7.11) we can find the propagation constant β of the mode.

From the above analysis we can make the following important observations:

(a) The characteristic equation has multiple solutions representing propagation of discrete modes.

(b) Since k_1, k_2, k_3 have to be real, from Eqns (7.43) and (7.45) we must have

$$\omega^2 \mu\, (\epsilon_d - \epsilon_0) > k_2^2 \tag{7.46}$$

and

$$\omega^2 \mu\, (\epsilon_d - \epsilon_s) > k_2^2 \tag{7.47}$$

As k_2^2 is always positive, Eqns (7.46) and (7.47) imply

$$\epsilon_d > \epsilon_0 \tag{7.48}$$

and

$$\epsilon_d > \epsilon_s \tag{7.49}$$

Equation (7.48) is a trivial condition as the permittivity of any dielectric medium is always greater than air (vacuum). The Condition (7.49) demands that the dielectric constant of the slab must be greater than that of the substrate. In general, therefore, the guided wave propagation in the slab is possible provided the slab dielectric constant is greater than that of both the media enclosing it.

(c) For a given frequency ω, k_2 can vary between 0 and $\omega\sqrt{\mu(\epsilon_d - \epsilon_s)}$ implying that the RHS of Eqn (7.41) has a finite range and therefore there are always finite number of solutions to Eqn (7.41). In other words, at a given frequency, only a finite number of modes can propagate in the waveguide.

(d) When k_2^2 becomes greater than $\omega^2\mu\,(\epsilon_d - \epsilon_s)$, k_3 becomes imaginary and the function $e^{k_3 x}$ does not represent a decaying field. Instead, it represents a wave traveling in $-x$ direction in the substrate, that is, it represents leakage of wave in the substrate. We can then say that, the mode is no more a purely guided mode. This condition is called the cut-off of the mode, and the frequency at which k_3 changes from real to imaginary, is called the 'cut-off' frequency of the mode. The cut-off frequency therefore corresponds to $k_3 = 0$ and is given as

$$\omega_c = \frac{k_2}{\sqrt{\mu(\epsilon_d - \epsilon_s)}} \tag{7.50}$$

Note that this cut-off condition is rather different than that for a metallic waveguide. In a metallic waveguide the mode is said to be cut-off when the propagation constant $\beta \to 0$. In the dielectric waveguide however, at the cut-off the β does not approach zero, i.e. the mode keeps propagating along the waveguide even below the cut-off except, it starts losing energy in the direction perpendicular to the waveguide inside the substrate.

(e) From Eqn (7.11) we have

$$\beta^2 = \omega^2 \mu \epsilon_d - k_2^2 \tag{7.51}$$

since k_2^2 can vary from 0 to $\omega^2\mu(\epsilon_d - \epsilon_s)$, β^2 has to lie between $\omega^2\mu\epsilon_s$ and $\omega^2\mu\epsilon_d$ i.e.

$$\omega\sqrt{\mu\epsilon_s} < \beta < \omega\sqrt{\mu\epsilon_d} \tag{7.52}$$

Recall that $\omega\sqrt{\mu\epsilon_s}$ is the wave number in the substrate and $\omega\sqrt{\mu\epsilon_d}$ is the wave number in the dielectric slab. The propagation constant β is, therefore, bounded by the wave numbers of the slab and the substrate. This seems intuitively correct since a part of the modal energy is in the slab and a part is in the substrate. Depending upon the energy distribution, β tends to $\omega\sqrt{\mu\epsilon_d}$ or $\omega\sqrt{\mu\epsilon_s}$. If there is relatively large energy in the slab, $\beta \to \omega\sqrt{\mu\epsilon_d}$ whereas if there is more energy in the substrate, $\beta \to \omega\sqrt{\mu\epsilon_s}$. The value of β relative to wave numbers of the slab and the substrate therefore provides a qualitative measure of the energy distribution in the two media.

7.1.2 Physical Picture of the Modal Propagation in a Slab Waveguide

The modal propagation inside a parallel plane waveguide could be visualized as the interference of plane waves bouncing between two conducting planes. A similar physical picture can be developed for the dielectric slab waveguide as well. In a parallel plane waveguide, however, the reflection of plane waves from the conducting plane was perfect. But in the dielectric waveguide, a perfect reflection can occur only if the wave is totally internally reflected. Indeed, that happens in the slab waveguide as can be seen from the following arguments.

(i) The Eqns (7.48) and (7.49) insist that for guided wave propagation, $\epsilon_d > \epsilon_0$ and ϵ_s. That is, the slab medium has to be denser compared to its surrounding media. This is the essential condition for the total internal reflection (TIR).

(ii) The fields in the substrate and in air are exponentially decaying. For a plane wave inside the slab, this is possible only if the wave is total internally reflected. As we know, at TIR, the fields in the rarer medium always decay exponentially.

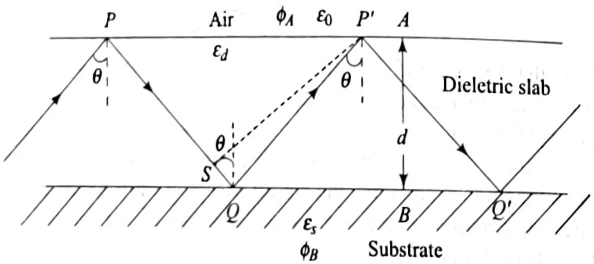


Fig. 7.2 *Physical picture of modal propagation inside a dielectric slab waveguide.*

So, the picture of modal propagation inside a slab waveguide is as shown in Fig. 7.2. A plane wave is total internally reflected at two sides A and B of the slab. For TIR on both the surfaces, the angle θ must be greater than the critical angles on both the surfaces. On surface A the critical angle is

$$\theta_{cA} = \sin^{-1}\sqrt{\frac{\epsilon_0}{\epsilon_d}} \tag{7.53}$$

and on surface B, the critical angle is

$$\theta_{cB} = \sin^{-1}\sqrt{\frac{\epsilon_s}{\epsilon_d}} \tag{7.54}$$

since θ_{cB} is greater than θ_{cA} (since $\epsilon_s > \epsilon_0$), we must have

$$\theta > \theta_{cB} = \sqrt{\frac{\epsilon_s}{\epsilon_d}} \tag{7.55}$$

Now from Section 5.5 we know that at the TIR the wave undergoes a phase change, which is a function of θ, ϵ_d, ϵ_s, and the wave polarization.

Let us draw a constant phase plane for the ray $P'Q'$ at point P' as shown by the dotted line in Fig. 7.2. Let the constant phase plane intersect the ray PQ at point S. The constant phase plane is common to the ray PQ as well as $P'Q'$. That is, the phase of the point P' on the ray $P'Q'$ is same as that of the point S on the ray PQ. In other words, the phase change which the ray undergoes from S to P' should be integer multiple of 2π radians.

The distance travelled by the ray from S to P' is $QP'(\cos 2\theta + 1)$, and $QP' = d\sec\theta$. The propagation phase, therefore, is

$$\phi_{prop} = \omega\sqrt{\mu\epsilon_d}\, d\sec\theta(\cos 2\theta + 1) = 2\omega\sqrt{\mu\epsilon_d}\, d\cos\theta \tag{7.56}$$

(Note: $\omega\sqrt{\mu\epsilon_d}$ is the phase constant of a plane wave in the slab.)

Now if the phase changes due to TIR at Q and P' are ϕ_B and ϕ_A respectively, the condition for sustained wave propagation inside the slab is

$$\phi_{prop} + \phi_A + \phi_B = 2m\pi, \quad m = 0, 1, 2, \ldots \tag{7.57}$$

$$\Rightarrow \quad 2\omega\sqrt{\mu\epsilon_d}\, d\cos\theta + \phi_A + \phi_B = 2m\pi \tag{7.58}$$

ϕ_A and ϕ_B are functions of θ.

In Eqn (7.58) since m is an integer, the equation has solution for discrete values of θ. That is, the waves only at specific angles can have sustained propagation inside the slab. This discrete nature is same as the modal propagation. In fact Eqn (7.41) can be derived from (7.58) by proper substitution for ϕ_A and ϕ_B. The derivation is beyond the scope of this book and interested readers can refer to advanced literature on the subject.

7.2 CIRCULAR DIELECTRIC WAVEGUIDE

As mentioned in the previous sections, a circular dielectric waveguide is often found in the form of an optical fiber although it can be used for guiding microwaves and millimeter waves as well. As the name suggests, for this waveguide the guiding structure is a circular rod of dielectric material. The energy propagates along the axis of the rod. Conceptually, the operation of the circular waveguide is same as that of a dielectric slab or metallic waveguide. In this case also, the modal propagation can be visualized as superposition of total internally reflected waves inside the dielectric rod. However, the analysis of this waveguide is relatively complex due to use of cylindrical coordinate system appropriate for the circular rod geometry. In the following, we carry out the analysis of a circular dielectric waveguide using the general approach of solving the wave equation.

A circular waveguide is schematically shown in Fig. 7.3. Let the dielectric rod be having radius, a, and permittivity ϵ_1. Also let us assume that the rod is surrounded by an infinitely large medium with permittivity ϵ_2. If the rod is placed in the free-space, $\epsilon_2 = \epsilon_0$.

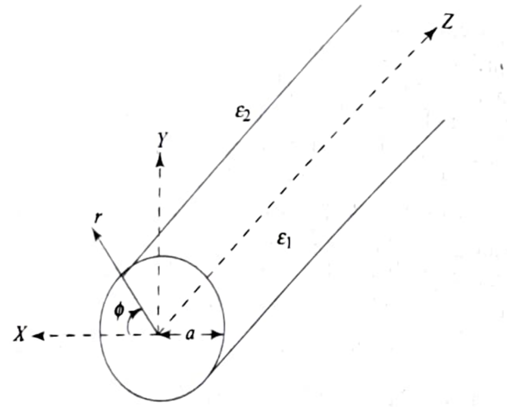


Fig. 7.3 *Circular dielectric waveguide.*

As discussed earlier, let us assume that the longitudinal field components E_z and H_z are independent, and the transverse components H_r, H_ϕ, E_r, E_ϕ are expressed in terms of E_z and H_z. In general, the approach of analysis would then be to solve the wave equation for E_z and H_z inside and outside the rod, find transverse components, apply boundary conditions at the surface of the rod, and find the modal fields and the propagation constant.

The wave equation for E_z and H_z in cylindrical coordinates can be written as

$$\nabla^2\{\psi\} + \omega^2\mu\epsilon\{\psi\} = 0 \qquad (7.59)$$

$$\Rightarrow \quad \left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial z^2} + \omega^2\mu\epsilon\right]\psi = 0 \qquad (7.60)$$

where $\psi = E_z$ or H_z and $\epsilon = \epsilon_1$ inside the rod, and $\epsilon = \epsilon_2$ outside the rod.

Equation (7.60) can be solved using separation of variables. Let the solution be represented as

$$\psi = R(r)\Phi(\phi)Z(z) \qquad (7.61)$$

Since, we are interested in investigating here a traveling mode along $+z$ direction (along the axis of the rod) the field should have z variation as $e^{-j\beta z}$, where β is the modal propagation constant, yet to be determined. We, therefore, have

$$Z(z) = e^{-j\beta z} \qquad (7.62)$$

Let us now consider a point (r, ϕ, z) in the waveguide. Keeping r and z fixed if we vary ϕ, the point moves along a circle in a plane transverse to the axis of the rod (z direction). For a change in ϕ by multiples of 2π we reach to the same point (r, ϕ, z), i.e. $(r, \phi, z) \equiv (r, \phi + 2m\pi, z)$; m is an integer. Consequently, we should have

$$\psi(r, \phi, z) \equiv \psi(r, \phi + 2\pi m, z) \qquad (7.63)$$

This can be achieved if we choose

$$\Phi(\phi) = e^{j\nu\phi} \qquad \nu \text{ is an integer} \qquad (7.64)$$

(Note: $e^{j\nu(\phi+2m\pi)} = e^{j\nu\phi}e^{j2\nu m\pi} = e^{j\nu\phi}$).

Substituting $Z(z)$ and $\Phi(\phi)$ in the wave function Eqn (7.61), we have

$$\psi(r, \phi, z) = R(r)e^{j\nu\phi}e^{-j\beta z} \qquad (7.65)$$

The wave Eqn (7.60) becomes,

$$\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial z^2} + \omega^2\mu\epsilon\right]R(r)e^{j\nu\phi}e^{-j\beta z} = 0 \qquad (7.66)$$

$$\Rightarrow \quad \frac{1}{r}\frac{d}{dr}\left(r\frac{dR}{dr}\right) - \frac{\nu^2 R}{r^2} - \beta^2 R + \omega^2\mu\epsilon R = 0 \qquad (7.67)$$

$$\Rightarrow \quad \frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + \left\{(\omega^2\mu\epsilon - \beta^2) - \frac{\nu^2}{r^2}\right\}R = 0 \qquad (7.68)$$

Equation (7.68) is the well known Bessel's equation. The equation cannot be solved in closed from. The series solutions of the equation are called the Bessel functions. General properties of Bessel functions are summarized in Appendix D. However, some discussion on the nature of the solution of the equation is in order at this point.

Let us define

$$q^2 = \omega^2\mu\epsilon - \beta^2 \qquad (7.69)$$

For a travelling field, β is real, and for lossless media μ and ϵ are real. Therefore q^2 is also real though it could be positive or negative. The quantity q could therefore be either purely real (if q^2 is positive) or purely imaginary (if q^2 is negative). Depending upon the sign of q^2, the Bessel's equation has different solutions. Equation (7.68) being a second order differential equation, has two solutions.

For $q^2 > 0$, the solutions are called the Bessel functions and the Neumann functions, and are denoted by $J_\nu(qr)$ and $N_\nu(qr)$ respectively. ν is called the order of the function and the quantity in the brackets is called the argument of the function. The general solution to Eqn (7.68) can then be written as a linear combination of two functions as

$$R(r) = \alpha_1 J_\nu(qr) + \alpha_2 N_\nu(qr) \qquad (7.70)$$

where α_1, α_2 are the arbitrary constants.

For $q^2 < 0$, the solutions are called the Modified Bessel functions, and are denoted $K_\nu(qr/j)$ and $I_\nu(qr/j)$ respectively. Since, in this case q is imaginary, (qr/j) is a real quantity. The solution to Eqn (7.68) in this case can be written as

$$R(r) = \eta_1 K_\nu(qr/j) + \eta_2 I_\nu(qr/j) \qquad (7.71)$$

where η_1 and η_2 are arbitrary constants.

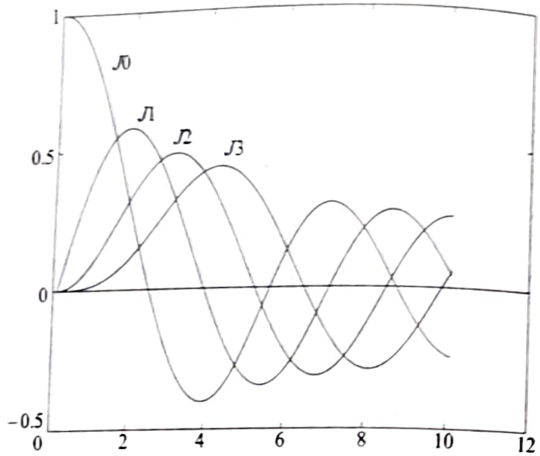


Fig. 7.4 *Bessel functions $J_\nu(x)$ as functions of x.*

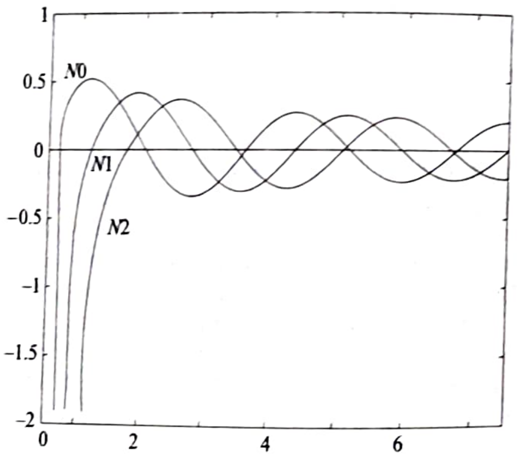


Fig. 7.5 *Neumann functions $N_\nu(x)$ as functions of x.*

Since, the propagation constant β is unknown, it is not clear at this point 'whether q is real or imaginary and which solution is to be chosen to the wave equation?' In fact, this ambiguity cannot be resolved unless we make use of the physical understanding of the problem, or some a priori knowledge about the nature of the fields. To this end, let us first investigate the behavior of different Bessel functions. Figure 7.4 to 7.7 show the variation of the functions $J_\nu(x)$, $N_\nu(x)$, $K_\nu(x)$ and $I_\nu(x)$ as a function of their argument, x. We can make following observations from Figures 7.4 to 7.7.

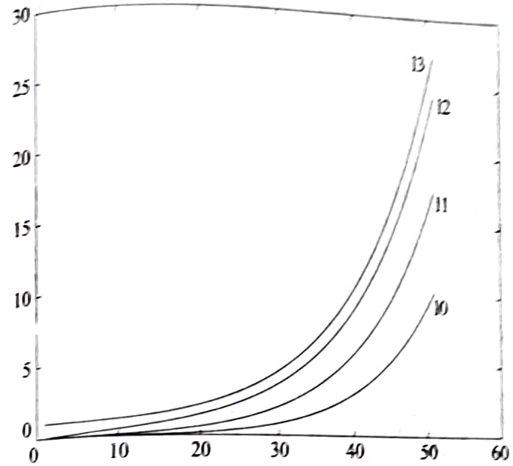


Fig. 7.6 *Modified Bessel functions $I_\nu(x)$ as functions of x.*

1. The function $J_\nu(x)$ is always finite for all values of x. Except $J_0(x)$ which is 1 at $x = 0$, all other $J_\nu(x)$ are zero at $x = 0$.
2. The functions $N_\nu(x)$ asymptotically diverge to $-\infty$ as the argument $x \to 0$, though it is finite at all values of $x \neq 0$.
3. Both functions $J_\nu(x)$ and $N_\nu(x)$ have oscillatory behavior as a function of x.
4. The functions $K_\nu(x)$ are monotonically decreasing functions of x, and they go to zero as $x \to \infty$.
5. The functions $I_\nu(x)$ are monotonically increasing functions of x, and they diverge to ∞ as $x \to \infty$.
6. Both functions $K_\nu(x)$ and $I_\nu(x)$ are monotonic tunctions of x.

Let us refer to the problem of guided mode propagation. We say a mode is guided when its fields are confined to the guide, and outside the guide the fields decay monotonically. Also, as seen in the ease of slab waveguide, the fields inside the guide are due to superposition of total internally reflected waves, and therefore, exhibit an amplitude variation going through maxima and minima in space due to constructive and distructive interference. For a guided mode we hence expect a spatially oscillatory field inside the dielectric rod, and a decaying field outside of it.

The choice of solution inside and outside the rod is quite obvious now. Inside the rod, we should have a solution given by Eqn (7.70), and outside the rod solution should be given by Eqn (7.71). Note that while doing this, we have unknowingly put some constraints on q and consequently on the propagation constant β. Inside the rod, q has to be real $(q^2 > 0)$ and therefore from Eqn (7.69)

$$\beta^2 < \omega^2\mu\epsilon_1 \qquad \text{for } 0 < r < a \tag{7.72}$$

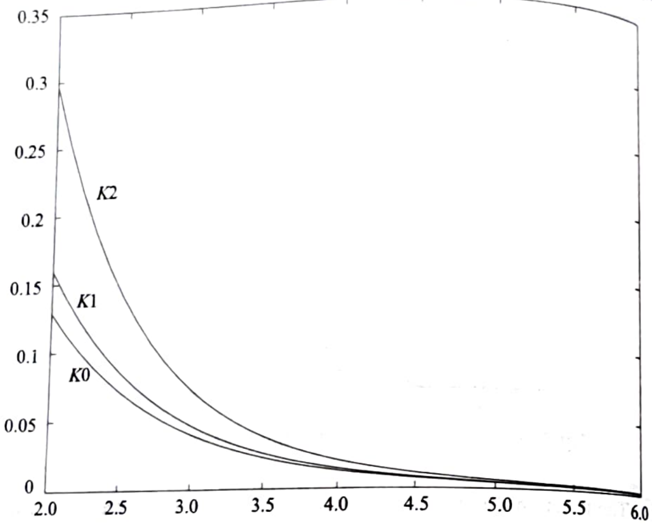


Fig. 7.7 *Modified Bessel functions $K_\nu(x)$ as functions of x.*

On the other hand, outside the rod, the q has to be imaginary $(q^2 < 0)$ and consequently from Eqns (7.69)

$$\beta^2 > \omega^2\mu\epsilon_2 \qquad \text{for } r > a \tag{7.73}$$

The propagation constant β therefore is bounded by the wave numbers of the two media, i.e.

$$\omega\sqrt{\mu\epsilon_2} < \beta < \omega\sqrt{\mu\epsilon_1} \tag{7.74}$$

This result is identical to what we obtained for the slab waveguide (See Eqn (7.52)).

Although the solution given by Eqn (7.70) is appropriate for $r < a$ and the solution given by Eqn (7.71) is appropriate for $r > a$, we note that both the terms in Eqn (7.70) and Eqn (7.71) do not correctly represent the field behavior. Since, the fields have to be finite everywhere in the space, α_2 and η_2 should be identically zero since N_ν and I_ν functions are not finite everywhere in the space.

The solution to the wave equation in the two regions, inside and outside the rod can hence be written as

$$\psi_1(r,\phi,z) = \alpha_1 J_\nu(ur)e^{j\nu\phi}e^{-j\beta z} \qquad \text{for } r < a \tag{7.75}$$

and

$$\psi_2(r,\phi,z) = \eta_1 K_\nu(wr)e^{j\nu\phi}e^{-j\beta z} \qquad \text{for } r > a \tag{7.76}$$

where we have defined

$$q \equiv u = \sqrt{\omega^2\mu\epsilon_1 - \beta^2} \qquad \text{for } r < a \tag{7.77}$$

and

$$q \equiv jw = j\sqrt{\beta^2 - \omega^2\mu\epsilon_2} \qquad \text{for } r > a \tag{7.78}$$

From Eqns (7.77) and (7.78) it is clear that u and w are real quantities.

Having developed the general solution to the wave equation, now we are in a position to investigate different kind of modes and their characteristics.

Using above analysis, we can find the longitudinal fields E_z and H_z inside and outside the dielectric rod. Substitution of E_z and H_z in Eqns (6.52) and (6.53) yields the transverse field components. The transverse fields in cylindrical coordinates can be written as

$$\mathbf{E}_\perp \equiv E_r\hat{\mathbf{r}} + E_\phi\hat{\boldsymbol{\phi}} \tag{7.79}$$

$$= \frac{-j\omega\mu}{h^2}\left\{\frac{1}{r}\frac{\partial H_z}{\partial\phi}\hat{\mathbf{r}} - \frac{\partial H_z}{\partial r}\hat{\boldsymbol{\phi}}\right\} - \frac{\gamma}{h^2}\left\{\frac{\partial E_z}{\partial r}\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial E_z}{\partial\phi}\hat{\boldsymbol{\phi}}\right\} \tag{7.80}$$

$$= \left\{\frac{-j\omega\mu}{h^2}\frac{1}{r}\frac{\partial H_z}{\partial\phi} - \frac{\gamma}{h^2}\frac{\partial E_z}{\partial r}\hat{\mathbf{r}}\right\} + \left\{\frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial r} - \frac{\gamma}{h^2}\frac{1}{r}\frac{\partial E_z}{\partial\phi}\right\}\hat{\boldsymbol{\phi}} \tag{7.81}$$

and

$$\mathbf{H}_\perp \equiv H_r\hat{\mathbf{r}} + H_\phi\hat{\boldsymbol{\phi}} \tag{7.82}$$

$$= \frac{j\omega\epsilon}{h^2}\left\{\frac{1}{r}\frac{\partial E_z}{\partial\phi}\hat{\mathbf{r}} - \frac{\partial E_z}{\partial r}\hat{\boldsymbol{\phi}}\right\} - \frac{\gamma}{h^2}\left\{\frac{\partial H_z}{\partial r}\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial H_z}{\partial\phi}\right\}\hat{\boldsymbol{\phi}} \tag{7.83}$$

$$= \left\{\frac{j\omega\epsilon}{h^2}\frac{1}{r}\frac{\partial E_z}{\partial\phi} - \frac{\gamma}{h^2}\frac{\partial H_z}{\partial r}\right\}\hat{\mathbf{r}} - \left\{\frac{j\omega\epsilon}{h^2}\frac{\partial E_z}{\partial r} + \frac{\gamma}{h^2}\frac{1}{r}\frac{\partial H_z}{\partial\phi}\right\}\hat{\boldsymbol{\phi}} \tag{7.84}$$

where for the dielectric waveguide $h = q$ and $\gamma = j\beta$.

7.2.1 Transverse Electric (TE) Mode

For a TE mode, $E_z = 0$ and the longitudinal field H_z is functionally given by Eqns (7.75) and (7.76), yielding

$$H_{z1} = \alpha_1 J_\nu(ur)e^{j\nu\phi}e^{-j\beta z} \qquad r < a \tag{7.85}$$

$$H_{z2} = \eta_1 K_\nu(wr)e^{j\nu\phi}e^{-j\beta z} \qquad r > a \tag{7.86}$$

Substituting for E_z and H_z in Eqns (7.81) and (7.84) we have the transverse fields as

For $r < a$ (inside the waveguide)

$$E_{r1} = \frac{-j\omega\mu}{u^2}\frac{1}{r}\frac{\partial H_{z1}}{\partial\phi} = \alpha_1\frac{\omega\mu}{u^2}\frac{\nu}{r}J_\nu(ur)e^{j\nu\phi}e^{-j\beta z} \tag{7.87}$$

$$E_{\phi 1} = \frac{j\omega\mu}{u^2}\frac{\partial H_{z1}}{\partial r} = \alpha_1\frac{j\omega\mu}{u}J'_\nu(ur)e^{j\nu\phi}e^{-j\beta z} \tag{7.88}$$

$$H_{r1} = \frac{-j\beta}{u^2}\frac{\partial H_{z1}}{\partial r} = \alpha_1 \frac{-j\beta}{u} J_\nu'(ur)e^{j\nu\phi}e^{-j\beta z} \tag{7.89}$$

$$H_{\phi 1} = \frac{-j\beta}{u^2}\frac{1}{r}\frac{\partial H_{z1}}{\partial \phi} = \alpha_1 \frac{\beta}{u^2}\frac{\nu}{r} J_\nu(ur)e^{j\nu\phi}e^{-j\beta z} \tag{7.90}$$

For $r > a$ (outside the waveguide)

$$E_{r2} = \frac{-j\omega\mu}{-w^2}\frac{1}{r}\frac{\partial H_{z2}}{\partial \phi} = \eta_1 \frac{-\omega\mu}{w^2}\frac{\nu}{r} K_\nu(wr)e^{j\nu\phi}e^{-j\beta z} \tag{7.91}$$

$$E_{\phi 2} = \frac{j\omega\mu}{-w^2}\frac{\partial H_{z2}}{\partial r} = \eta_1 \frac{-j\omega\mu}{w} K_\nu'(wr)e^{j\nu\phi}e^{-j\beta z} \tag{7.92}$$

$$H_{r2} = \frac{-j\beta}{-w^2}\frac{\partial H_{z2}}{\partial r} = \eta_1 \frac{j\beta}{w} K_\nu'(wr)e^{j\nu\phi}e^{-j\beta z} \tag{7.93}$$

$$H_{\phi 2} = \frac{-j\beta}{-w^2}\frac{1}{r}\frac{\partial H_{z2}}{\partial \phi} = \eta_1 \frac{-\beta}{w^2}\frac{\nu}{r} K_\nu(wr)e^{j\nu\phi}e^{-j\beta z} \tag{7.94}$$

Here $'$ represents the derivative with respect to the whole argument, i.e. $J_\nu'(x) = \frac{\partial}{\partial x}\{J_\nu(x)\}$ and $K_\nu'(x) = \frac{\partial}{\partial x}\{K_\nu(x)\}$.

Application of the boundary conditions that the tangential components of E and H are continuous at the surface of the rod (i.e. at $r = a$) yields

$$H_{z1} = H_{z2} \tag{7.95}$$

and

$$E_{\phi 1} = E_{\phi 2} \tag{7.96}$$

Substituting for H_{z1}, H_{z2}, $E_{\phi 1}$ and $E_{\phi 2}$ we get

$$\alpha_1 J_\nu(ua) = \eta_1 K_\nu(wa) \tag{7.97}$$

and

$$\frac{\alpha_1}{u} J_\nu'(ua) = -\frac{\eta_1}{w} K_\nu'(wa) \tag{7.98}$$

Instead of the z component of the magnetic field, if we take the ϕ component, which is also tangential to the surface, we get (at $r = a$)

$$H_{\phi 1} = H_{\phi 2} \tag{7.99}$$

$$\Rightarrow \quad -\frac{\nu\alpha_1}{u^2} J_\nu(ua) = \frac{\nu\eta_1}{w^2} K_\nu(wa) \tag{7.100}$$

Now we have a difficulty. Equation (7.100) is inconsistent with Eqn (7.97) and Eqn (7.98). The inconsistency can be resolved only if $H_{\phi 1}$ and $H_{\phi 2}$ are identically made zero. This is possible only if, either, β or η_1 or ν is zero. η_1 can not become zero as it represents the field amplitude, and β can not become zero because then the mode will not be a traveling mode. The only option therefore is, $\nu = 0$. Since, ν represents the field variation in the azimuth, $\nu = 0$ means circularly symmetric fields.

This indeed is a very profound conclusion as it implies that the fields of the TE-modes essentially have to be circularly symmetric. Substituting $\nu = 0$ in the field Eqns (7.87) and (7.94) we get

$$E_{r1} = E_{r2} \equiv 0 \tag{7.101}$$

and

$$H_{\phi 1} = H_{\phi 2} \equiv 0 \tag{7.102}$$

The TE-mode therefore has only H_z, H_r and E_ϕ components as

Inside the waveguide ($r < a$)

$$H_{z1} = \alpha_1 J_0(ur)e^{-j\beta z} \tag{7.103}$$

$$E_{\phi 1} = \alpha_1 \frac{j\omega\mu}{u} J_0'(ur)e^{-j\beta z} \tag{7.104}$$

$$H_{r1} = -\alpha_1 \frac{j\beta}{u} J_0'(ur)e^{-j\beta z} \tag{7.105}$$

Outside the waveguide ($r > a$)

$$H_{z2} = \eta_1 K_0(wr)e^{-j\beta z} \tag{7.106}$$

$$E_{\phi 2} = \eta_1 \frac{j\omega\mu}{w} K_0'(wr)e^{-j\beta z} \tag{7.107}$$

$$H_{r2} = -\eta_1 \frac{j\beta}{w} K_0'(wr)e^{-j\beta z} \tag{7.108}$$

Dividing Eqn (7.98) by Eqn (7.97) we can eliminate the arbitrary constants to yield the so called 'Characteristic Equation' of the TE mode as

$$\frac{J_0'(ua)}{u J_0(ua)} = -\frac{K_0'(wa)}{w K_0(wa)} \tag{7.109}$$

Since $J_0'(ua)$ and $J_0(ua)$ are oscillatory functions of ua, the characteristic Eqn (7.109) has multiple solutions. This aspect will be discussed in detail a little later.

A TE mode is designated by TE_{0m}, where the first suffix 0 corresponds to $\nu = 0$, and m represents the solution number (i.e. first solution, second solution, etc).

7.2.2 Transverse Magnetic (TM) Mode

For the transverse magnetic mode, $H_z \equiv 0$ and the transverse fields are expressed in terms of E_z. On the lines similar to that of the TE mode, it can be shown that in this case also we have to have $\nu = 0$. That is, a TM mode also has to be circularly symmetric. Substituting $\nu = 0$, the longitudinal component of the electric field can be written as

$$E_{z1} = \alpha_1 J_0(ur)e^{-j\beta z} \tag{7.110}$$

$$E_{z2} = \eta_1 K_0(wr)e^{-j\beta z} \tag{7.111}$$

The corresponding transverse components are

Inside the waveguide ($r < a$)

$$E_{r1} = \frac{-j\beta}{u^2}\frac{\partial E_{z1}}{\partial r} = \frac{-j\beta}{u}\alpha_1 J_0'(ur)e^{-j\beta z} \tag{7.112}$$

$$H_{\phi 1} = \frac{-j\omega\epsilon_1}{u^2}\frac{\partial E_{z1}}{\partial r} = \frac{-j\omega\epsilon_1}{u}\alpha_1 J_0'(ur)e^{-j\beta z} \tag{7.113}$$

Outside the waveguide ($r > a$)

$$E_{r2} = \frac{-j\beta}{-w^2}\frac{\partial E_{z2}}{\partial r} = \frac{j\beta}{w}\eta_1 K_0'(wr)e^{-j\beta z} \tag{7.114}$$

$$H_{\phi 2} = \frac{-j\omega\epsilon_2}{-w^2}\frac{\partial E_{z2}}{\partial r} = \frac{j\omega\epsilon_2}{w}\eta_1 K_0'(wr)e^{-j\beta z} \tag{7.115}$$

Continuity of the tangential components at the waveguide surface ($r = a$) yields,

$$E_{z1} = E_{z2} \tag{7.116}$$

and $$H_{\phi 1} = H_{\phi 2} \tag{7.117}$$

Substituting for E_{z1}, E_{z2}, $H_{\phi 1}$, $H_{\phi 2}$, we get

$$\alpha_1 J_0(ua) = \eta_1 K_0(wa) \tag{7.118}$$

and $$-\frac{\epsilon_1}{u}\alpha_1 J_0'(ua) = \frac{\epsilon_2}{w}\eta_1 K_0'(wa) \tag{7.119}$$

Dividing Eqn (7.119) by Eqn (7.118) we get the characteristic equation of the TM mode as

$$\frac{\epsilon_1}{u}\frac{J_0'(ua)}{J_0(ua)} = -\frac{\epsilon_2}{w}\frac{K_0'(wa)}{K_0(wa)} \tag{7.120}$$

This equation is similar to that of the TE mode, and has multiple solutions. A TM mode also is designated as TM_{0m} mode, where m is the solution number.

7.2.3 Cut-off of TE and TM Modes

As discussed in the case of the slab wave guide, a mode is said to be cut-off if the fields outside the waveguide do not remain of decaying nature. In case of the circular waveguide, the decaying fields are appropriately represented by $K_0(wr)$ function. The modified Bessel function $K_0(wr)$ remains the modified Bessel function as long as w is real. If, w becomes imaginary, the function $K_0(wr)$ becomes a simple Bessel function (not modified Bessel function), and therefore, does not represent a decaying field. A mode, therefore, is said to be approaching cut-off when $w \to 0$.

From Eqns (7.77) and (7.78) we note that

$$u^2 + w^2 = \omega^2\mu\epsilon_1 - \omega^2\mu\epsilon_2 \tag{7.121}$$

For a given waveguide and a given operating frequency, the RHS of Eqn (7.121) is fixed implying

$$u^2 + w^2 = constant \tag{7.122}$$

i.e. both u and w cannot be zero simultaneously. When $w \to 0$, $u \to u_c = \omega\sqrt{\mu(\epsilon_1 - \epsilon_2)}$.

From the characteristic equations of the TE and TM modes (Eqns (7.109) and (7.120)) we can see that as $w \to 0$, since u cannot approach zero, $J_0(ua)$ should approach zero. $J_0(ua) = 0$, therefore, corresponds to the cut-off of the TE or TM mode.

$$\Rightarrow \quad u_c a = \chi_{0m} \tag{7.123}$$

where χ_{0m} is the m^{th} root of the J_0 Bessel function. Substituting for $u_c = \omega\sqrt{\mu(\epsilon_1 - \epsilon_2)}$, we get the modal cut-off frequency as

$$\omega_c = \frac{\chi_{0m}}{a\sqrt{\mu(\epsilon_1 - \epsilon_2)}} \tag{7.124}$$

The first root of J_0 function, $\chi_{01} = 2.4$ and consequently, the lowest frequency which can propagate in TE or TM mode is

$$\omega_{c_{01}} = \frac{2.4}{a\sqrt{\mu(\epsilon_1 - \epsilon_2)}} \tag{7.125}$$

7.2.4 Hybrid Modes

In the previous sections we concluded that in a circular dielectric waveguide, the TE and TM modes are circularly symmetric ($\nu = 0$). Then the question is, 'are the fields always circularly symmetric in a circular dielectric waveguide?' Our earlier understanding shows that if the size of any structure is comparable to the wavelength, there has to be a spatial variation in the field. It, therefore, looks unlikely that we never get the fields which are not circularly symmetric. We may, therefore, say that the circular symmetry of the fields is required only if H_z or E_z is zero. It is worthwhile to investigate the fields if both E_z and H_z are non-zero. If these fields exist, they would neither be transverse electric nor transverse magnetic. We would then call them 'hybrid modal fields'. Let us assume that the hybrid fields exist, and in general, they are not circularly symmetric. The longitudinal fields in the two media can then be written as

Inside the waveguide ($r < a$)

$$E_{z1} = A J_\nu(ur)e^{j\nu\phi - j\beta z} \tag{7.126}$$

$$H_{z1} = B J_\nu(ur)e^{j\nu\phi - j\beta z} \tag{7.127}$$

Outside the waveguide ($r > a$)

$$E_{z2} = C K_\nu(wr)e^{j\nu\phi - j\beta z} \tag{7.128}$$

$$H_{z2} = D K_\nu(wr)e^{j\nu\phi - j\beta z} \tag{7.129}$$

A, B, C, D denote the amplitudes of the respective field components.

Substituting for E_{z1}, H_{z1}, E_{z2}, H_{Z2} from Eqns (7.126) to (7.129) into Eqns (7.81) and (7.84) we can get the transverse fields as

Inside the waveguide ($r < a$)

$$E_{r1} = \frac{\omega\mu}{u^2}\frac{\nu}{r} B J_\nu(ur)e^{j\nu\phi} - \frac{j\beta}{u} A J_\nu'(ur)e^{j\nu\phi} \tag{7.130}$$

$$E_{\phi 1} = \frac{j\omega\nu}{u} B J_\nu'(ur)e^{j\nu\phi} + \frac{\beta}{u^2}\frac{\nu A}{r} J_\nu(ur)e^{j\nu\phi} \tag{7.131}$$

$$H_{r1} = -\frac{\omega\epsilon_1}{u^2}\frac{\nu}{r} A J_\nu(ur)e^{j\nu\phi} - \frac{j\beta}{u} B J_\nu'(ur)e^{j\nu\phi} \tag{7.132}$$

$$H_{\phi 1} = \frac{-j\omega\epsilon_1}{u} A J_\nu'(ur)e^{j\nu\phi} + \frac{\beta}{u^2}\frac{\nu}{r} B J_\nu(ur)e^{j\nu\phi} \tag{7.133}$$

Outside the waveguide ($r > a$)

$$E_{r2} = -\frac{\omega\mu}{w^2}\frac{\nu}{r} D K_\nu(wr)e^{j\nu\phi} + \frac{j\beta}{w} C K_\nu'(wr)e^{j\nu\phi} \tag{7.134}$$

$$E_{\phi 2} = -\frac{j\omega\mu}{w} D K_\nu'(wr)e^{j\nu\phi} - \frac{\beta}{w^2}\frac{\nu}{r} C K_\nu(wr)e^{j\nu\phi} \tag{7.135}$$

$$H_{r2} = \frac{\omega\epsilon_2}{w^2}\frac{\nu}{r} C K_\nu(wr)e^{j\nu\phi} + \frac{j\beta}{w} D K_\nu'(wr)e^{j\nu\phi} \tag{7.136}$$

$$H_{\phi 2} = \frac{j\omega\epsilon_2}{w} C K_\nu'(wr)e^{j\nu\phi} - \frac{\beta}{w^2}\frac{\nu}{r} D K_\nu(wr)e^{j\nu\phi} \tag{7.137}$$

Since, there are four unknown constants A, B, C, D to be eliminated, we need four boundary conditions.

Taking at $r = a$

$$E_{z1} = E_{z2} \tag{7.138}$$

$$H_{z1} = H_{z2} \tag{7.139}$$

$$E_{\phi 1} = E_{\phi 2} \tag{7.140}$$

$$H_{\phi 1} = H_{\phi 2} \tag{7.141}$$

We get four equations which can be written in a matrix form as

$$\begin{bmatrix} J_\nu(ua) & 0 & -K_\nu(wa) & 0 \\ 0 & J_\nu(ua) & 0 & -K_\nu(wa) \\ \frac{\beta\nu}{au^2} J_\nu(ua) & \frac{j\omega\mu}{u} J_\nu'(ua) & \frac{\beta\nu}{aw^2} K_\nu(wa) & \frac{j\omega\mu}{w} K_\nu'(wa) \\ \frac{-j\omega\epsilon_1}{u} J_\nu'(ua) & \frac{\beta\nu}{au^2} J_\nu(ua) & \frac{-j\omega\epsilon_2}{w} K_\nu'(wa) & \frac{\beta\nu}{aw^2} K_\nu(wa) \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = 0 \tag{7.142}$$

Equation (7.142) is a homogeneous equation and the solution of the equation exists if the determinant of the 4×4 matrix is identically zero. Equating the determinant to zero we get the characteristic equation of the hybrid mode as

$$\left\{\frac{J_\nu'(ua)}{uJ_\nu(ua)} + \frac{K_\nu'(wa)}{wK_\nu(wa)}\right\}\left\{\frac{\omega^2\mu\epsilon_1}{u}\frac{J_\nu'(ua)}{J_\nu(ua)} + \frac{\omega^2\mu\epsilon_2}{w}\frac{K_\nu'(wa)}{K_\nu(wa)}\right\} = \left(\frac{\beta\nu}{a}\right)^2\left\{\frac{1}{u^2}+\frac{1}{w^2}\right\}^2 \tag{7.143}$$

Note that, the Eqn (7.143) is a general equation for any value of E_z and H_z including zero. If we substitute $\nu = 0$ in Eqn (7.143), the RHS goes to zero and the equation splits into two equations,

$$\frac{J_0'(ua)}{uJ_0(ua)} + \frac{K_0'(wa)}{wK_0(wa)} = 0 \tag{7.144}$$

and

$$\frac{\epsilon_1}{u}\frac{J_0'(ua)}{J_0(ua)} + \frac{\epsilon_2}{w}\frac{K_0'(wa)}{K_0(wa)} = 0 \tag{7.145}$$

which are precisely the characteristic equations of TE and TM modes respectively as given by Eqns (7.109) and (7.120).

The above analysis clearly indicates that for a circular dielectric waveguide, there is a special class of modes with $\nu \neq 0$, and these modes are the hybrid modes. These modes are designated as $HE_{\nu m}$ modes, where m indicates that the mode corresponds to the m^{th} root of the characteristic equation.

7.2.5 Cut-off of a Hybrid Mode

Finding the cut-off frequency of a hybrid mode is not as straight-forward as that of a TE or a TM mode. One has to carry out some algebraic manipulations of the characteristic equation using recurrence relations of the Bessel functions given in Appendix D. We can expand the derivative of the Bessel function in terms of higher and lower order Bessel functions as

$$J_\nu'(x) = \frac{1}{2}\{J_{\nu-1}(x) - J_{\nu+1}(x)\} \tag{7.146}$$

$$K_\nu'(x) = -\frac{1}{2}\{K_{\nu-1}(x) + K_{\nu+1}(x)\} \tag{7.147}$$

Now substituting for derivatives of the Bessel functions in Eqn (7.143) and defining

$$J^- = \frac{J_{\nu-1}(ua)}{uJ_\nu(ua)} \tag{7.148}$$

$$J^+ = \frac{J_{\nu+1}(ua)}{uJ_\nu(ua)} \tag{7.149}$$

$$K^- = \frac{K_{\nu-1}(wa)}{wK_\nu(wa)} \tag{7.150}$$

$$K^+ = \frac{K_{\nu+1}(wa)}{wK_\nu(wa)} \tag{7.151}$$

The characteristic Eqn (7.143) can be written as

$$\{(J^- - J^+) - (K^- + K^+)\}\{\epsilon_1(J^- - J^+) - \epsilon_2(K^- + K^+)\}$$
$$= \frac{4}{\omega^2\mu}\left(\frac{\beta\nu}{a}\right)^2\left\{\frac{1}{u^2} + \frac{1}{w^2}\right\}^2 \tag{7.152}$$

Rearrangement of the terms leads to

$$-(J^- - K^-)(\epsilon_1 J^+ + \epsilon_2 K^+) - (J^+ + K^+)(\epsilon_1 J^- - \epsilon_2 K^-)$$
$$+(J^- - K^-)(\epsilon_1 J^- - \epsilon_2 K^-) + (J^+ + K^+)(\epsilon_1 J^+ + \epsilon_2 K^+)$$
$$= \frac{4}{\omega^2\mu}\left(\frac{\beta\nu}{a}\right)^2\frac{(u^2+w^2)^2}{u^4w^4} \tag{7.153}$$

which can be further modified to

$$-2(J^- - K^-)(\epsilon_1 J^+ + \epsilon_2 K^+) - 2(J^+ + K^+)(\epsilon_1 J^- - \epsilon_2 K^-)$$
$$+\{(J^- - K^-) + (J^+ + K^+)\}\{\epsilon_1(J^- + J^+) + \epsilon_2(K^+ - K^-)\}$$
$$= \frac{4}{\omega^2\mu}\left(\frac{\beta\nu}{a}\right)^2\frac{(u^2+w^2)^2}{u^4w^4} \tag{7.154}$$

Now using another relation of Bessel functions, i.e.

$$J_{\nu+1}(x) + J_{\nu-1}(x) = \frac{2\nu}{x}J_\nu(x) \tag{7.155}$$

and
$$K_{\nu+1}(x) - K_{\nu-1}(x) = \frac{2\nu}{x}K_\nu(x) \tag{7.156}$$

we obtain

$$J^+ + J^- = \frac{2\nu}{u^2a} \tag{7.157}$$

and
$$K^+ - K^- = \frac{2\nu}{w^2a} \tag{7.158}$$

Therefore, Eqn (7.154) becomes

$$-2(J^- - K^-)(\epsilon_1 J^+ + \epsilon_2 K^+) - 2(J^+ + K^+)(\epsilon_1 J^- - \epsilon_2 K^-)$$
$$+\left[\frac{2\nu}{a}\left\{\frac{1}{u^2} + \frac{1}{w^2}\right\}\frac{2\nu}{a}\left\{\frac{\epsilon_1}{u^2} + \frac{\epsilon_2}{w^2}\right\}\right]$$
$$= \frac{4}{\omega^2\mu}\left(\frac{\beta\nu}{a}\right)^2\frac{(u^2+w^2)^2}{u^4w^4} \tag{7.159}$$

Now the expression inside the square brackets in Eqn (7.159) can be simplified as

$$\frac{4\nu^2}{a^2}\frac{(u^2+w^2)}{u^4w^4}\left(\epsilon_1 w^2 + \epsilon_2 u^2\right) = \frac{4\nu^2}{a^2}\frac{(u^2+w^2)}{u^4w^4}\beta^2(\epsilon_1 - \epsilon_2)$$
$$= \frac{4\nu^2}{a^2}\frac{(u^2+w^2)}{u^4w^4}\beta^2\frac{(u^2+w^2)}{\omega^2\mu} \tag{7.160}$$
$$= \text{RHS of Eqn (7.159)}.$$

Therefore, Eqn (7.159) reduces to

$$(J^- - K^-)(\epsilon_1 J^+ + \epsilon_2 K^+) + (J^+ + K^+)(\epsilon_1 J^- - \epsilon_2 K^-) = 0 \tag{7.161}$$

This equation is fully equivalent to the original characteristic equation (7.143). This form, however, is more suited for finding the cut-off of the modes.

We know that the cut-off of a mode corresponds to $w \to 0$. For small arguments (i.e. for $x << 1$) the $K_\nu(x)$ functions can be approximated by

$$K_0(x) \approx -[\ln x + \ln C - \ln 2] = -\ln\left(\frac{xC}{2}\right) \tag{7.162}$$

and
$$K_\nu(x) \approx \frac{2^{\nu-1}(\nu-1)!}{x^\nu} \tag{7.163}$$

where C is the Euler's constant = 1.781672.
This yields for $(wa \to 0)$

$$K^+ = \frac{K_{\nu+1}(wa)}{wK_\nu(wa)} = \frac{2\nu}{w^2a} \qquad \nu = 1, 2, 3\ldots \tag{7.164}$$

$$K^- = \frac{K_{\nu-1}(wa)}{wK_\nu(wa)} = -a\ln\left(\frac{Cwa}{2}\right) \qquad \text{for } \nu = 1 \tag{7.165}$$

$$K^- = \frac{K_{\nu-1}(wa)}{wK_\nu(wa)} = \frac{a}{2(\nu-1)} \qquad \text{for } \nu = 2, 3, \ldots \tag{7.166}$$

Substituting for K^+ and K^- in Eqn (7.161) and letting $w \to 0$, we get For $\nu = 1$

$$\{2(ua)J_1(ua)\}^2\ln\left(\frac{Cwa}{2}\right) = 0 \tag{7.167}$$

$$\Rightarrow \qquad J_1(ua) = 0 \tag{7.168}$$

Other solution $ua = 0$ is included in Eqn (7.168).

For $\nu > 1$ on the other hand, we have

$$J_\nu(ua)\left\{\frac{\epsilon_1+\epsilon_2}{\epsilon_1}J_{\nu-1}(ua) - \frac{ua}{\nu-1}J_\nu(ua)\right\} = 0 \tag{7.169}$$

$$\Rightarrow J_\nu(ua) = 0;\ ua \neq 0; \qquad \text{for } \nu = 2, 3, \ldots \tag{7.170}$$

and
$$\left(\frac{\epsilon_1+\epsilon_2}{\epsilon_1}\right)J_{\nu-1}(ua) - \frac{ua}{\nu-1}J_\nu(ua) = 0 \quad \text{for } \nu = 2, 3, \ldots \tag{7.171}$$

From the above discussion it is clear that, the lowest order HE mode is the HE_{11} mode which is the first solution of the Eqn (7.168) for $\nu = 1$, and its cut-off is given by $u_c = 0$. The cut-off frequency corresponding $u_c = 0$ is 0. Therefore, we conclude that the HE_{11} mode does not have any cut-off frequency, i.e. it can propagate at any frequency up to DC. It is then interesting to note that at DC, the fields inside a dielectric rod are hybrid and not transverse. The electro-static fields (which correspond to DC) inside a dielectric rod essentially must have a field component along the axis of the rod.

Finally summarizing the cut-off behavior of all the modes we find that the first mode which can propagate on a dielectric waveguide is the HE_{11} mode and therefore, the HE_{11} mode is the dominant mode of the circular dielectric waveguide. This is the mode through which the transmission of light takes place inside a single mode optical fiber over hundreds of kilometers.

7.3 SUMMARY

In this chapter the we observed the analysis of the slab as well as circular dielectric waveguides. The slab waveguides are used in integrated optical devices, thin film technology, lasers, etc. The circular waveguides are used in the form of optical fibers in optical communication. Both types of dielectric waveguides play an important role in modern electronics and photonics. The chapter gives a flavour of the kind of analysis one needs for investigating modern waveguiding devices.

Review Questions

7.1 What is a dielectric waveguide?

7.2 What are the advantages and disadvantages of a dielectric waveguide compared to a metallic waveguide?

7.3 Give an example of dielectric waveguide.

7.4 What are the trivial conditions for propagation of an EM wave inside a dielectric waveguide?

7.5 What is the range of the phase constant for a guided mode inside a waveguide?

7.6 What is cut-off frequency of a mode inside a dielectric waveguide?

7.7 How does the cut-off condition for a dielectric waveguide differ from that for the metallic waveguide?

7.8 What happens to the number of modes if the dielectric constant of a dielectric waveguide is increased?

7.9 What is the functional form of the modal fields inside a circular waveguide?

7.10 What are the behavioral differences between the four Bessel function $J_\nu(x)$, $N_\nu(x)$, $I_\nu(x)$, $K_\nu(x)$?

7.11 Why is the Bessel function $J_\nu(x)$ appropriate for representing the fields inside the waveguide, and the modified Bessel functions $K_\nu(x)$ is appropriate for representing the fields outside the waveguide?

7.12 What is hybrid mode?

7.13 Why are TE and TM modes azimuthally symmetric?

7.14 What is the field variation outside the waveguide at the cut-off?

7.15 What is the lowest order mode which propagates on a circular dielectric waveguide?

7.16 What are the cut-off frequencies of the lowest order and TE_{01} and TM_{01} modes?

CHAPTER **8**

Antennas

In the previous chapters, we discussed the propagation of electromagnetic waves in unbound and bound media. (However, we have deliberately avoided any discussion on generation of these electromagnetic waves.) We started with an assumption that, the electromagnetic energy is put into the medium and then investigated how the energy survives in the medium. We also established that the time varying electric and magnetic fields have to always coexist in the medium constituting a wave phenomenon.

In this chapter, we first develop the foundation of generation of electromagnetic waves, and subsequently analyse structures, called antennas, which can efficiently generate electromagnetic waves. Broadly, the problem of electromagnetic wave generation is to establish a relation between the fields of an electromagnetic wave and the sources responsible for them.

We know from our basic knowledge of electrostatics that the electric field is produced by a charge. However, the same charge when placed in uniform motion, constitutes a direct current which produces magnetic field. Both of these cases are static cases as the charge or the current does not vary with time. The question we then have to ask is "If the charge and current were varying with time, what kind of fields would be generated?"

Since, a current is the rate of change of charge, a time varying current corresponds to acceleration or deceleration of charges. The question then reduces to "what fields are generated by accelerating charges?". The basic laws of physics tell us that the accelerating charges radiate energy, i.e. they throw energy in their surrounding medium. We should, therefore, expect a radiation from time varying currents. Two things should be emphasized at this point,

(i) Radiation is a phenomenon related to time varying currents. For non-time varying currents, there is no acceleration of charges and consequently no radiation.

(ii) No matter how slowly the current varies, as long as it is not constant as a function of time, there is always a possibility of radiation. Furthermore,

as the frequency of the current increases, radiation will also increase. The word possibility has been used because it is not necessary that for a time varying current there will be a radiation. This is due to the fact that, if there are two identical time varying currents flowing in opposite directions placed close to each other, the radiation fields will cancel each other and will consequently not give a net radiation. For efficient radiation, there has to be spatial imbalance of time varying currents. That is, even if there are currents flowing in opposite directions they should not be physically close to each other. The physical proximity in electromagnetics is always measured in terms of the wavelength. One can, therefore, say that if the two oppositely flowing currents are placed within a distance which is much smaller than the wavelength, the radiation from these currents will be practically negligible. On the other hand, one would get radiation if there is only uni-directional current or two oppositely flowing currents but separated by a distance comparable to the wavelength. In a transmission line, the two conductors carry equal and opposite currents and the separation between them is much smaller compared to the wavelength and hence a transmission line even with open conductors does not show significant radiation. Whereas, when the same transmission line opens up (as shown in Fig. 8.1) the separation between the two currents increases and the structure starts radiating.

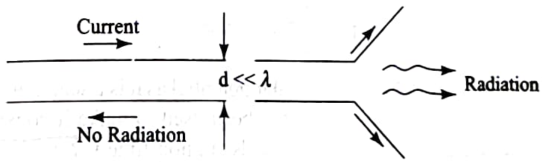


Fig. 8.1 *Radiation from a flared two-conductor transmission line.*

In this chapter, we first develop the basic theory of radiation due to a time varying current and then study the characteristics of linear antennas. Maxwell's equations are to be solved with appropriate sources like charges and currents. However, solving non-homogeneous Maxwell's equations directly for electric and magnetic fields is rather a difficult task. Instead, one can define potential functions related to the electric and magnetic fields and obtain their solutions from the Maxwell's equations. With the knowledge of the potential functions we can then obtain the electric and the magnetic fields.

8.1 POTENTIAL FUNCTIONS

Maxwell's equations for the time varying fields are given as

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon} \tag{8.1}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{8.2}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\mu \frac{\partial \mathbf{H}}{\partial t} \tag{8.3}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J} + \epsilon \frac{\partial \mathbf{H}}{\partial t} \tag{8.4}$$

where $\mathbf{J}$ and ρ are the source current density and source charge density respectively.

Equation (8.2) is identically satisfied if $\mathbf{B}$ is curl of some vector, say $\mathbf{A}$ ($\nabla \cdot \nabla \times \mathbf{A} \equiv 0$ for any $\mathbf{A}$). Then, by definition we have

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{8.5}$$

$$\Rightarrow \qquad \mathbf{H} = \frac{1}{\mu} \nabla \times \mathbf{A} \tag{8.6}$$

$\mathbf{A}$ is called the magnetic vector potential, as it is a vector and is related to the magnetic flux density $\mathbf{B}$. Substituting for $\mathbf{B}$ from Eqn (8.5) to Eqn (8.3) we get

$$\nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0 \tag{8.7}$$

Equation (8.7) can be identically satisfied if the quantity inside the brackets is gradient of some scalar function say V, ($\nabla \times \nabla V \equiv 0$ for all V). Then again by definition, we have

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla V \tag{8.8}$$

The quantity 'V' is called the electric scalar potential as it is a scalar and is related to the electric field. The negative sign has been used to make it consistent with the electrostatic case. That is, when the fields are non-time varying, $\frac{\partial \mathbf{A}}{\partial t} \equiv 0$ and the electric field should be equal to the negative of the gradient of the electrostatic potential. Rearranging Eqn (8.8) we get

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla V \tag{8.9}$$

Through Eqns (8.6) and (8.8) the electric and magnetic fields are expressed in terms of potential functions. Substitution of Eqn. $\mathbf{E}$ and $\mathbf{H}$ from Eqns (8.6) and (8.9) into Maxwell's Eqns (8.1) and (8.4) respectively yields,

$$\nabla \cdot \left(-\frac{\partial \mathbf{A}}{\partial t} - \nabla V \right) = \frac{\rho}{\epsilon} \tag{8.10}$$

$$\Rightarrow \qquad \boxed{\nabla \cdot \left(\frac{\partial \mathbf{A}}{\partial t} \right) + \nabla^2 V = -\frac{\rho}{\epsilon}} \tag{8.11}$$

$$\text{and} \qquad \frac{1}{\mu} \nabla \times \nabla \times \mathbf{A} = \mathbf{J} + \epsilon \frac{\partial}{\partial t} \left(-\frac{\partial \mathbf{A}}{\partial t} - \nabla V \right) \tag{8.12}$$

Using vector identity, $\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$, and rearranging, Eqn (8.12) becomes

$$\boxed{\nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu \mathbf{J} + \mu\epsilon \frac{\partial}{\partial t} \left\{ -\frac{\partial \mathbf{A}}{\partial t} - \nabla V \right\}} \tag{8.13}$$

Equations (8.11) and (8.13) are coupled equations for $\mathbf{A}$ and V, and therefore are quite difficult to solve. Moreover, the equations do not have a unique solution as $\mathbf{A}$ itself is not completely defined. This is due to the Helmholtz theorem, which states that a vector field is uniquely defined if and only if its both divergence and curl are specified. We have defined the curl of $\mathbf{A}$ which is the magnetic flux density (Eqn (8.5)) but the divergence of $\mathbf{A}$ is still undefined. Obviously, depending upon the selection of the divergence of $\mathbf{A}$, we will get different solutions to the equations. At this point, we may re-state the type of behavior we are expecting for the fields and hence for the potentials.

We know from the source-free analysis of the Maxwell's equations that we should get wave type of solution to the fields if we make all sources equal to zero. With this guide line the choice of divergence of $\mathbf{A}$ gets highly restricted. Examination of Eqns (8.11) and (8.13) suggest that if we define the divergence of $\mathbf{A}$ as

$$\boxed{\nabla \cdot \mathbf{A} = -\mu\epsilon \frac{\partial V}{\partial t}} \rightarrow \text{Laurentz gauge cond}^n \tag{8.14}$$

The Eqns (8.11) and (8.13) not only get decoupled but also reduce to the wave equations as

$$\nabla^2 \mathbf{A} - \mu\epsilon \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu \mathbf{J} \tag{8.15}$$

$$\text{and} \qquad \nabla^2 V - \mu\epsilon \frac{\partial^2 V}{\partial t^2} = -\frac{\rho}{\epsilon} \tag{8.16}$$

In the absence of sources like ρ and $\mathbf{J}$, Eqns (8.15), (8.16) reduce to the source-free wave equations discussed in Chapter 4. Also for non-time varying potentials, the Eqns (8.15), (8.16) reduce to static equations like Poisson's equation as

$$\nabla^2 \mathbf{A} = -\mu \mathbf{J} \tag{8.17}$$

$$\nabla^2 V = -\frac{\rho}{\epsilon} \tag{8.18}$$

Poisson's eqn

These arguments, therefore, suggest that the selection of divergence of $\mathbf{A}$ through Eqn (8.14) is indeed correct. Equation (8.14) is known as the *Laurentz gauge condition.*

Few important observations on Eqns (8.15) and (8.16) are in order at this point. Firstly, the electric scalar potential is related to the charges only whereas the magnetic vector potential is related to the current only. Secondly, in Eqn (8.15) since ∇^2 operator is scalar, for a given current density, the direction of the vector potential everywhere in the space is same and it is same as the direction of the current. Later, we will observe that the same behavior is not true for the electric and magnetic fields, and will appreciate the use of potential functions for solving the Maxwell's equations with sources.

8.1.1 Solution for Potential Functions

Before we proceed to solve Eqns (8.15) and (8.16) for **A** and V, we may note that although **A** and V have two different equations, they are not independent. The two are related through the Laurentz gauge condition. It is therefore not required to solve both the Eqns (8.15) and (8.16). We can solve one of the two equations to get solution for either **A** or V and then use the Laurentz gauge condition to find the solution for the other. However, we may note that if we solve Eqn (8.16) for V, obtaining **A** from Eqn (8.14) involves the spatial integration, whereas if we solve for **A** using Eqn (8.15), finding V from Eqn (8.14) involves only a time integration. For time-harmonic potentials (varying as $e^{j\omega t}$) integration with respect to time is much simpler as it is equivalent to multiplying the quantity by $1/j\omega$. We, therefore, generally solve Eqn (8.15) for the magnetic vector potential, **A**.

Assuming that the potentials and fields have a time variation $e^{j\omega t}$, we have

$$\frac{\partial^2}{\partial t^2} \equiv -\omega^2$$

and Eqn (8.15) becomes

$$\nabla^2 \mathbf{A} + \omega^2 \mu \epsilon \mathbf{A} = -\mu \mathbf{J} \tag{8.19}$$

Recognizing that $\omega\sqrt{\mu\epsilon} = \beta$, the propagation constant in the medium (see Chapter 4), Eqn (8.19) yields

$$\nabla^2 \mathbf{A} + \beta^2 \mathbf{A} = -\mu \mathbf{J} \tag{8.20}$$

As mentioned above, the operator ∇^2 is a scalar operator, therefore, Eqn (8.20) is basically a scalar differential equation. The vector nature is only in the variable **A** and the source function **J**. One can, therefore, solve the equation treating **A** as some scalar quantity. After we get the solution of the differential equation, we can multiply the solution by appropriate vector constant.

The approach to solve the equation is called the Green's function technique. The Green's function basically is a spatial impulse response of a system represented by the differential equation. The Green's function G therefore is a solution to the differential Eqn (8.20) with RHS equal to δ-function. The Green's function completely characterizes the system and once the Green's function is known, the potential for any arbitrary excitation can be obtained by simple superposition.

The Greens function, therefore, satisfies equation

$$\nabla^2 G + \beta^2 G = \delta(space) \tag{8.21}$$

G is a scalar quantity.

Now, for the radiation problem the most appropriate coordinate system is the spherical coordinate system, as the wave expands radially outwards from the source in the three dimensional space. We, therefore, obtain here the Green's

function in the spherical coordinate system. In spherical coordinates, Eqn (8.21) becomes,

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial G}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial G}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 G}{\partial\phi^2} + \beta^2 G = \delta(space) \tag{8.22}$$

Without losing generality, we can assume that the impulse excitation is located at the origin, i.e. at $r = 0$, and the problem is symmetric in θ and ϕ. Consequently, the derivatives with respect to θ and ϕ identically go to zero, i.e.

$$\frac{\partial}{\partial\theta} \equiv 0 \qquad \frac{\partial}{\partial\phi} \equiv 0$$

and Eqn (8.22) reduces to

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dG}{dr}\right) + \beta^2 G = \delta(r) \tag{8.23}$$

Note, the partial derivatives with respect to r are changed to full derivatives since G is now a function of r only.

As we know, the impulse response of a differential equation is just the complementary solution with arbitrary constants appropriately evaluated with the initial or boundary conditions. If we now define a temporary variable $\psi = Gr$, Eqn (8.23) yields

$$\frac{d^2\psi}{dr^2} + \beta^2\psi = \delta(r) \tag{8.24}$$

Equation (8.24) is a simple second order differential equation with constant coefficients, and its complementary solution can be written as

$$\psi = Ce^{-j\beta r} + De^{+j\beta r} \tag{8.25}$$

C and D are arbitrary constants to be evaluated.

This solution is similar to that of a transmission line and one can immediately recognize that the two terms on the RHS of Eqn (8.25) represent waves traveling in r direction, one away from $r = 0$ and other towards $r = 0$ respectively. Since, we have assumed the infinite space with only one source at the origin, the wave traveling towards $r = 0$ should be identically zero, i.e. $D \equiv 0$. The solution Eqn (8.25), therefore, yields

$$\psi = Ce^{-j\beta r} \tag{8.26}$$

$$\Rightarrow \qquad G = C\frac{e^{-j\beta r}}{r} \tag{8.27}$$

From Eqn (8.27) we can note that the amplitude of G decreases inversely with the radial distance r. The constant phase surfaces correspond to $r = constant$, therefore, are spherical in shape. Consequently, this wave is called a 'spherical wave'. The phase fronts (constant phase surfaces) for a spherical wave travel radially outwards like a freely expanding balloon as shown in Fig. 8.2.

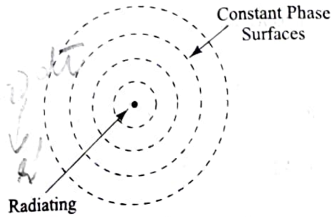


Fig. 8.2 *Spherical wave.*

The arbitrary constant C can be evaluated by substituting Eqn (8.27) into Eqn (8.23) and then integrating both sides over a volume around $r = 0$ with $r \to 0$. We, therefore, have

$$\lim_{r\to 0}\left\{\int_0^r\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\frac{1}{r^2}\frac{d}{dr}\left\{r^2\frac{d}{dr}\left(\frac{Ce^{-j\beta r}}{r}\right)\right\}r^2\sin\theta\, dr\, d\theta\, d\phi\right.$$

$$\left.+\beta^2\int_0^r\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}C\frac{e^{-j\beta r}}{r}r^2\sin\theta\, dr\, d\theta\, d\phi = \int_V \delta(r)dr\right\} \quad (8.28)$$

$$\lim_{r\to 0}\left\{4\pi[-C(j\beta r+1)e^{-j\beta r}] - 4\pi\beta^2\int_0^r e^{-j\beta r}r\,dr\right\} = 1 \quad (8.29)$$

In the second term of Eqn (8.29), $e^{-j\beta r} \to 1$ as $r \to 0$ and consequently the integral $\to 0$ as $r \to 0$. In the first term of Eqn (8.29) the term containing r goes to zero as $r \to 0$. The integration of δ-function on the right hand side is 1 by the definition of the function. Equation (8.29) hence yields the arbitrary constant as

$$C = -\frac{1}{4\pi} \quad (8.30)$$

Thus the Green's function is

$$G = -\frac{e^{-j\beta r}}{4\pi r} \quad (8.31)$$

Since, the medium surrounding the source is linear, the response to any arbitrary input can be obtained by convolution of the impulse response (Green's function) and the input source function, (-current density multiplied by μ) i.e. the vector potential can be obtained as

$$\mathbf{A} = \int_V \mu\mathbf{J}(\mathbf{r}')\frac{e^{-j\beta|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}dv' \quad (8.32)$$

The prime coordinates denote the source space and the unprimed coordinates denote the observation space as shown in Fig 8.3.

Note that, in normal convolution we have $(\mathbf{r} - \mathbf{r}')$. The modulus sign in this case is to make sure that the quantity $(\mathbf{r} - \mathbf{r}')$ is positive. This is due to the fact that the radial distance in spherical coordinates is always positive.

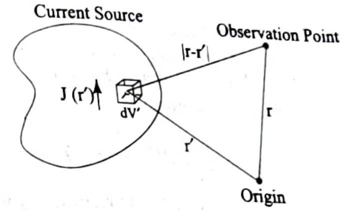


Fig. 8.3 *Source space and observation space.*

From Eqn (8.32) it is clear that once the spatial current distribution $\mathbf{J}(\mathbf{r}')$ is known, the magnetic vector potential and subsequently the electric and magnetic fields can be obtained through a straight forward algebra. How a particular current distribution is to be realized is a rather complex issue and is beyond the scope of the discussion here.

8.2 RADIATION FROM THE HERTZ DIPOLE

An infinitesimally small current element is called the Hertz dipole. Although an infinitesimally small current element is not of much practical importance, it forms the basis of any complex radiating structure. After all any radiating structure can always be thought of as a collection of small current elements. The analysis of a Hertz dipole, therefore, is an important subject in the antenna theory.

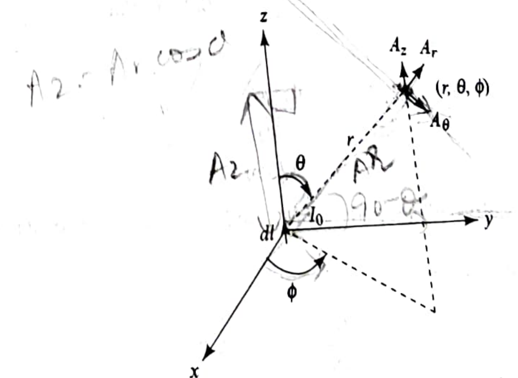


Fig. 8.4 *Hertz dipole located at the origin and oriented along the z-axis.*

Let us consider an infinitesimal current element of length dl carrying an alternating current I_0. Without losing generality, let us assume that the current element is located at the origin and is oriented along the z axis as shown in Fig. 8.4. The instantaneous current in the dipole is given as

$$I(t) = I_0 e^{j\omega t}\hat{\mathbf{z}} \quad (8.3$$

where ω is the angular frequency of the current.

Since, the current is along the z direction, the magnetic vector potential is also along the z direction every where in the space. From Eqn (8.32) the magnetic vector potential at some point P can be given as

$$\boxed{\mathbf{A} = A_z\hat{\mathbf{z}} = \frac{\mu}{4\pi} I_0 dl \frac{e^{-j\beta r}}{r} e^{j\omega t}\hat{\mathbf{z}}} \tag{8.34}$$

Note that in Eqn (8.32) the integral of the current density over the source volume is $I_0 dl$. I_0 directly gives the integral of the current density over the area of cross section of the current element.

EXAMPLE 8.1 A 0.1 m long thin wire is carrying 10 A peak current at 30 MHz, and is oriented along the z-axis. Find the magnetic vector potential at a distance of (i) 1 m (ii) 10 m (iii) 100 m from the wire.

Solution:

From Eqn (8.34) the magnetic vector potential due to z-oriented current element is

$$\mathbf{A} = A_z\hat{\mathbf{z}} = \frac{\mu}{4\pi} I_0 dl \frac{e^{-j\beta r}}{r} e^{j\omega t}\hat{\mathbf{z}}$$

$$\omega = 2\pi \times 30 \times 10^6 = 6\pi \times 10^7 \text{ rad/s}$$

For the free space, $\mu = \mu_0 = 4\pi \times 10^{-7}$ H/m

$$\beta = \frac{\omega}{c} = \frac{6\pi \times 10^7}{3 \times 10^8} = 0.2\pi \text{ rad/m}$$

(i) Vector potential at $r = 1$m

$$A_z = \frac{4\pi \times 10^{-7}}{4\pi} \times \frac{10 \times 0.1}{1.0} \times e^{-j0.2\pi} e^{j\omega t}$$

$$= 10^{-7} e^{-j0.2\pi} e^{j\omega t} \text{ Wb/m}$$

Since $|A_z|$ is inversally proportional to r, and phase is proportional to r, we get

(ii) Vector potential at $r = 10$ m

$$A_z = \frac{10^{-7}}{10} e^{-j0.2\pi \times 10} e^{j\omega t}$$

$$= 10^{-8} e^{-j2\pi} e^{j\omega t} \text{ Wb/m}$$

(iii) Vector potential at $r = 100$ m

$$A_z = 10^{-9} e^{-j20\pi} e^{j\omega t} \text{ Wb/m}$$

EXAMPLE 8.2 If the current element in Example 8.1 is oriented along the y-direction, and is located at the origin, find the components of the magnetic vector potential at point (10, 20, 30)m. Also find the components in the spherical co-ordinate system.

Solution:

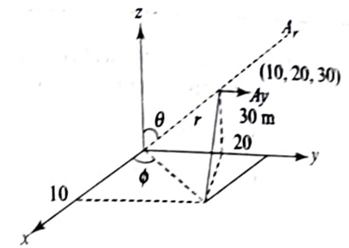


Fig. 8.5

Radial distance

$$r = \sqrt{(10)^2 + (20)^2 + (30)^2} = 10\sqrt{1+4+9} = 10\sqrt{14} \text{ m}$$

The vector potential everywhere is oriented in y-direction since the current element is oriented along the y-direction.

$$\mathbf{A} = A_y\hat{\mathbf{y}} = \frac{4\pi \times 10^{-7}}{4\pi} \times \frac{10 \times 0.1 \times e^{-j0.2\pi \times 10\sqrt{14}}}{10\sqrt{14}} \cdot e^{j\omega t}\hat{\mathbf{y}} \text{ Wb/m}$$

$$A_y = \frac{10^{-8}}{\sqrt{14}} e^{-j2\pi\sqrt{14}} \cdot e^{j\omega t} \text{ Wb/m}$$

From Fig. 8.5 we find that

$$\cos\phi = \frac{1}{\sqrt{5}}, \qquad \sin\phi = \frac{2}{\sqrt{5}}$$

$$\cos\theta = \frac{30}{10\sqrt{14}} = \frac{3}{\sqrt{14}}$$

$$\sin\theta = \sqrt{1 - \frac{9}{14}} = \sqrt{\frac{5}{14}}$$

Using relations in Appendix C, we get

$$A_r = A_y \sin\theta \sin\phi = \frac{10^{-8}}{\sqrt{14}} \cdot \sqrt{\frac{5}{14}} \cdot \frac{2}{\sqrt{5}} e^{-j2\pi\sqrt{14}} e^{j\omega t} \text{ Wb/m}$$

$$= \frac{10^{-8}}{7} e^{-j2\pi\sqrt{14}} e^{j\omega t} \text{ Wb/m}$$

$$A_\theta = A_y \cos\theta \sin\phi = \frac{10^{-8}}{\sqrt{14}} \frac{3}{\sqrt{14}} \cdot \frac{2}{\sqrt{5}} e^{-j2\pi\sqrt{14}} e^{j\omega t} \text{ Wb/m}$$

$$= \frac{3\times 10^{-8}}{7\sqrt{5}} e^{-j2\pi\sqrt{14}} e^{j\omega t} \text{ Wb/m}$$

$$A_\phi = A_y \cos\phi = \frac{10^{-8}}{\sqrt{14}} \cdot \frac{1}{\sqrt{5}} . e^{-j2\pi\sqrt{14}} e^{j\omega t} \text{ Wb/m}$$

Since, the radiation is investigated in the spherical coordinate system first we have to convert **A** in (r, θ, ϕ) components. From Fig. 8.4 we get

$$A_r = A_z \cos\theta \tag{8.35}$$

$$A_\theta = -A_z \sin\theta \tag{8.36}$$

$$A_\phi = 0 \tag{8.37}$$

Substitution of **A** in (8.6) yields the magnetic field as

$$\mathbf{H} = \frac{1}{\mu} \nabla \times \mathbf{A} \tag{8.38}$$

$$= \frac{1}{\mu} \frac{1}{r^2 \sin\theta} \begin{vmatrix} \hat{\mathbf{r}} & r\hat{\theta} & r\sin\theta\hat{\phi} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi} \\ A_r & rA_\theta & r\sin\theta A_\phi \end{vmatrix} \tag{8.39}$$

For the Hertz dipole we have symmetry in ϕ giving $\frac{\partial}{\partial\phi} \equiv 0$. Substitution of Eqns (8.35)–(8.37) in Eqn (8.39) yields

$$\mathbf{H} = \frac{1}{\mu r^2 \sin\theta} \begin{vmatrix} \hat{\mathbf{r}} & r\hat{\theta} & r\sin\theta\hat{\phi} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & 0 \\ A_z\cos\theta & -rA_z\sin\theta & 0 \end{vmatrix} \tag{8.40}$$

$$\Rightarrow \quad H_r = 0 \tag{8.41}$$

$$H_\theta = 0 \tag{8.42}$$

$$H_\phi = \frac{1}{\mu}\frac{1}{r}\left\{\frac{\partial}{\partial r}(-rA_z\sin\theta) - \frac{\partial}{\partial\theta}(A_z\cos\theta)\right\} \tag{8.43}$$

Substitution of A_z from Eqn (8.34) in Eqn (8.43) yields

$$H_\phi = -\frac{I_0 dl e^{j\omega t}}{4\pi r}\left\{\frac{\partial}{\partial r}(e^{-j\beta r}\sin\theta) + \frac{\partial}{d\theta}(\frac{e^{-j\beta r}\cos\theta}{r})\right\} \tag{8.44}$$

$$= \frac{I_0 dl e^{j\omega t}}{4\pi r}\left\{ j\beta e^{-j\beta r}\sin\theta + \frac{e^{-j\beta r}}{r}\sin\theta\right\} \tag{8.45}$$

$$H_\phi = \frac{I_0 dl e^{j\omega t}\sin\theta}{4\pi r} e^{-j\beta r}\left\{ j\beta + \frac{1}{r}\right\} \tag{8.46}$$

The Hertz dipole has only ϕ component of the magnetic field, i.e. the magnetic field circulates around the dipole.

The electric field at point P can be obtained by substituting H into the source free maxwell's equation for the free space as

$$E = \frac{1}{j\omega\epsilon} \nabla \times \mathbf{H} \tag{8.47}$$

$$= \frac{1}{j\omega\epsilon}\frac{1}{r^2\sin\theta} \begin{vmatrix} \hat{\mathbf{r}} & r\hat{\theta} & r\sin\theta\hat{\phi} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & 0 \\ 0 & 0 & H_\phi r\sin\theta \end{vmatrix} \tag{8.48}$$

Note, $\frac{\partial}{\partial\phi} \equiv 0$ and H_r and H_θ are also zero for the Hertz dipole. We hence get

$$E_r = \frac{1}{j\omega\epsilon}\frac{1}{r^2\sin\theta}\frac{\partial(H_\phi r\sin\theta)}{\partial\theta}$$

$$= \frac{1}{j\omega\epsilon}\frac{2I_0 dl e^{j\omega t}\cos\theta e^{-j\beta r}}{4\pi r^2}\left\{j\beta + \frac{1}{r}\right\}$$

$$\boxed{E_r = \frac{I_0 dl e^{j\omega t}\cos\theta e^{-j\beta r}}{4\pi\omega\epsilon}\left\{\frac{\beta}{r^2} - \frac{j}{r^3}\right\}} \tag{8.49}$$

and

$$E_\theta = -\frac{1}{j\omega\epsilon}\frac{1}{r\sin\theta}\frac{\partial}{\partial r}(r\sin\theta H_\phi)$$

$$= \frac{-1}{j\omega\epsilon}\frac{1}{r}\frac{\partial}{\partial r}(rH_\phi)$$

$$\boxed{E_\theta = \frac{I_0 dl\sin\theta e^{j\omega t}e^{-j\beta r}}{4\pi\epsilon}\left\{\frac{j\beta^2}{\omega r} + \frac{\beta}{\omega r^2} - \frac{j}{\omega r^3}\right\}} \tag{8.50}$$

From Eqns (8.46), (8.49) and (8.50) we see that the electric field E lies in the (r, θ) plane whereas the magnetic field H lies in the ϕ plane. That is to say that the electric and magnetic fields are perpendicular to each other at every point in space. One can make many detail observations of Eqns (8.46), (8.49) and (8.50) and some of them are very important.

1. Firstly, if we look at the expressions of the electric and magnetic fields, we note that the fields can be classified in three catagories on the basis of their variation as a function of distance. The three fields have spatial variations as $\frac{1}{r}$ (radiation field), $\frac{1}{r^2}$ (induction field), and $\frac{1}{r^3}$ (electrostatic field). Noting that $\beta = \omega\sqrt{\mu\epsilon}$ i.e. β is proportional to ω, we observe that the magnitude of the field which varies as $\frac{1}{r^3}$ is inversely proportional to the frequency. The field which varies as $\frac{1}{r^2}$ is independent of frequency, and the

field which varies as $\frac{1}{r}$ is proportional to the frequency. The $\frac{1}{r^2}$ and $\frac{1}{r^3}$ fields are dominant only for small values of r whereas $\frac{1}{r}$ fields are dominant for large values of r. It is then interesting to note that at low frequencies, the fields are more localized around the sources whereas, at high frequencies the fields reach out at far away distances from the source. In other words, for same current I_0 and same distance r, the $\frac{1}{r}$ field becomes stronger as the frequency increases.

This is the main cause of interference at high frequencies. The widely separated electronic components, which do not have any coupling at low frequency, start showing coupling as the frequency increases. Obviously, the electronic circuit performance gets altered due to coupling at high frequencies.

The three types of fields are plotted in Fig. 8.6. From Fig. 8.6, we can see that the three fields become equal is magnitude when $\frac{\beta^2}{r} = \frac{\beta}{r^2} = \frac{1}{r^3}$, i.e. when $r = \frac{1}{\beta}$. Since, $\beta = \frac{2\pi}{\lambda}$, the three fields become equal in magnitude at a distance of approximately $\frac{\lambda}{6}$ (since $2\pi \approx 6$) from the source. For a distance $<< \frac{\lambda}{6}$, the electrostatic field ($\frac{1}{r^3}$) dominates whereas for a distance $>> \frac{\lambda}{6}$ the radiation field dominates.

2. Secondly, we note that the near-field is only the electric field. There is, no magnetic field component which varies as $\frac{1}{r^3}$. Since, the electric field is primarily related to the charges (Gauss Law) and the magnetic field is primarily related to the currents (Biot-Savart Law), it is worthwhile to investigate why there is electric field with $\frac{1}{r^3}$ variation and not the magnetic field.

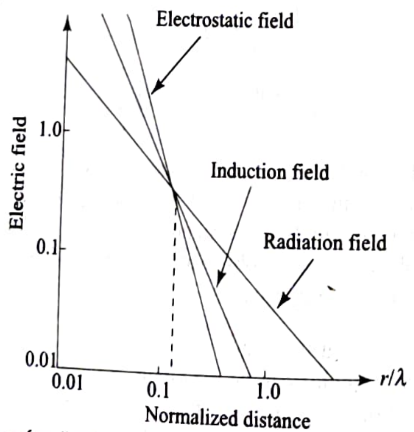


Fig. 8.6 *Variation of radiation, induction and electrostatic fields as a function of distance from an antenna.*

At some instant of time, let the current in the Hertz dipole flow from bottom to top as shown in Fig. 8.7. That is the electrons flow from top to bottom. Since, the two ends of the dipole are not connected to any circuit, a negative charge will accumulate at the lower end of the dipole (point B), whereas a positive charge will accumulate at the upper end of the dipole (point A). The two charges will be equal and opposite at every instant of time. Let at some instant of time the charges be $+Q$ and $-Q$. We then have two charges separated by a distance dl. The total electrostatic field at some point P at a distance, r, from the center of the dipole can be obtained by vector addition of the fields due to two charges. For simplicity, if we assume that the point P is equi-distant from A and B, the total electrostatic field can be written as

$$\mathbf{E} = \mathbf{E}^+ + \mathbf{E}^- \tag{8.51}$$

$$= \frac{Q\cos\theta}{4\pi R^2\epsilon} + \frac{Q\cos\theta}{4\pi R^2\epsilon} = \frac{2Q\cos\theta}{4\pi\epsilon R^2} \tag{8.52}$$

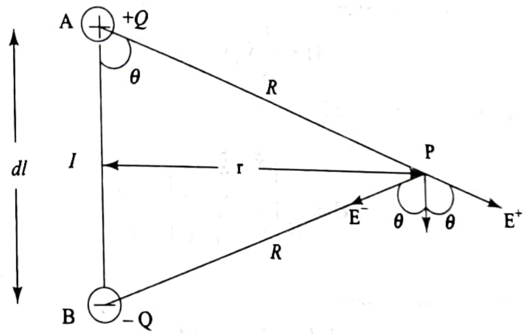


Fig. 8.7 *Electrostatic field due to a Hertz dipole.*

From Fig. 8.7, $\cos\theta = \frac{dl/2}{R}$, and $R^2 = (\frac{dl}{2})^2 + r^2$. Assuming that $r >> dl$, i.e. $R^2 \approx r^2$, Eqn (8.52) yields

$$E \approx \left\{\frac{2Q(\frac{dl}{2})}{4\pi\epsilon r^3}\right\} = \left\{\frac{Q dl}{4\pi\epsilon r^3}\right\} \tag{8.53}$$

This field varies as $\frac{1}{r^3}$.

It then clearly indicates that the $\frac{1}{r^3}$ field is due to the accumulated charges at the tips of the dipole. For any alternating current the charges oscillate between A and B as the direction of the current reverses every half period of the current. However, one thing is clear that for a given value of I, as the frequency reduces there is more accumulation of charge and consequently higher field. For a DC current (if DC current keeps flowing) the charges will build up to infinite value as time $\rightarrow \infty$ (steady state) and consequently the field will also be infinite. The inverse proportionality of the field with the frequency is, therefore, quite evident in this case.

The magnetic field is due to the current and not due to accumulated charges and consequently it does not have a variation of $\frac{1}{r^3}$.

EXAMPLE 8.3 Find the magnetic field intensity in spherical coordinate system due to the current element in Example 8.2. Derive the expressions in two ways:

(i) Using the curl operator in the Cartesian coordinate system and then transforming the H-vector to the spherical coordinate system.

(ii) Using A_r, A_θ, A_ϕ components and curl operator in spherical coordinate system.

Solution:

The magnetic field is given as

$$\mathbf{H} = \frac{1}{\mu}\nabla \times \mathbf{A}$$

(i) In Cartesian system, we have

$$\mathbf{H} = \frac{1}{\mu}\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ 0 & A_y & 0 \end{vmatrix}$$

since $A_x = A_z = 0$ in this case.

$$\Rightarrow \qquad \mathbf{H} = \frac{1}{\mu}\left\{-\frac{\partial A_y}{\partial z}\hat{\mathbf{x}} + \frac{\partial A_y}{\partial x}\hat{\mathbf{z}}\right\}$$

Now $$A_y = \frac{\mu I_0 dl}{4\pi r}e^{-j\beta r}e^{j\omega t}$$

$$\frac{\partial A_y}{\partial x} = \frac{\mu I_0 dl}{4\pi}e^{j\omega t}\left\{\frac{r(-j\beta)e^{-j\beta r} - e^{-j\beta r}}{r^2}\right\}\frac{dr}{dx}$$

since $$r = \sqrt{x^2 + y^2 + z^2}$$

$$\frac{dr}{dx} = \frac{x}{r} = \frac{r\sin\theta\cos\phi}{r} = \sin\theta\cos\phi$$

similarly $$\frac{dr}{dz} = \frac{z}{r} = \cos\theta$$

$$\Rightarrow \qquad \frac{\partial A_y}{\partial x} = \frac{\mu I_0 dl}{4\pi}\left\{\frac{-j\beta r - 1}{r^2}\right\}e^{-j\beta r}\sin\theta\cos\phi e^{j\omega t}$$

and $$\frac{\partial A_y}{\partial z} = \frac{\mu I_0 dl}{4\pi}\left\{\frac{-j\beta r - 1}{r^2}\right\}e^{-j\beta r}\cos\theta e^{j\omega t}$$

$$\mathbf{H} = \frac{-I_0 dl}{4\pi}\left\{\frac{j\beta r + 1}{r^2}\right\}e^{-j\beta r}e^{j\omega t}\{-\cos\theta\hat{\mathbf{x}} + \sin\theta\cos\phi\hat{\mathbf{z}}\}$$

$$\Rightarrow \qquad H_x = \frac{I_0 dl\cos\theta}{4\pi}\left\{\frac{j\beta r + 1}{r^2}\right\}e^{-j\beta r}e^{j\omega t}\text{ Amp/m}$$

$$H_y = 0$$

$$H_z = \frac{-I_0 dl\sin\theta\cos\phi}{4\pi}\left\{\frac{j\beta r + 1}{r^2}\right\}e^{-j\beta r}e^{j\omega t}\text{ Amp/m}$$

Now from Appendix C we have,

$$\begin{bmatrix} H_r \\ H_\theta \\ H_\phi \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{bmatrix}\begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix}$$

$$H_r = H_x\sin\theta\cos\phi + H_z\cos\theta$$

$$= \frac{I_0 dl}{4\pi}\left\{\frac{j\beta r + 1}{r^2}\right\}e^{-j\beta r}e^{j\omega t}\{\sin\theta\cos\theta\cos\phi - \sin\theta\cos\theta\cos\phi\}$$

$$= 0$$

$$H_\theta = H_x\cos\theta\cos\phi - H_z\sin\theta$$

$$= \frac{I_0 dl}{4\pi}\left\{\frac{j\beta r + 1}{r^2}\right\}e^{-j\beta r}e^{j\omega t}\{\cos^2\cos\phi + \sin^2\theta cos\phi\}$$

$$= \frac{I_0 dl}{4\pi}\left\{\frac{j\beta r + 1}{r^2}\right\}e^{-j\beta r}e^{j\omega t}\cos\phi$$

$$H_\phi = -H_x\sin\phi$$

$$= -\frac{I_0 dl}{4\pi}\left\{\frac{j\beta r + 1}{r^2}\right\}e^{-j\beta r}e^{j\omega t}\cos\theta\sin\phi$$

(ii) Knowing A_y we can find A_r, A_θ, A_ϕ as

$$A_r = A_y\sin\theta\sin\phi = \frac{\mu I_0 dl}{4\pi r}e^{-j\beta r}e^{j\omega t}\cdot\sin\theta\sin\phi$$

$$A_\theta = A_y\cos\theta\sin\phi = \frac{\mu I_0 dl}{4\pi r}e^{-j\beta r}e^{j\omega t}\cos\theta\sin\phi$$

$$A_\phi = A_y\cos\phi = \frac{\mu I_0 dl}{4\pi r}e^{-j\beta r}e^{j\omega t}\cos\phi$$

We, therefore, have the magnetic field as

$$\mathbf{H} = \frac{1}{\mu}\nabla \times A = \frac{1}{r^2\sin\theta}\begin{vmatrix} \hat{\mathbf{r}} & r\hat{\theta} & r\sin\theta\hat{\phi} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi} \\ A_r & rA_\theta & r\sin\theta A_\phi \end{vmatrix}$$

$$\Rightarrow \quad H_r = \frac{1}{r^2 \sin\theta}\left\{\frac{\partial}{\partial\theta}(r\sin\theta A\phi) - \frac{\partial}{\partial\phi}(rA_\theta)\right\}$$

$$= \frac{I_0 dl e^{-j\beta r}}{4\pi r^3 \sin\theta} e^{j\omega t}\{r\cos\theta\cos\phi - r\cos\theta\cos\phi\} = 0$$

$$H_\theta = \frac{1}{r\sin\theta}\left\{\frac{\partial A_r}{\partial\phi} - \frac{\partial}{\partial r}(r\sin\theta A_\phi)\right\}$$

$$= \frac{I_0 dl}{4\pi r\sin\theta}\left\{\frac{\sin\theta\cos\phi e^{-j\beta r}}{r} - \frac{\partial}{\partial r}(e^{-j\beta r})\sin\theta\cos\phi\right\}e^{j\omega t}$$

$$= \frac{I_0 dl}{4\pi r^2}\cos\phi(1+j\beta r)e^{-j\beta r}e^{j\omega t}$$

$$H_\phi = \frac{I_0 dl}{4\pi r}\left\{\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial}{\partial\theta}(A_r)\right\}$$

$$= \frac{I_0 dl}{4\pi r^2}\{-j\beta r\cos\theta\sin\phi - \cos\theta\sin\phi\}e^{-j\beta r}e^{j\omega t}$$

$$= -\frac{I_0 dl}{4\pi r^2}(1+j\beta r)\cos\theta\sin\phi e^{-j\beta r}e^{j\omega t}$$

EXAMPLE 8.4 Find the near and far electric fields for the y-oriented dipole in Example 8.2. If the peak current is 1 Amp and the frequency of the current is 100 MHz, find the instantaneous electric field at 1 m away along the 45° line in the yz-plane at $t = 1\ \mu$s.

Solution:

Here $\theta = 45^\circ$ and $\phi = 90^\circ$ (yz-plane)

The angular frequency $\omega = 2\pi f = 2\pi \times 10^8$ rad/s

The phase constant $\quad \beta = \frac{\omega}{c} = \frac{2\pi \times 10^8}{3\times 10^8} = \frac{2\pi}{3}$ rad/m

The electric field is $\quad \mathbf{E} = \frac{1}{j\omega\epsilon}\nabla\times\mathbf{H}$

From Example 8.2, we have $H_r = 0$. Therefore we get

$$\mathbf{E} = \frac{1}{j\omega\epsilon}\frac{1}{r^2\sin\theta}\begin{vmatrix}\hat{r} & r\hat{\theta} & r\sin\theta\hat{\phi} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi} \\ 0 & rH_\theta & r\sin\theta H_\phi\end{vmatrix}$$

$$\mathbf{H} = K\left\{\frac{e^{-j\beta r}}{r^2}\cos\phi\hat{\theta} - \frac{e^{-j\beta r}}{r^2}\cos\theta\sin\phi\hat{\phi}\right\}\{1+j\beta r\}$$

where $\quad K \equiv \frac{I_0 dl}{4\pi}e^{j\omega t}$

$$E_r = \frac{1}{j\omega\epsilon r^2\sin\theta}\left\{\frac{\partial}{\partial\theta}(r\sin\theta H_\phi) - \frac{\partial}{\partial\phi}(rH_\theta)\right\}$$

$$= \frac{(1+j\beta r)\sin\phi\sin\theta e^{-j\beta r}e^{j\omega t}I_0 dl}{2\pi j\omega\epsilon r^3}$$

$$= -0.0043 - 0.00301j \text{ V/m}$$

Hence, the value of electrical field is -0.0043 V/m at $t = 1\ \mu$s. Here, the negative sign shows that the direction is reversed, i.e. radially inwards.

3. Thirdly, we find that the near field which is dominated by the electrostatic field is far more complex than the radiation field. The radiation field is also referred as the far-field.

Near Field Much within the distance of $\lambda/6$ from the current element, the dominant field is the $1/r^3$ field. The electric field, therefore, is given by

$$E_r \approx \frac{-j2I_0 dl\cos\theta e^{j\omega t}}{4\pi\epsilon\omega r^3}e^{-j\beta r} \tag{8.54}$$

$$\approx \frac{-j2I_0 dl\cos\theta}{4\pi\epsilon\omega r^3}e^{j\omega t} \tag{8.55}$$

and

$$E_\theta \approx \frac{-jI_0 dl\sin\theta e^{j\omega t}}{4\pi\omega\epsilon r^3}e^{-j\beta r} \tag{8.56}$$

$$\approx \frac{-jI_0 dl\sin\theta}{4\pi\omega\epsilon r^3}e^{j\omega t} \tag{8.57}$$

Here, we have assumed that for $r << \frac{\lambda}{6}$, $e^{-j\beta r} \approx 1$. The magnitude of the electric field in the near field region therefore is given as

$$|E| = \sqrt{|E_r^2| + |E_\theta^2|} = \frac{I_0 dl}{4\pi\epsilon\omega r^3}\sqrt{4\cos^2\theta + \sin^2\theta} \tag{8.58}$$

$$= \frac{I_0 dl}{4\pi\omega\epsilon r^3}\sqrt{1+3\cos^2\theta} \tag{8.59}$$

A polar plot of the near field is shown in Fig. 8.8. The field is maximum along $\theta = 0$ and $\theta = \pi$ direction, and it is minimum along $\theta = \frac{\pi}{2}$ direction. However, it is important to note that in no direction is the field zero.

Far Field In the far field zone, that is for $r >> \frac{\lambda}{6}$, the field is dominated by the radiation field. The electric and magnetic fields therefore are given as

$$E_\theta = \frac{jI_0 dl\beta^2\sin\theta e^{-j\beta r}e^{j\omega t}}{4\pi\epsilon\omega r} \tag{8.60}$$

and

$$H_\phi = \frac{jI_0 dl\sin\theta\beta e^{-j\beta r}e^{j\omega t}}{4\pi r} \tag{8.61}$$

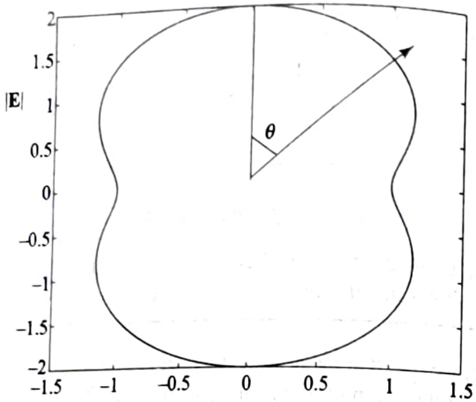


Fig. 8.8 *Angular variation of electrical field due to the Hertz dipole for $r << \lambda/6$.*

The electric and magnetic fields are in phase with each other and they are 90° out of phase with respect to the current. The phase difference between the current and the field is expected, because, as mentioned earlier, the radiation is related to the acceleration of charges, i.e. to the rate of change of current.

The constant phase surfaces are given by βr = constant and therefore are spherical in shape. The wave travels in $+r$ direction away from the current element. This wave is called a spherical wave.

It is interesting to note that the ratio of the electric and the magnetic field is constant, i.e.

$$\frac{E_\theta}{H_\phi} = \frac{\beta}{\omega\epsilon} = \frac{\omega\sqrt{\mu\epsilon}}{\omega\epsilon} = \sqrt{\frac{\mu}{\epsilon}}$$

$$= \eta \text{ (intrinsic impedance of the medium)} \tag{8.62}$$

We, hence, conclude that an oscillating current element generates a spherical wave travelling radially outwards and the direction of the electric field, magnetic field and the wave vector are perpendicular to each other at every point in space (see Fig. 8.9). The spherical wave therefore is a transverse electromagnetic wave. It should be noted that the direction of the E field (and consequently H-field) is different at different locations but the TEM nature of the wave holds true at every point.

Both E and H fields are zero along $\theta = 0$ and they are maximum along $\theta = \frac{\pi}{2}$. A current element therefore radiates in a direction perpendicular to the current direction and no radiation goes along the direction of the current.

4. Fourthly, the power flow in the medium is solely due to the radiation field. This can be shown by computing the average Poynting vector at some point in space. The average Poynting vector

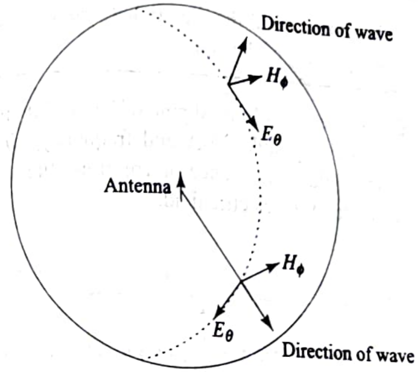


Fig. 8.9 *Spherical wave direction with respect to the electric and the magnetic fields due to an antenna.*

$$\mathbf{P}_{av} = \frac{1}{2}Re(\mathbf{E} \times \mathbf{H}^*) \tag{8.63}$$

$$= \frac{1}{2}Re(E_\theta H_\phi^* \hat{\mathbf{r}} - E_r H_\phi^* \hat{\theta})$$

From Eqns (8.46), (8.49) and (8.50) we note that E_r and H_ϕ^* are 90° out of phase and their product is purely imaginary. The Poynting vector, therefore, reduces to

$$\mathbf{P}_{av} = \frac{1}{2}Re\left\{E_\theta H_\phi^*\right\}\hat{\mathbf{r}} \tag{8.64}$$

$$= \frac{1}{2}Re\left\{\left(\frac{I_0 dl \sin\theta}{4\pi r}\right)^2 \frac{1}{\omega\epsilon}\left(j\beta^2 + \frac{\beta}{r} - \frac{j}{r^2}\right)\left(-j\beta + \frac{1}{r}\right)\right\}\hat{\mathbf{r}} \tag{8.65}$$

$$= \frac{1}{2}\left(\frac{I_0 dl \sin\theta}{4\pi r}\right)^2 \frac{\beta^3}{\omega\epsilon}\hat{\mathbf{r}} \tag{8.66}$$

In Eqn (8.66) we note that the net real product is only due to the radiation terms (i.e. $-j\beta^2$ and $-j\beta$). The other fields give net imaginary product only and therefore do not contribute to an average power flow. It is, therefore, apparent, that if at all there is a power flow it should be only through radiation. Since, the electrostatic and the induction fields contribute to only reactive power, they are also called the reactive fields.

Having this general understanding of the various fields created by the Hertz dipole, now we can focus our attention on only the radiation fields.

Generally, in a practical radiating system, one is interested in two things,

(i) the total power radiated by the antenna,

(ii) the directional dependence of the field.

The total radiated power can be obtained by integrating the Poynting vector over a sphere enclosing the antenna and the directional dependence of the fields is given by the 'Radiation pattern' of the antenna.

EXAMPLE 8.5 A z-oriented Hertz dipole of length 10 cm is excited with a sinusoidal current of amplitude 20 A and frequency 10 MHz. Find the instantaneous electric field at a distance of 1m along the x- axis at 1 μsec. Also find the orientation of the electric field.

Solution:

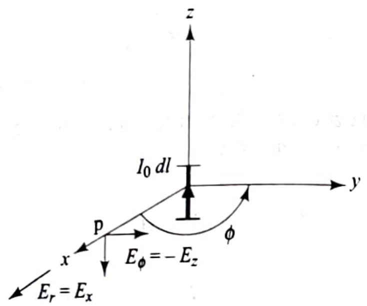


Fig. 8.10

The frequency of the current is 10 MHz.

$$\Rightarrow \qquad \omega = 2\pi \times 10^7 \text{ rad/s.}$$

$$\text{The phase constant } \beta = \frac{\omega}{c} = \frac{2\pi \times 10^7}{3 \times 10^8} = \frac{\pi}{15} \text{ rad/m}$$

Since, the observation point is along the x-axis, $r = x = 1$ m, as shown in Fig. 8.10

$$\Rightarrow \qquad \beta r = \frac{\pi}{15} \text{ rad}$$

$$\text{And} \qquad \omega t = 2\pi \times 10^7 \times 10^{-6} = 20\pi$$

$$E_r = E_x = Re\left\{\frac{I_0 dl \cos\theta}{4\pi\omega\epsilon}\left[\frac{\beta}{r^2} - \frac{j}{r^3}\right]e^{j\omega t - j\beta r}\right\}$$

$$= \frac{I_0 dl \cos\theta}{4\pi\omega\epsilon}\left\{\frac{\beta}{r^2}\cos(\omega t - \beta r) + \frac{1}{r^3}\sin(\omega t - \beta r)\right\}$$

Since, the observation point is on the x-axis, $\theta = \frac{\pi}{2}$.

$$\Rightarrow \qquad \cos\theta = 0 \qquad \sin\theta = 1$$

$$\Rightarrow \qquad E_r = E_x = 0$$

The θ component can be obtained from Eqn (8.50) as

$$-E_z = E_\theta = \frac{I_0 dl}{4\pi\epsilon} Re\left\{e^{j\omega t - j\beta r}\left[\frac{j\beta^2}{\omega r} + \frac{\beta}{\omega r^2} - \frac{j}{\omega r^3}\right]\right\}$$

$$= \frac{I_0 dl}{4\pi\epsilon}\left\{-\left(\frac{\beta^2}{\omega r} - \frac{1}{\omega r^3}\right)\sin(\omega t - \beta r) + \frac{\beta}{\omega r^2}\cos(\omega t - \beta r)\right\}$$

$$= 0.1932 \text{ V/m}$$

and $\qquad E_\phi = E_y = 0$

The electric field is therefore oriented in the -z direction.

8.2.1 Total Power Radiated by the Hertz Dipole

The total power radiated by the antenna is obtained by integrating the Poynting vector over a spherical surface of say radius r as shown in Fig. 8.11. The elemental surface area is given as $r^2 \sin\theta d\theta d\phi$ and the total power can be written as

$$W = \int\int \mathbf{P}_{\mathbf{av}} r^2 \sin\theta d\theta d\phi$$

$$= \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\frac{1}{2}\left(\frac{I_0 dl \sin\theta}{4\pi r}\right)^2\frac{\beta^3}{\omega\epsilon}r^2\sin\theta d\theta d\phi \qquad (8.67)$$

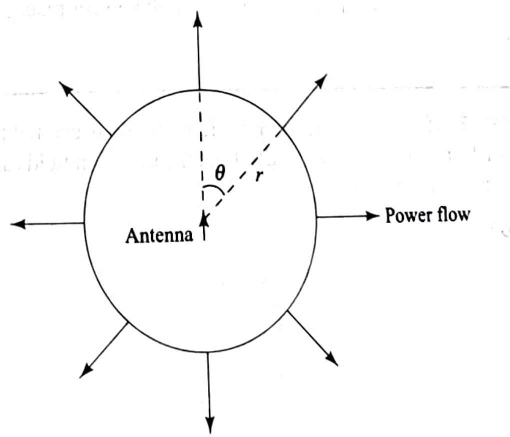


Fig. 8.11 *Power flow from an antenna.*

Note that the integral is no more a function of r as the total power is independent of the radius of the sphere chosen for integration. Also the Poynting vector is not a function of ϕ giving

$$W = \frac{1}{2}\left(\frac{I_0 dl}{4\pi}\right)^2\frac{\beta^3}{\omega\epsilon}\int_0^{2\pi}d\phi\int_0^{\pi}\sin^3\theta d\theta \qquad (8.68)$$

$$= \frac{1}{2}\left(\frac{I_0 dl}{4\pi}\right)^2 \frac{\beta^3}{\omega\epsilon}.2\pi \int_0^{\pi} (1-\cos^2\theta)\sin\theta d\theta \tag{8.69}$$

(The integral can be solved by substituting $\cos\theta = t$.) We therefore get

$$W = \pi \frac{I_0^2 dl^2}{16\pi^2}\frac{\beta^3}{\omega\epsilon}.\frac{4}{3} \tag{8.70}$$

Using relation $\beta = \omega\sqrt{\mu\epsilon}$, we get

$$\frac{\beta^3}{\epsilon\omega} = \frac{4\pi^2}{\lambda^2}.\sqrt{\frac{\mu}{\epsilon}}$$

$$= \frac{4\pi^2}{\lambda^2}\eta_0$$

where η_0 is the intrinsic impedance of the medium, and its value for the free space is 120π (see Chapter 4). The Eqn (8.70) yields

$$\boxed{W = 40\pi^2 I_0^2 (\frac{dl}{\lambda})^2} \tag{8.71}$$

From Eqn (8.71) it is worth noting that the power radiated by the dipole is proportional to the square of the dipole length normalized with respect to the wavelength. The radiated power, for a given length of the dipole, dl, and a given current I_0, is inversely proportional to the wavelength or directly proportional to the frequency. In other words, an electrical component starts radiating more and more power as the frequency increases, and the power reaches out to farther distances.

EXAMPLE 8.6 A 1m long vertical Hertz dipole is excited to have a peak current of 100 A at $\omega = 3 \times 10^6$ rad/s. Find the radiation fields at a distance of 1 km at an elevation of 30°.

Solution:

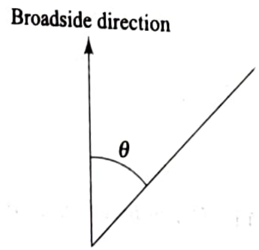


Fig. 8.12

Here we have

$$\theta = 90° - 30° = 60°$$

$$\beta = \frac{\omega}{c} = \frac{3\times 10^6}{3\times 10^8} = \frac{1}{100}\ \text{rad/m}$$

$$\beta r = \frac{1}{100} \times 10^3 = 10\ \text{rad}$$

The radiation fields from Eqns (8.60), (8.61) are

$$|E| = |E_\theta| = \frac{100 \times 1 \times 10^{-4} \times \sin(60°)}{4\pi \times \frac{1}{36\pi} \times 10^{-9} \times 3 \times 10^6 \times 10^3}$$

$$= 3\frac{\sqrt{3}}{2} \times 10^{-2}\ \text{V/m}$$

$$= 0.026\ \text{mV/m}$$

$$|H| = \frac{|E|}{120\pi}$$

$$= 0.0869\ \mu\text{A/m}$$

EXAMPLE 8.7 A vertical Hertz dipole radiates 1 kW power. Find the electric field and the Poynting vector at a distance of 10 km from the dipole in the horizontal plane passing through the dipole. What is the direction of the electric field at the observation point?

Solution:

The total power radiated by a Hertz dipole is

$$W = 1\ \text{kW} = 40\pi^2 I_0^2 \left(\frac{dl}{\lambda}\right)^2$$

$$\Rightarrow \quad I_0\left(\frac{dl}{\lambda}\right) = \left(\frac{1\times 10^3}{40\pi^2}\right)^{1/2} = \frac{5}{\pi}$$

In the horizontal plane $\theta = 90°$ for any point. Therefore, from Eqn (8.60)

$$|E| = \frac{I_0 dl \beta^2}{4\pi\epsilon\omega r}$$

$$= \frac{I_0 dl}{4\pi\epsilon r}\frac{\beta}{\omega}\cdot\beta$$

$$= \frac{I_0 dl}{4\pi\epsilon r c}\cdot\frac{2\pi}{\lambda}$$

$$|E| = \frac{1}{2\epsilon r c}\cdot\left(\frac{I_0 dl}{\lambda}\right)$$

Substituting for $\frac{I_0 dl}{\lambda}$, we get,

$$|E| = \frac{1}{2 \cdot \frac{1}{36\pi} \times 10^{-9} \times (10\ \text{km})(3 \times 10^8\ \text{m/s})} \cdot \frac{5}{\pi}$$

$$= 30\ \text{mV/m}$$

The poynting vector is

$$P = \frac{|E|^2}{\eta} = \frac{30 \times 10^{-3}}{120\pi}\ \text{W/m}^2$$

$$= \frac{3 \times 10^{-4}}{40\pi}\ \text{W/m}^2$$

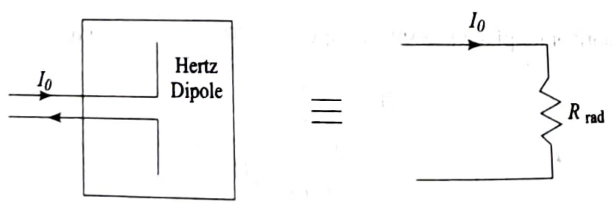
The observation point is $\theta = \pi/2$, hence, the electric field is vertical.

8.2.2 Radiation Resistance of the Hertz Dipole

Now, consider the dipole as a black box, fed with a peak current I_0. Since, the antenna radiates power W, we can equivalently say that the black box consumes W Watt of power. In other words, the black box (i.e. Antenna) appears like a resistance when seen from the feed side. This resistance is called the radiation resistance of the dipole and is generally denoted by R_{rad}. Now, from our basic circuit analysis we know that the power loss in a resistance carrying a peak current I_0 is $\frac{1}{2} I_0^2 R_{rad}$. Equating this loss to the power radiated by the dipole W, we get

$$W = 40\pi^2 I_0^2 \left(\frac{dl}{\lambda}\right)^2 = \frac{1}{2} I_0^2 R_{rad} \tag{8.72}$$

$$\Rightarrow \quad R_{rad} = 80\pi^2 \left(\frac{dl}{\lambda}\right)^2 \tag{8.73}$$



Black Box

Fig. 8.13 *Hertz dipole equivalent to a resistance when seen from the input terminals of the antenna.*

Keeping in mind that for the Hertz dipole $dl << \lambda$, we note that the R_{rad} and consequently radiated power is generally small for the Hertz dipole. For a $0.1\,\lambda$ long dipole, the radiation resistance is $\sim 8\ \Omega$ only.

EXAMPLE 8.8 Two Hertz dipoles, one horizontal and other vertical, are co-located. The vertical dipole has a length of 1 m and the horizontal dipole has a length of 2 m. The two dipoles are excited with currents which have same amplitudes of 1 A each but the vertical current leads the horizontal current by 60°. Find the state of polarization of the radiated wave at a distance of 1 km in the horizontal plane along the direction which is perpendicular to both the dipoles. Also find the magnitude and direction of the electric field at $t = 1\ \mu$s. The $t = 0$ refers to the time when the horizontal current passes through 0 phase point. Angular frequency of the wave is 1 Mrad/s.

Solution:

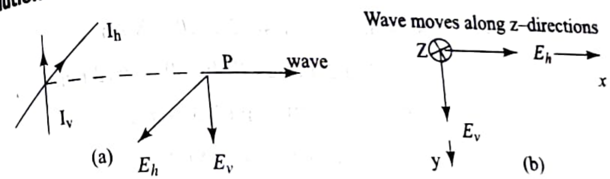


Fig. 8.14

Since, the radiation field is proportional to the currents, the fields E_v due to vertical dipole leads the field E_h due to the horizontal dipole. Also, since the field is proportional to the dipole length, we have

$$\frac{|E_h|}{|E_v|} = 2$$

Assuming $\quad |E_v| = E_0,$

we get $\quad |E_h| = 2E_0$

Using right handed coordinate system, at the observation point as shown in Fig. 8.14(a), we have

$E_x = 2E_0\angle 0 \qquad$ and $\qquad E_y = E_0\angle 60°$

Since, $|E_x| \neq |E_y|$, the wave is elliptically polarized. Also since E_y leads E_x, the sense of rotation is LH.

The equation of ellipse of polarization is (see Chapter 4)

$$\frac{E_x^2}{4E_0^2} - \frac{2E_x E_y \cos 60°}{2E_0^2} + \frac{E_y^2}{E_0^2} = \sin^2 60°$$

$$\Rightarrow \quad \frac{E_x^2}{4} - \frac{E_x E_y}{2} + E_y^2 = \frac{3}{4}E_0^2$$

$$\Rightarrow \quad E_x^2 - 2E_x E_y + 4E_y^2 = 3E_0^2$$

Now, for the wave

$$\beta = \frac{\omega}{c} = \frac{10^6}{3 \times 10^8} = \frac{1}{300} \text{ rad/m}$$

At $t = 1\ \mu$ s, we get $\omega t = 10^6 \times 10^{-6} = 1$ rad

The vertical and horizontal fields at the point of observation are ($\theta = \pi/2$ for both dipoles)(from Eqn (8.60))

$$E_h = Re\left\{ j \frac{1(A)2(m)(1/300)^2}{4\pi \times \frac{10^{-9}}{36\pi} \times 10^6 \times (1\text{ km})} e^{j(1-\frac{1000}{300})} \right\}$$

$$= Re\left\{ 2 \times 10^{-4} j e^{-j\frac{7}{3}} \right\}$$

$$= 2 \times 10^{-4} \sin(7/3) = 1.4462 \times 10^{-4} \text{ V/m}$$

Now, magnitudewise E_v is $E_h/2$ and it leads E_h by $60° = \pi/3$ rad.

$$\Rightarrow \quad E_v = -1 \times 10^{-4} \sin(7/3 + \pi/3)$$

$$= -0.2367 \times 10^{-4} \text{ V/m}$$

The total electric field

$$E = \sqrt{|E_h|^2 + |E_v|^2} = 1.465 \times 10^{-4} \text{ V/m}$$

Angle which **E** makes with the horizontal direction $= \tan^{-1}\left(\frac{E_v}{E_h}\right) = 9.3°$

8.2.3 Radiation Pattern of the Hertz Dipole

The radiation pattern is a plot of magnitude of the radiation electric field as a function of θ and ϕ. It essentially describes the way the power flow varies as a function of direction. Since, we are plotting the electric field magnitude as a function of θ and ϕ, it is clear that the plot is a surface in three-dimension. Moreover, since the purpose of the radiation pattern is to provide the directional dependence of the radiation field, the absolute field amplitudes are not of much concern. Consequently, the radiation field can be normalized with respect to its maximum value. The radiation pattern for the Hertz dipole therefore is

$$F(\theta, \phi) = \sin\theta \tag{8.74}$$

Equation (8.74) when plotted in the spherical coordinate system (ρ, θ, ϕ) with $\rho = |F(\theta, \phi)|$ generates a shape as shown in Fig. 8.15. This shape is similar to an apple. The radiation pattern for a Hertz dipole, therefore, appears like an apple.

Though it is always correct to visualize the radiation pattern as a three dimensional figure, many times, just the two principal sections of the figure are adequate to describe the gross radiation characteristics of an antenna. To this end,

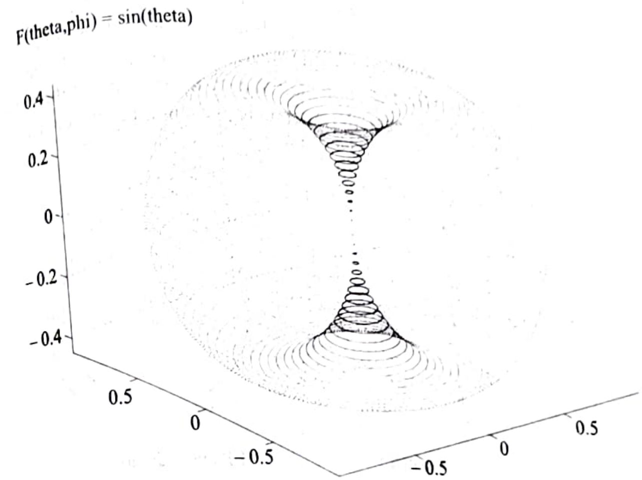


Fig. 8.15 *Radiation pattern of a Hertz dipole.*

the principal sections are obtained by cutting the three dimensional figure with two planes perpendicular to each other. To obtain a good description of the figure like apple, the two appropriate planes would be one vertical and one horizontal both passing through the antenna center and the point of maximum radiation as shown in Fig. 8.16.

One can note that a vertical plane oriented along any ϕ, would contain the electric field vectors and therefore we call the section shown in Fig. 8.16(a) as the *E* plane radiation pattern. The horizontal plane contains the magnetic field component, H_ϕ and consequently, the section shown in Fig. 8.16(b) is called the *H* plane radiation pattern.

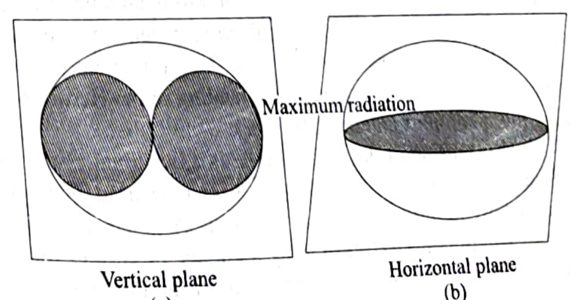


Fig. 8.16 *Sections of the three dimensional radiation pattern shown in Fig. 8.15 in two perpendicular planes.*

The E and H plane radiation patterns for the Hertz dipole are shown in Fig. 8.17.

Readers should be reminded that the length ρ in Fig. 8.17 does not represent distance but the magnitude of the electric field. Also, the physical size of the antenna does not appear in the radiation pattern plot. Irrespective of the physical complexity and size of the antenna, the antenna is just a point at the center of the radiation pattern.

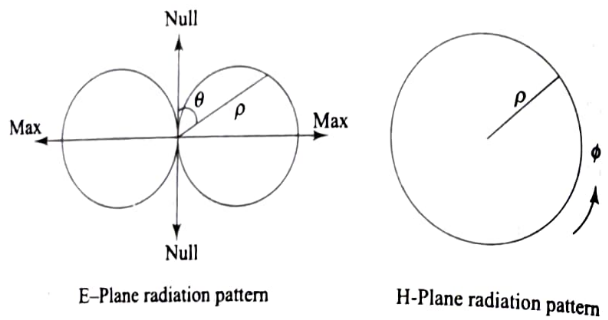


Fig. 8.17 *E and H plane radiation patterns of a Hertz dipole.*

The radiation pattern is a very useful description of an antenna as it indicates the directions of maximum radiation as well as the directions of no radiation. The directions of no radiation are called the 'nulls' of the radiation pattern. For the Hertz dipole, there are two nulls at $\theta = 0$ and $\theta = 180°$ and the direction of maximum radiation corresponds to $\theta = 90°$ for all angles ϕ. The radiation, therefore, goes equally in all directions in a plane perpendicular to the dipole, and there is no radiation along the axis of the dipole.

8.3 THIN LINEAR ANTENNA

In the previous section we investigated the radiation from an infinitesimal current element, the Hertz dipole. We noted that a small current element is not an efficient radiator as its radiation resistance is very small. A current element therefore is not of much practical use. However, the understanding of the Hertz dipole helps us in analyzing more complex antenna structures.

The next logical extension of the Hertz dipole is a linear antenna or a dipole of substantial physical length. This dipole then can be used in practice for transmitting and receiving electromagnetic radiation. A linear dipole is simply a piece of wire of length (say) $2H$ excited by a voltage or current source at its center as shown in Fig. 8.18. The gap at the center of the dipole is assumed to be small. The current spreads on the length of the dipole and attains a steady state distribution. Since the dipole is thin, the current flows only along the length of the dipole. It is not very obvious at this stage, what distribution the current attains along the dipole. In fact, finding the current distribution along an antenna is quite a complicated problem. As will however be seen shortly, once the current

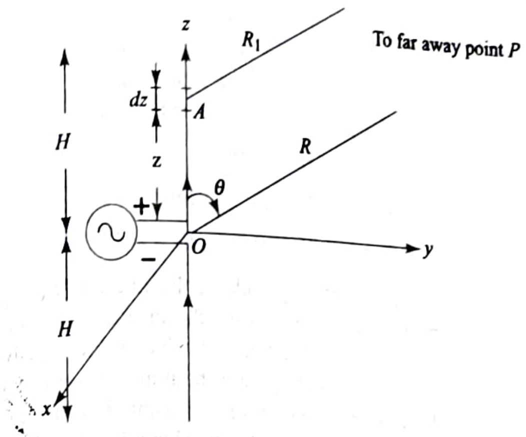


Fig. 8.18 *Linear dipole antenna of length $2H$.*

distribution is known, finding the radiation characteristics of the antenna is a relatively simpler problem.

Although the exact current distribution is not obvious, we can say that due to symmetry of the dipole with respect to the feed point, the current distribution will be symmetric with respect to the center of the dipole. We can also say that at the ends of the dipole since there is no path for the current to flow, the current will be zero at the ends of the dipole. A detail derivation of the current distribution is beyond the scope of this book. For the time being, we can use some crude arguments to deduce the current distribution on a thin dipole. To this end, we visualize a dipole as a flared up version of a transmission line as shown in Fig. 8.19.

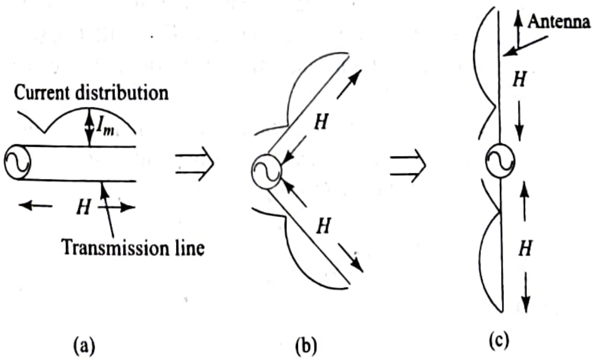


Fig. 8.19 *A dipole antenna as a flared version of a transmission line. Current distribution on a centre-fed dipole of length $2H$.*

Consider an open circuited section of a transmission line of length H as shown in Fig. 8.19(a). The voltage reflection coefficient at the open circuited end of the line is +1 and consequently we have a fully developed standing wave along the line. The current at the open circuited end is zero and its magnitude has a sinusoidal variation (see 'the standing waves' in Chapter 2).

The standing wave current distribution on the line is given as

$$I(z) = I_m \sin(\beta(H - |z|)) \tag{8.75}$$
$$= I_m \sin(\beta(H - z)) \qquad \text{for } z > 0$$
$$= I_m \sin(\beta(H + z)) \qquad \text{for } z < 0 \tag{8.76}$$

where I_m is the maximum current amplitude at the line and $\beta (= \frac{2\pi}{\lambda})$ is the propagation constant in the free space. z is the distance measured on the line from the generator. z is positive on the upper conductor and negative on the lower conductor of the line. Now, slowly flare the transmission line as shown in Fig. 8.19(b). Indeed we have a non-uniform transmission line now, as the separation between the two conductors is not same along their lengths. However, the current distribution still remains same. If we further flare it to become a dipole antenna as in Fig. 8.19(c), the current distribution still remains same as that given by Eqn (8.75). One may wonder at this point whether it is justified to use the current distribution obtained from a uniform transmission line to its completely flared up version, the dipole antenna. The answer is strictly 'No'. However, one can show by analysis that the current distribution on a thin dipole is sinusoidal and is given by Eqn (8.75). At this point, therefore, we will proceed with our analysis assuming the current distribution to be sinusoidal, without asking, how we got it.

Once the current distribution along the dipole is given, the radiation properties of the dipole can be analysed by dividing the dipole into infinitesimally small current elements and by applying superposition. Without losing generality let us assume that the origin of the coordinate system lies at the center of the dipole and the dipole is oriented along the z axis as shown in Fig. 8.18. Consider a small current element of length dz at a height z from the origin. The current in the element is given by Eqn (8.76).

Here, we are interested only in the radiation or the far field of the dipole. The electric field due to the current element at a far away point P (ideally the point P is at ∞) can be written using Eqn (8.60) as

$$dE_\theta = j\frac{\beta^2 \sin\theta}{4\pi\omega\epsilon}\frac{I(z)dze^{-j\beta R_1}}{R_1} \tag{8.77}$$

and the magnetic field is

$$dH_\phi = \frac{dE_\theta}{\eta} \tag{8.78}$$

Since, the observation point P is far away from the antenna, the line AP and OP are almost parallel and we have

$$R_1 = R - z\cos\theta \tag{8.79}$$

In Eqn (8.77) the distance appears in absolute sense in the denominator, whereas in the argument of the exponential function it appears as $\beta R_1 = 2\pi\frac{R_1}{\lambda}$, that is relative to the wavelength λ. Consequently, replacing R_1 by R in the denominator does not cause any significant error, but the same approximation causes significant phase change in the exponential term. We, therefore, write

$$dE_\theta = j\frac{\beta^2 \sin\theta}{4\pi\omega\epsilon R} I(z)e^{-j\beta R + j\beta z\cos\theta}dz \tag{8.80}$$

The total field due to the entire dipole can be obtained by integrating dE_θ over the dipole length to give

$$E_\theta = j\frac{\beta^2 \sin\theta e^{-j\beta R}}{4\pi\omega\epsilon R}\int_{-H}^{H} I(z)e^{j\beta z\cos\theta}dz \tag{8.81}$$

$$= j\frac{\beta^2 \sin\theta e^{-j\beta R}}{4\pi\omega\epsilon R}\left\{\int_{-H}^{0} I_1(z)dz + \int_{0}^{H} I_2(z)dz\right\} \tag{8.82}$$

$$= j60I_m\frac{e^{-j\beta R}}{R}F(\theta) \tag{8.83}$$

where

$$I_1(z) = I_m \sin\beta(H + z)e^{j\beta z\cos\theta}$$
$$I_2(z) = I_m \sin\beta(H - z)e^{j\beta z\cos\theta}$$

Here, we have used $\beta = \omega\sqrt{\mu\epsilon}$ to get

$$\frac{\beta^2}{4\pi\omega\epsilon} = \frac{\omega^2\mu\epsilon}{4\pi\omega\epsilon} = \frac{\omega\mu}{4\pi} = \frac{\beta\eta_o}{4\pi} = \frac{120\pi\beta}{4\pi} = 30\beta \tag{8.84}$$

and the $F(\theta)$ is given as

$$\boxed{F(\theta) = \frac{\cos(\beta H\cos\theta) - \cos(\beta H)}{\sin\theta}} \tag{8.85}$$

$F(\theta)$ describes the relative variation of the electric field as a function of θ and therefore is the E-plane radiation pattern. In the H plane, E_θ is constant (as it is not a function of ϕ) and hence the H plane radiation pattern of the dipole is a circle which is same as that of the Hertz dipole. The E-plane pattern however is different than that of the Hertz dipole, and it varies with the length of the dipole.

Figure 8.20 shows the current distributions on dipoles of different lengths and the corresponding E-plane radiation patterns. The H-plane radiation pattern for a dipole is always a circle. The three dimensional radiation patterns can be obtained by revolving the E plane patterns around the axis of the dipole, i.e. the z axis. These are shown in Figs. 8.21 to 8.23.

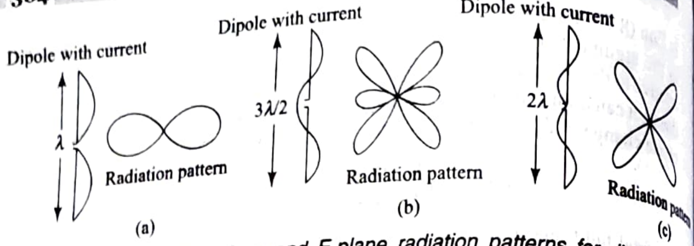


Fig. 8.20 *Current Distributions and E-plane radiation patterns for dipoles of different lengths.*

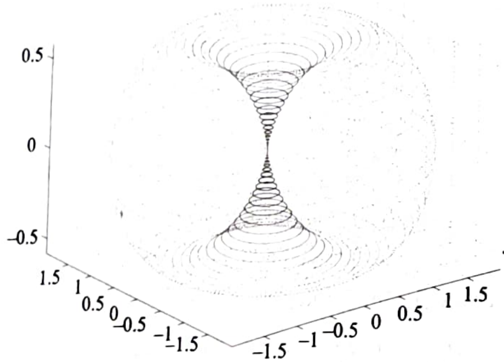


Fig. 8.21 *Three dimensional radiation pattern for a λ long dipole.*

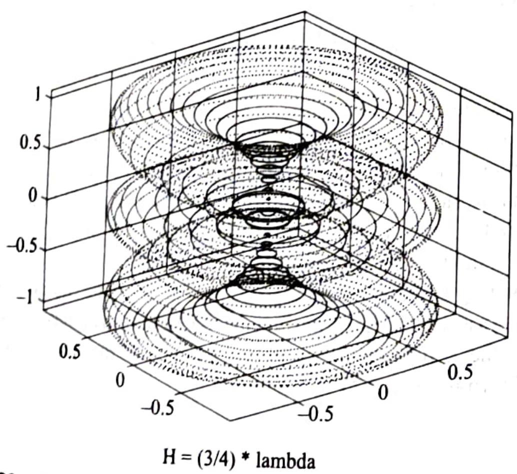


Fig. 8.22 *Three dimensional radiation pattern for a 1.5λ long dipole.*

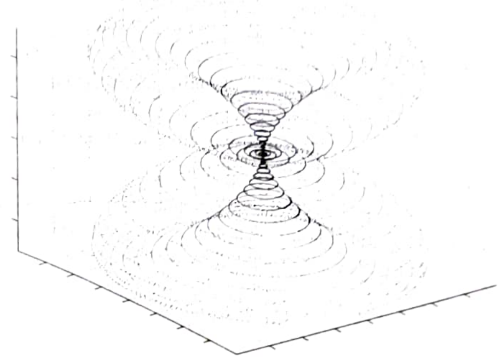


Fig. 8.23 *Three dimensional radiation pattern for a 2λ long dipole.*

From these figures, we can make following important observations:

1. The input current at the antenna feed terminal is

$$I_{in} = I_m \sin\beta H \tag{8.86}$$

and the input impedance of the dipole is

$$Z_{in} = \frac{V}{I_{in}} = \frac{V}{I_m \sin\beta H} \tag{8.87}$$

The input impedance is a function of the dipole length and can vary from $\frac{V}{I_m}$ to ∞. When $H =$ odd multiple of $\frac{\lambda}{4}$, $\beta H = m\pi/2$ and $|\sin\beta H| = 1$. In this case the input impedance is $\frac{V}{I_m}$. When $H =$ even multiple of $\frac{\lambda}{4}$ (i.e. multiple of $\frac{\lambda}{2}$), $\sin\beta H \to 0$ making impedance $Z_{in} \to \infty$.

2. Since with the increasing length of the dipole more and more current becomes available for radiation (assuming I_m remains same for all lengths) the total power radiated by the dipole is expected to increase monotonically with the length. Since, the input impedance is not a monotonic function of length anymore, there is no obvious relation between the input resistance and the power radiated by the dipole.

3. The electric field has only E_θ component making the radiation linearly polarized in the $\hat{\theta}$ direction. A dipole antenna therefore generates a linearly polarized radiation irrespective to its length.

4. The radiation pattern tends to have more nulls as the length of the dipole increases. This means, although the radiated power increses with the dipole length, the power does not uniformly go in all directions. There are certain preferred directions in which the power is radiated.

5. One hardly has any control over the input impedance and also the radiation pattern of the dipole. Moreover, the two cannot be varied independently, i.e. it is not possible to change the radiation pattern without affecting the input impedance of the dipole and vice versa. In practice however, impedance is not a major issue as one can always modify the impedance by impedance-transformation techniques (see Chapter 2 on Transmission Lines). In the design of a dipole and for that matter even other type of antenna, therefore the main focus is on the radiation pattern.

EXAMPLE 8.9 A small dipole of length 0.1 λ is excited with a peak current of 5 A. How much power will be radiated by the antenna?

Solution:

Radian resistance of the dipole is

$$R_{rad} = 80\pi^2\left(\frac{dl}{\lambda}\right)^2 = 80\pi^2\left(\frac{0.1\lambda}{\lambda}\right)^2 = 0.8\pi^2\ \Omega$$

$$\text{The power radiated}\quad W = \frac{1}{2}i_0^2 R_{rad} = \frac{1}{2}(5)^2 \times 0.8\pi^2 = 98.696\ \text{W}$$

From Eqn (8.85) we can obtain the directions of the nulls by equating $F(\theta)$ to zero. That is

$$\frac{\cos(\beta H\cos\theta_{null}) - \cos(\beta H)}{\sin\theta_{null}} = 0 \tag{8.88}$$

$$\Rightarrow\quad \cos(\beta H\cos\theta_{null}) = \cos(\beta H) \tag{8.89}$$

$$\Rightarrow\quad \beta H\cos\theta_{null} = \pm\beta H \pm 2m\pi,\qquad m = 0, 1, 2\ldots \tag{8.90}$$

$$\Rightarrow\quad \cos\theta_{null} = \pm 1 \pm \frac{2m\pi}{\beta H} = \pm 1 \pm \frac{m\lambda}{H} \tag{8.91}$$

For $m = 0$, $\cos\theta_{null} = \pm 1$ giving $\theta_{null} = 0$ or π. In this case, although the numerator of Eqn (8.88) goes to zero, the denominator also goes to zero and it becomes imperative to find whether $m = 0$ corresponds to a null. This can be investigated by taking limit of $F(\theta)$ as $\theta \to 0$ or π

$$\lim_{\theta\to 0 \text{ or } \pi} F(\theta) = \lim_{\theta\to 0 \text{ or } \pi} \frac{\cos(\beta H\cos\theta) - \cos\beta H}{\sin\theta} \tag{8.92}$$

Expanding the cosine function in a power series we get

$$\lim_{\theta\to 0 \text{ or } \pi} F(\theta) = \lim_{\theta\to 0 \text{ or } \pi} \frac{\left[1 - \frac{(\beta H\cos\theta)^2}{2!} + \frac{(\beta H\cos\theta)^4}{4!}\right] - \left[1 - \frac{(\beta H)^2}{2!} + \frac{(\beta H)^4}{4!}\right]}{\sin\theta} \tag{8.93}$$

$$= \lim_{\theta\to 0 \text{ or } \pi} \frac{(\beta H)^2\frac{(1-\cos^2\theta)}{2!} - \frac{(\beta H)^4(1-\cos^4\theta)}{4!}}{\sin\theta} \tag{8.94}$$

$$= \lim_{\theta\to 0 \text{ or } \pi}\left\{(\beta H)^2\frac{\sin\theta}{2} - (\beta H)^4\sin\theta(1+\cos^2\theta)\right\} = 0 \tag{8.95}$$

From Eqn (8.95) we conclude that $\theta = 0$ and $\theta = \pi$ are nulls of the radiation pattern. In fact, these are the two nulls which always remain irrespective of the length of the dipole. We may recall that these are the nulls of the Hertz dipole and since a linear dipole is a superposition of collinear Hertz dipoles, the original nulls of the Hertz dipole remain in the final radiation pattern. Equation (8.91), hence, gives the direction of nulls for all integer values of m including 0. Of course, only those values of m have to be considered which give $\cos\theta_{null} \le 1$.

Since, between every two adjacent nulls the magnitude of radiation field must have gone through a maximum, one can approximately get the directions of maximum radiation as the mean of the adjacent null directions. Although, this does not give very precise direction of the maximum radiation, this approach is commonly used as it is less complex compared to finding the maximum of $|F(\theta)|$ by equating $\frac{d|F(\theta)|}{d\theta}$ to zero.

EXAMPLE 8.10 A 1.2 λ long dipole has 1 Amp peak input current. Find the maximum peak current seen on the dipole. If the dipole is oriented along the z-axis, find the radiation electric and magnetic fields at a distance of 100 m along $\theta = 60°$.

Solution:

From Eqn (8.86) we have,

$$I_{in} = 1.0 = I_m|\sin\beta H|$$

$$\text{Here}\quad H = \frac{1.2\lambda}{2} = 0.6\lambda \quad\Rightarrow\quad \beta H = \frac{2\pi}{\lambda}\times 0.6\lambda = 1.2\pi\ \text{rad}$$

The maximum peak current therefore is

$$I_m = \frac{I_{in}}{|\sin(1.2\pi)|} = \frac{1.0}{\sin(1.2\pi)} = 1.70\ \text{A}$$

The radiation field from Eqn (8.83) is

$$|E_\theta| = \frac{60 I_m}{R}\left\{\frac{\cos(1.2\pi\cos\theta) - \cos(1.2\pi)}{\sin\theta}\right\}$$

Here we have $R = 100\,m, \theta = \pi/3$

$$\Rightarrow \qquad |E_\theta| = \frac{60 I_m}{100}\left\{\frac{\cos(0.6\pi) - \cos(1.2\pi)}{\sin(\pi/3)}\right\} = 0.09 \text{ V/m}$$

The magnetic field $|H_\phi| = \dfrac{|E_\theta|}{\eta} = \dfrac{|E_\theta|}{120\pi} = 0.24$ mA/m

8.4 RADIATION PARAMETERS OF AN ANTENNA

The radiation pattern of an antenna describes the directional dependence of the power radiated by the antenna. In general a radiation pattern has a direction of maximum radiation. Then, there are certain directions called nulls, along which no radiation goes, and there are directions in which the radiation is locally maximum. The local maxima in the radiation pattern are called the 'side lobes' of the radiation pattern. A typical radiation pattern in some plane, therefore, appears like that shown in Fig. 8.24.

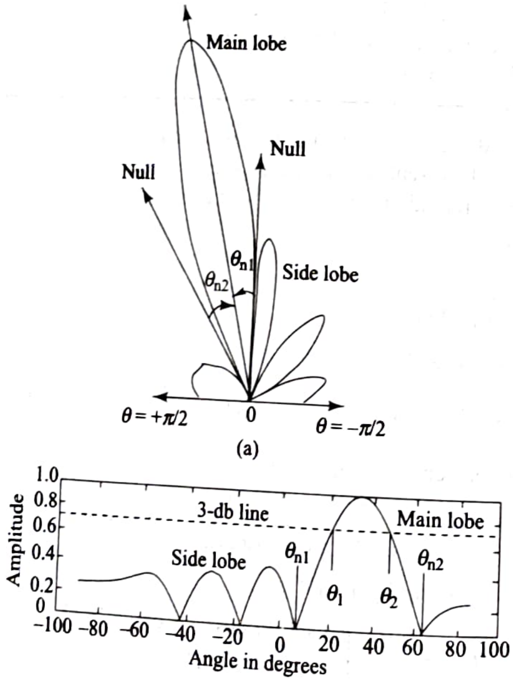


Fig. 8.24 *Radiation pattern as a (a) polar plot (b) Cartesian plot.*

Figure 8.24(a) shows the radiation pattern as the polar plot and Fig. 8.24(b) shows the radiation pattern as the Cartesian plot. Figure 8.24(b) is rather a representation of the actual radiation pattern taking θ as an independent parameter and plotting $|E|$ as a function of θ in a rectangular plot. The radiation pattern indeed is periodic over 360° i.e. 2π radians.

Although the radiation pattern is a complete description of the directional properties of antenna radiation, its utility is rather restricted due to its too detailed nature. One would instead like to derive certain quantitative parameters from the radiation pattern which can be used for comparison of different antennas. These parameters are as follows:

8.4.1 Direction of the Main Beam

It is the direction in which the radiation field strength is maximum. Generally it is denoted by θ_{max}.

8.4.2 Half-Power Beam Width (HPBW)

The main beam is the angular region where effective radiation from an antenna goes. The effective width of the main beam is the angular width of the pattern between the points on the radiation pattern where the field magnitude reduces to $\frac{|E_{max}|}{\sqrt{2}}$, where E_{max} is the maximum field. From Fig. 8.24(b), the HPBW = $(\theta_2 - \theta_1)$. Since, the field in directions θ_1 and θ_2 reduces to $1/\sqrt{2}$ of its maximum value, the Poynting vector in these direction reduces to $\frac{1}{2}$ or −3dB compared to that in the direction of the maximum radiation. (Poynting vector $\propto |E|^2$ or $|H|^2$). The HPBW, therefore, is many times referred to as 3dB Beam width of the antenna. Generally, the HPBWs are measured in the E and H planes of the radiation pattern.

8.4.3 Beam Width between First Nulls (BWFN)

Occasionally, the width of the main beam (also called main lobe) is measured by angular separation between the first nulls on either side of the direction of maximum radiation. From Fig. 8.24(b), BWFN $= \theta_{n2} - \theta_{n1}$. It should be mentioned here that the HPBW is a better measure of the effective width of the main beam compared to the BWFN as there are cases where the HPBW changes but the BWFN remains fixed.

8.4.4 Side Lobe Level (SLL)

The presence of side lobes essentially indicates the leakage of power in the undesired directions. An antenna system is primarily designed to transmit radiation in the direction along the main beam. Ideally, we would like to have no radiation outside the main beam. However, since in any practical antenna system we have side lobes, the total radiated power is not focused into the main beam but a part of it leaks in the directions of the side lobes. Obviously, one would like to keep this leakage minimal. In other words, the side lobe amplitudes should be as small as possible compared to the amplitude of the main beam. The amplitude of the highest side lobe compared to that of the main beam is called the side lobe level (SLL). For good satellite communication antennas the SLL is −30 dB to −40 dB. Generally the amplitude of the sidelobes reduces as we go away from the main beam and hence the first side lobe on either side of the main beam defines the SLL.

8.4.5 Directivity

The directivity is a parameter which quantifies the radiation focusing capability of an antenna. The HPBW in some sense has similar information but it does not quantify how much of the total radiated power has been confined into the main beam. As discussed earlier, a part of the radiated power leaks through the side lobes and the remaining power is spread over the HPBW. Since, here we are investigating the angular distribution of the radiated power, it is appropriate to define a quantity called the Radiation Intensity as the power per unit solid angle. (Note, that the radiation pattern is a three dimensional surface). The radiation intensity in general is a function of θ and ϕ and is given by

$$U(\theta,\phi) = \frac{\text{Power along direction } (\theta,\phi) \text{ in a solid angle } d\Omega}{\text{solid angle}(d\Omega)} \tag{8.96}$$

Now the solid angle on the surface of a sphere is defined as

$$d\Omega = \frac{dA}{r^2} \tag{8.97}$$

where dA is the area on the surface of the sphere and r is the radius of the sphere. $d\Omega = \sin\theta d\theta d\phi$. Substituting Eqn (8.97) into Eqn (8.96) we get

$$U(\theta,\phi) = \frac{\text{Power along}(\theta,\phi)}{dA} r^2 \tag{8.98}$$

$$= (\text{Power Density})\, r^2 \tag{8.99}$$

The power density of a radiated wave is given by the magnitude of the Poynting vector $P(\theta,\phi) = \frac{|E(\theta,\phi)|^2}{\eta}$. We, therefore, get

$$U(\theta,\phi) = \frac{|E(\theta,\phi)|^2}{\eta} r^2 = P(\theta,\phi) r^2 \tag{8.100}$$

The total power radiated W by the antenna can be obtained by integrating the radiation intensity over the total 4π solid angle i.e.

$$W = \int_{\Omega=4\pi} U(\theta,\phi) d\Omega = \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} U(\theta,\phi)\sin\theta d\theta d\phi \tag{8.101}$$

If W watts of power is uniformly radiated in all directions, i.e. over 4π solid angle, we get the average radiation intensity as

$$U_{av} = \frac{W}{4\pi} = \frac{1}{4\pi}\int\int U(\theta,\phi) d\Omega \tag{8.102}$$

One can then ask, "for a given antenna, how much is the increase or decrease in the radiation intensity in a certain direction compared to U_{av}". 'The directive gain' of an antenna is defined as

$$G(\theta,\phi) = \frac{U(\theta,\phi)}{U_{av}} = \frac{4\pi U(\theta,\phi)}{\int\int U(\theta,\phi) d\Omega} \tag{8.103}$$

It is obvious that $G(\theta,\phi)$ can be less than or greater than unity with no bounds on either side. That is, theoretically $G(\theta,\phi)$ can vary from 0 to ∞. In the direction of the nulls $E(\theta,\phi)$ and consequently $U(\theta,\phi)$ and $G(\theta,\phi)$ are zero, whereas, in the direction of the main beam $G(\theta,\phi)$ is maximum.

The maximum directive gain is called the 'Directivity' of the antenna and is denoted by D.

$$D = \frac{U_{max}}{U_{av}} = \frac{4\pi U_{max}}{\int\int U(\theta,\phi) d\Omega}$$

$$D = Max(G(\theta,\phi)) \tag{8.104}$$

Using Eqn (8.100) we get,

$$D = \frac{4\pi E_{max}^2}{\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} |E(\theta,\phi)|^2 \sin\theta d\theta d\phi} \tag{8.105}$$

The directivity D lies between 1 and ∞. For $D = 1$, $U_{max} = U_{av}$ and the antenna does not exhibit any directional characteristic. This antenna is called the 'isotropic antenna'. It may be noted that an isotropic antenna is only hypothetical and in practice one cannot realize an isotropic antenna. Hence, for a practical antenna D is always greater than 1. As the directivity D increases, the radiation gets more and more focused is the direction of the main beam.

If we define the normalized radiation pattern as $E_n(\theta,\phi) = \frac{E(\theta,\phi)}{E_{max}}$, the directivity can be written as

$$D = \frac{4\pi}{\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} |E_n(\theta,\phi)|^2 \sin\theta d\theta d\phi} \tag{8.106}$$

$|E_n(\theta,\phi)|$ by definition has maximum value of unity. For large antennas with low side lobe level, most of the radiated power is confined to the main beam and the integration can be approximated by

$$\int\int_{\text{Main Beam}} |E_n(\theta,\phi)|^2 d\Omega \tag{8.107}$$

For narrow beam antennas

$$\oint |E_n(\theta,\phi)|^2 d\Omega \simeq \theta_{HP}\phi_{HP} = \Omega_{MB} \tag{8.108}$$

where θ_{HP} and ϕ_{HP} are the half power beam widths in radians in the θ and ϕ planes respectively. This can be seen from Fig. 8.25. From Fig. 8.25 we can note that the volume under the actual main beam is approximately equal to the volume of the box with height unity and lengths and widths respectively θ_{HP} and

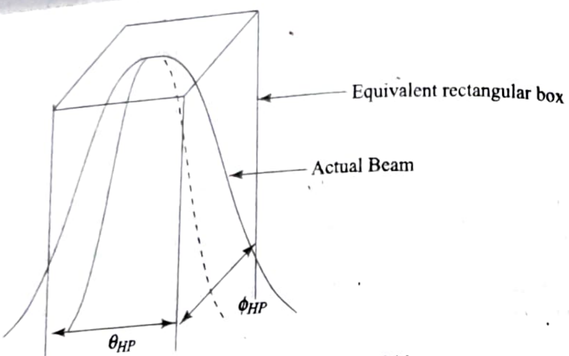


Fig. 8.25 *Half power beam widths.*

ϕ_{HP}. One should keep in mind however, that the approximation is valid for θ_{HP}, $\phi_{HP} << 1$ rad. As θ_{HP}, ϕ_{HP} become comparable to or greater than 1 rad, the approximation cannot be applied and the denominator in Eqn (8.106) has to be evaluated through proper integration. For narrow beam antennas we, therefore, get

$$D \simeq \frac{4\pi}{\Omega_{MB}} = \frac{4\pi}{\theta_{HP}\phi_{HP}} \tag{8.109}$$

8.4.6 Antenna Gain

As seen above, the directivity is a parameter which totally depends upon the radiation pattern. Indirectly it assumes that the total power radiated, W, is same as that supplied to the antenna input. In practice however, the antennas are not made of ideal conductor and consequently when the current flows along the antenna surface, there is a an ohmic loss. A part of the power supplied to the antenna input is therefore lost in heating the antenna due to the ohmic loss. If we assume that power actually radiated by the antenna is W, the power supplied to the input of the antenna will be

$$P_i = W + P_l$$

where P_l is the ohmic loss due to finite conductivity of the antenna. Let us define power efficiency of an antenna as

$$\eta_r = \frac{W}{P_i} = \frac{P_i - P_l}{P_i} = \frac{W}{W + P_l} \tag{8.110}$$

The antenna power gain is defined as

$$G_P = \frac{U_{max}(\text{actual})}{U_{av} \text{ for a lossless case}} \tag{8.111}$$

$$= \frac{4\pi U_{max}(\text{actual})}{P_i} \tag{8.112}$$

$$= \frac{4\pi U_{max}(\text{actual})}{W}\frac{W}{P_i} \tag{8.113}$$

$$G_P = D\eta_r \tag{8.114}$$

we hence have, **Power gain = Directivity × Efficiency.**

EXAMPLE 8.11 A dipole antenna is 1.8 λ long. Find the directions of the nulls.

Solution:

From Eqn (8.91) the directions of nulls are given by

$$\cos\theta_{\text{null}} = \pm 1 \pm \frac{m\lambda}{H}$$

Here $$H = \frac{1.8\lambda}{2} = 0.9\lambda$$

$$\Rightarrow \quad \cos\theta_{\text{null}} = \pm 1 \pm \frac{m\lambda}{0.9\lambda} = \pm 1 \pm \frac{m}{0.9}$$

We accept all values of m for which $\cos\theta_{\text{null}}$ is ≤ 1

(i) For $m = 0$, $\cos\theta_{\text{null}} = \pm 1 \Rightarrow \theta_{\text{null}} = 0, \pi$.

(ii) For $m = 1$, $\cos\theta_{\text{null}} = \pm 1 \pm \frac{1}{0.9} = 1 - \frac{1}{0.9}, -1 + \frac{1}{0.9}$.
i.e. $\cos\theta_{\text{null}} = \pm 1/9$.
The range of θ is from 0 to π. Therefore, $\theta_{\text{null}} = 83.62°, 96.38°$

(iii) For $m = 2$, $\cos\theta_{\text{null}} = \pm 1 \pm \frac{2}{0.9}$
i.e. $|\cos\theta_{\text{null}}| \geq 1$, and hence no physical angle.

The nulls of the radiation pattern, therefore, are
$\theta_{\text{null}} = 0°, 83.62°, 96.39°, 180°$.

EXAMPLE 8.12 The field radiation pattern of an antenna is given by

$$F(\theta) = \frac{\sin(10\cos\theta)}{\sin(2\cos\theta)} \qquad 0 \leq \theta \leq \pi$$

Find the directions of the nulls, direction of maximum radiation and HPBW of the antenna.

Solution:

The antenna has nulls when

$$F(\theta) = 0$$

$$\Rightarrow \quad \sin(10\cos\theta_{\text{null}}) = 0 \qquad \text{provided} \quad \sin(2\cos\theta_{\text{null}}) \neq 0$$

$$\Rightarrow \quad 10\cos\theta_{\text{null}} = \pm m\pi \qquad m = 1, 2, 3, \ldots$$

$$\Rightarrow \qquad \cos\theta_{\text{null}} = \pm\frac{m\pi}{10}$$

$|\cos\theta_{\text{null}}|$ has to be ≤ 1. We therefore get

$$\theta_{\text{null}} = \cos^{-1}\left(\pm\frac{m\pi}{10}\right) \quad \text{and} \quad \frac{m\pi}{10} \leq 1$$

giving $m \leq \frac{10}{\pi}$.

Since, m is an integer, we get, $\qquad m = 1, 2, 3.$

The directions of nulls therefore are

$\theta_{\text{null}} = \cos^{-1}(\pm\frac{\pi}{10}), \cos^{-1}(\pm\frac{\pi}{5}), \cos^{-1}(\pm\frac{3\pi}{10})$

Since, $0 \leq \theta \leq \pi$, the nulls are at

$\theta_{\text{null}} = 19.53°, 51.07°, 71.69°, 108.31°, 128.93°$ and $160.47°$.

At $\theta = \frac{\pi}{2}$, the denominator of $F(\theta)$ is zero making $\theta = \frac{\pi}{2}$ the direction of maximum radiation.

To obtain the HPBW of the antenna, we have to find two angles θ_1, and θ_2 around the direction of maximum radiation, $\theta = \frac{\pi}{2}$, along which the field is $\frac{1}{\sqrt{2}}$ of its maximum value.

The maximum value of $F(\theta)$ is along $\theta = \frac{\pi}{2}$ and can be obtained as

$$|F(\theta)|_{max} = \left|\frac{\sin(10\cos\theta)}{\sin(2\cos\vartheta)}\right| \text{ at } \qquad \theta = \frac{\pi}{2}$$

Since, at $\theta = \frac{\pi}{2}$, $F(\theta)$ has 0/0 form, we evaluate it in the form of a limit.

$$|F(\theta)|_{max} = \lim_{\theta\to\frac{\pi}{2}} \left|\frac{\sin(10\cos\theta)}{\sin(2\cos\theta)}\right|$$

$$= \lim_{\theta\to\frac{\pi}{2}} \left|\left\{\frac{10\cos\theta - \frac{(10\cos\theta)^3}{3!} + \ldots\ldots}{2\cos\theta - \frac{(2\cos\theta)^3}{3!} + \ldots\ldots\ldots}\right\}\right| = 5$$

For the half power points we then have

$$F(\theta) = \frac{5}{\sqrt{2}}$$

$$\Rightarrow \qquad \frac{\sin(10\cos\theta)}{\sin(2\cos\theta)} = \frac{5}{\sqrt{2}}$$

Solving numerically, we get HP angles θ_1 and θ_2 as

$\theta_1 = 82.08°$ and $\theta_2 = 97.93°$

The HPBW $= \theta_2 - \theta_1 \approx 16°$.

EXAMPLE 8.13 The field radiation pattern of an antenna is given as

$$F(\theta) = \cos^4\theta \qquad 0 \leq \theta \leq \frac{\pi}{2} \qquad 0 \leq \phi \leq 2\pi$$

and zero elsewhere. Find the directivity of the antenna.

Solution:

Power pattern of the antenna is

$P(\theta) = |F(\theta)|^2 = \cos^8\theta$

Since the maximum value of $P(\theta)$ is 1, the directivity can be written as

$$D = \frac{4\pi}{\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\frac{\pi}{2}} \cos^8\theta \sin\theta d\theta d\phi}$$

substituting $\cos\theta = t \Rightarrow -\sin\theta d\theta = dt$, we get

$$D = \frac{4\pi}{2\pi\int_0^1 t^8 dt} = \frac{2}{\left[\frac{t^9}{9}\right]_0^1} = 18$$

Directivity of the antenna is $D = 18 = 10\log 18 = 12.55$ dB

8.5 RECEIVING ANTENNA

In the previous sections we looked upon an antenna as a transmitter of electromagnetic radiation. We investigated radiation characteristics of an antenna and established relationships between the currents and the radiated electromagnetic fields. We observed that, depending upon the antenna lengths, the direction of maximum radiation and the directions of the nulls vary in the three-dimensional space. That is, in some directions more power is radiated compared to others. We also found that, from the input side, an antenna appears like an impedance in general and the resistive part of the impedance essentially corresponds to the power radiated by the antenna. So for a transmitting antenna, the cause is the current and the effect is the electromagnetic radiation. For a receiving antenna the two are interchanged, i.e. the cause is the electromagnetic radiation and the effect is the currents induced on the antenna surface and the voltage induced between the antenna terminals. The basic question for a receiving antenna is, "how does the terminal voltage vary as a function of direction and polarization of the incoming radiation, and if the antenna is connected to a load how much maximum power can be received by the load?". Also it would be of interest to know how the parameters of an antenna in the transmitting mode are related to its parameters in receiving mode.

From the reciprocity theorem one can show that an antenna is a reciprocal device, i.e. an antenna has identical radiation and circuit characteristics in transmitting and receiving modes. The proof of the reciprocity theorem is beyond the scope of this book. However, we will try to qualitatively verify the statement of the reciprocity theorem through simple cases.

Let us consider the case of a Hertz dipole. The Hertz dipole which is a small length of wire with an infinitesimal gap at the center is placed in the radiation

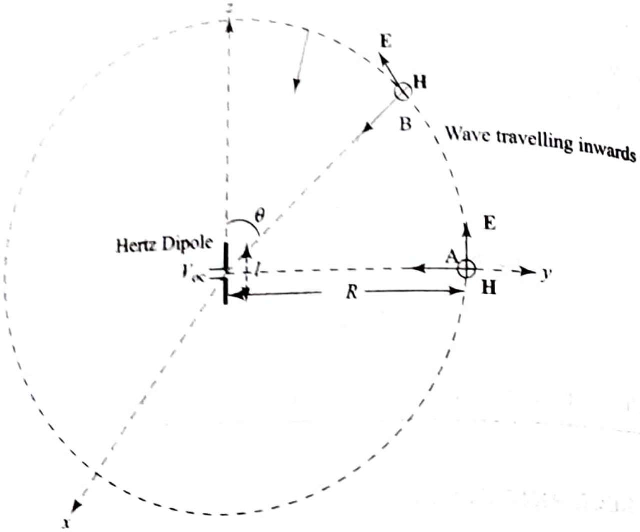


Fig. 8.26 *Hertz dipole in receiving mode.*

field as shown in Fig. 8.26. The dipole is placed along the z axis, at the origin. Let us now consider a source of electromagnetic wave located at a very large distance R from the dipole which generates an electromagnetic wave which is linearly polarized in the plane of the paper. Let us place the source at location A (Fig. 8.26) which is on a line perpendicular to the Hertz dipole axis ($\theta = \frac{\pi}{2}$). In this case, the electric field will be polarized along the z direction, therefore, will be parallel to the Hertz dipole. The open circuit voltage between the terminals of the Hertz dipole will be

$$V_{OC} = V_A = \mathbf{E} \cdot \mathbf{l}$$

$$= El \qquad \text{(since } \mathbf{E} \text{ and } \mathbf{l} \text{ are parallel)} \qquad (8.115)$$

Let us now move the source to position B along the arc of radius R such that the direction of radiation is radially inwards towards the Hertz dipole. The angle between the electric field $\mathbf{E}$ and the Hertz dipole now is $(\frac{\pi}{2} - \theta)$ and consequently, the voltage between the terminals is

$$V_{OC} = V_B = \mathbf{E} \cdot \mathbf{l}$$

$$= El \cos(\frac{\pi}{2} - \theta) = El \sin\theta \qquad (8.116)$$

Equations (8.115) and (8.116) suggest that voltage induced at the receiving antenna terminals depends upon the direction of the incoming radiation, θ. For $\theta = \frac{\pi}{2}$, V_{OC} is maximum and it is zero for $\theta = 0$ or π. The direction dependency of the terminal voltage V_{OC} is given by the factor $\sin\theta$.

We may recall that, the radiation patterr of the Hertz dipole while transmitting radiation is $\sin\theta$. This means, whatever directional dependence the Hertz dipole shows while it is transmitting, the same dependence it exhibits when it is set for reception. In fact, this behavior is true for any radiating system. The radiation patterns of a transmitting and receiving antennas are identical. That is, if an antenna radiates maximum radiation in a certain direction while in transmitting mode, it will have maximum reception from the same direction when used in the receiving mode, and if an antenna gives no radiation in certain direction it can not receive also any radiation coming from that direction.

We may further note that, for a fixed location of the transmitting and receiving antennas (say A), as the direction of the $\mathbf{E}$ changes in the xz plane, the terminal voltage V_{OC} changes and it becomes zero when $\mathbf{E}$ is oriented in x direction. In other words, the receiving antenna voltage varies as a function of polarization of the incoming radiation.

For the Hertz dipole, V_{OC} will be maximum when $\mathbf{E}$ will be along θ direction, i.e. $\mathbf{E} = E_\theta\hat{\theta}$. We may again recall that Hertz dipole generates a linear polarization with electric field $\mathbf{E} = E_\theta\hat{\theta}$. It can therefore be concluded that an antenna maximally responds to that polarization while in receiving mode, which it generates while in transmitting mode. We can then say that a receiving antenna has a state of polarization which is same as that which it is capable of generating. Or said differently, the state of polarization of a receiving antenna is the state of polarization of that wave to which the antenna responds maximally.

From the above discussion, we conclude that, the radiation characteristics of an antenna like radiation pattern and polarization are identical while in transmitting and receiving mode. In the following sections we discuss the similarity of the circuit characteristics of transmitting and receiving antennas.

A transmitting antenna is equivalent to an impedance (Fig. 8.27) where R is related to the power radiated by the antenna and X is related to the capacitive and inductive fields around the antenna.

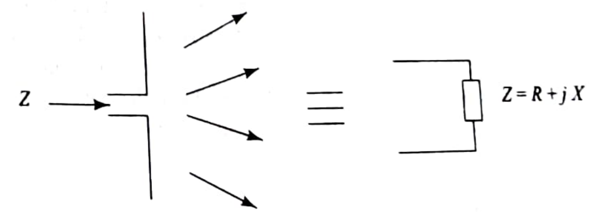


Fig. 8.27 *Transmitting antenna and its equivalent circuit.*

In the receiving mode the antenna appears like a voltage source with open circuit voltage V_{OC} and an internal impedance as $Z = R + jX$. (see Fig. 8.28).

If an impedance Z_L is connected between the receiving antenna terminals power is delivered to the load. The power is maximum when Z_L and Z are

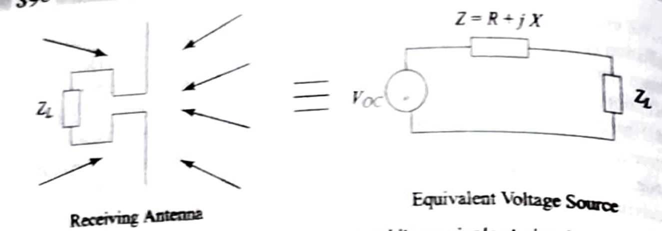


Fig. 8.28 *Receving antenna and its equivalent circuit.*

conjugate of each other, i.e. $Z_L = R - jX$. The maximum power received by the antenna and delivered to the load, therefore, is

$$P_L = \frac{|V_{OC}|^2}{4R} \tag{8.117}$$

The open circuit voltage V_{OC} depends upon the amplitude of the incident field, its polarization, direction of arrival and so on.

Let us assume that, the radiation incident on the receiving antenna has a Poynting vector S. That is, the incident wave has a power density of S Watts/m^2. The receiving antenna taps P_L power from the wave. The antenna therefore has an effective aperture, A_e where

$$A_e = \frac{P_L}{S} \tag{8.118}$$

The effective aperture is a parameter special to the receiving antenna. It describes the power capturing ability of a receiving antenna. Higher the effective aperture, more is the power delivered to the conjugately matched load.

Now consider a system of two antennas, one transmitting and other receiving, separated by a distance r as shown in Fig. 8.29. Deliberately we have not assumed any specific type of antennas as the discussion is true for any type of antennas.

The antennas 1 and 2 have internal impedances of Z_1 and Z_2 respectively. Antenna 2 is connected to a conjugate load Z_2^*. From the circuit point of view we may define the interaction between the two antennas through a mutual impedance Z_{mutual}. The open circuit voltage V_{OC} therefore can be written as

$$V_{OC} = Z_{\text{mutual}} I_1 \tag{8.119}$$

Substituting in Eqn (8.117), we get the power received by antenna 2 and delivered to the load Z_2^* as

$$P_L = \frac{|Z_{\text{mutual}}|^2 |I_1|^2}{4R_2} \tag{8.120}$$

The power transmitted by antenna 1 is

$$P_t = |I_1|^2 R_1 \tag{8.121}$$

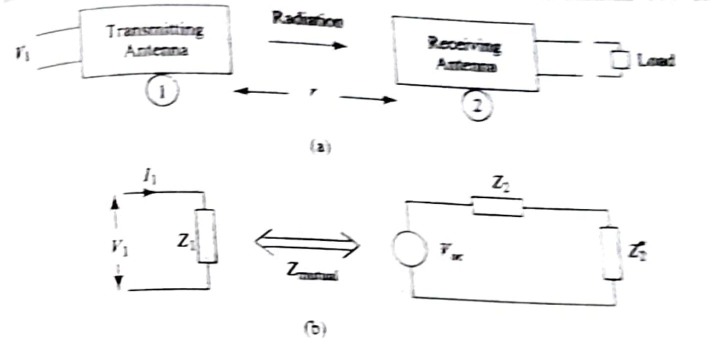


Fig. 8.29 *System of transmitting and receiving antennas.*

Now, if we assume that, the transmitting antenna has a gain of G_1 and the two antennas are aligned in the direction of their maximum radiation, the power density due to antenna 1 at the location of antenna 2 is given as

$$S = \frac{P_t G_1}{4\pi r^2} = \frac{|I_1|^2 R_1 G_1}{4\pi r^2} \tag{8.122}$$

If the effective aperture of antenna 2 is A_{e2}, the power received by the antenna is

$$P_r = S A_{e2} = \frac{|I_1|^2 R_1 G_1}{4\pi r^2} A_{e2} \tag{8.123}$$

Since, the power received by the antenna P_r is same as the power delivered to the load, P_L, we have from Eqns (8.120) and (8.123)

$$\frac{|Z_{\text{mutual}}|^2 |I_1|^2}{4R_2} = \frac{|I_1|^2 R_1 G_1}{4\pi r^2} A_{e2} \tag{8.124}$$

$$\Rightarrow \quad |Z_{\text{mutual}}|^2 = \frac{R_1 R_2 G_1 A_{e2}}{\pi r^2} \tag{8.125}$$

Now, let us interchange the roles of antennas 1 and 2, i.e. make antenna 1 the receiving antenna and antenna 2 the transmitting antenna. Using reciprocal property of the antennas, the mutual impedance Z_{mutual} in this case is same as that in the previous case. However, in this case we have

$$|Z_{\text{mutual}}|^2 = \frac{R_1 R_2 G_2 A_{e1}}{\pi r^2} \tag{8.126}$$

where G_2 is the directive gain of antenna 2, and A_{e1} is the effective aperture of antenna 1.

From Eqns (8.125) and (8.126) we obtain a very important result, i.e.

$$\boxed{G_1 A_{e2} = G_2 A_{e1}} \tag{8.127}$$

$$\Rightarrow \quad \boxed{\frac{G_1}{A_{e1}} = \frac{G_2}{A_{e2}}} \tag{8.128}$$

Since, in our discussion, we have not assumed specific antennas, Eqn (8.128) is true for every antenna pair, and every separation between them. This is possible only when

$$\frac{G_1}{A_{e1}} = \frac{G_2}{A_{e2}} = \text{constant } K \text{ (say)} \tag{8.129}$$

Constant K has to be independent of antenna parameters. Since, K is independent of type of antenna, it can be evaluated by determining the gain and the effective aperture of any antenna, and the simplest would be the Hertz dipole. We, therefore, evaluate K from the parameters of the Hertz dipole.

Gain of the Hertz Dipole The normalized radiation pattern of the Hertz dipole is given as (see Eqn (8.74)).

$$E(\theta) = \sin\theta$$

Assuming lossless dipole, we have gain equal to the directivity, i.e.

$$G = D = \frac{4\pi}{\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} \sin^2\theta . \sin\theta d\theta d\phi} \tag{8.130}$$

$$G = \frac{3}{2} \tag{8.131}$$

Effective Aperture of Hertz Dipole Let the Hertz dipole of length l be placed in an radiation electric field E. The power density of the wave is

$$S = \frac{|E|^2}{\eta} = \frac{|E|^2}{120\pi} \tag{8.132}$$

The maximum voltage developed between antenna terminals (from Eqn (8.115))

$$V_{OC} = El \tag{8.133}$$

Since, the radiation resistance of the Hertz dipole is $R_{rad} = 80\pi^2(\frac{l}{\lambda})^2$ (from Eqn (8.73)), the maximum power delivered to the matched load is

$$P_L = \frac{V_{OC}^2}{4R_{rad}} = \frac{|E^2|l^2}{4 \times 80\pi^2}\left(\frac{\lambda}{l}\right)^2 \tag{8.134}$$

Substituting for S and P_L from (8.132) and (8.134) in Eqn (8.118) we get the effective aperture of the Hertz dipole as

$$A_e = \frac{P_L}{S} = \frac{|E^2|l^2}{320\pi^2}\frac{120\pi}{|E^2|}\left(\frac{\lambda}{l}\right)^2$$

$$= \frac{3}{8}\frac{\lambda^2}{\pi} \tag{8.135}$$

We may make an important observation here that the effective aperture of the Hertz dipole is independent of its length and is a function of operating wavelength only. This is true only for the Hertz dipole. For any other antenna, the effective aperture depends upon the physical size of the antenna.

Substituting for G and A_e for the Hertz dipole in Eqn (8.129), we get the constant K as

$$K = \frac{G}{A_e} = \frac{\frac{3}{2}\pi}{\frac{3}{8}\lambda^2} = \frac{4\pi}{\lambda^2} \tag{8.136}$$

Equation (8.136) is a very important result as it relates the gain G which is a parameter of a transmitting antenna to the effective aperture A_e, which is a parameter of a receiving antenna.

For a general antenna therefore we have

$$\boxed{G = \frac{4\pi A_e}{\lambda^2}} \tag{8.137}$$

The gain of an antenna is directly proportional to its effective aperture. Higher the gain of the antenna while transmitting, higher is the effective aperture while receiving.

It may be pointed out here that although while deriving the Eqn (8.137) we assumed the two antennas to be aligned along the direction of their maximum radiation, the relation is true for any arbitrary direction. That is, if the directive gain of an antenna in the direction (θ, ϕ) is $G(\theta, \phi)$, and the effective aperture of the antenna for a radiation arriving from direction (θ, ϕ), is $A_e(\theta, \phi)$, we have in general

$$\boxed{G(\theta, \phi) = \frac{4\pi A_e(\theta, \phi)}{\lambda^2}} \tag{8.138}$$

EXAMPLE 8.14 An antenna is fed with 100 W power. The efficiency of the antenna is 80%. If the power radiation pattern of the antenna is

$$p(\theta) = \sin^2\theta \sin^2\phi \qquad 0 \le \theta \le \pi$$
$$0 \le \phi \le \pi$$

and zero elsewhere, find the radiation intensity in the direction of maximum radiation. Also find the power density at a distance of 10 km in the direction of the maximum radiation.

Solution:

Power radiated by the antenna

W = Power supplied to the antenna × efficiency = 100 × 0.8 = 80 W

Now we know that radiation intensity in the direction of maximum radiation

$$U_{max} = U_{av} \cdot D = \frac{80}{4\pi} D$$

Directivity
$$D = \frac{4\pi}{\int_{\phi=0}^{\pi} \int_{\theta=0}^{\pi} \sin^2\theta \sin^2\phi \sin\theta d\theta d\phi}$$

$$= \frac{4\pi}{\int_0^{\pi} \sin^2\phi d\phi \int_0^{\pi} \sin^2\theta \sin\theta d\theta}$$

$$= \frac{4\pi}{\int_0^{\pi} \left(\frac{1-\cos 2\phi}{2}\right) d\phi \int_0^{\pi} (1-\cos^2\theta) \sin\theta d\theta}$$

$$D = \frac{4\pi}{\frac{\pi}{2} \cdot \frac{2}{3}} = 12$$

The maximum radiation intensity

$$U_m = \frac{80}{4\pi} \times 12 = \frac{240}{\pi} \quad \text{W/St rad}$$

⇒ Power density along the direction of maximum radiation

$$P = \frac{U_{max}}{r^2} = \frac{240}{\pi(10^4)^2} = 0.764 \quad \mu\text{W/m}^2$$

EXAMPLE 8.15 A large circular parabolic dish antenna has the directivity of 30 dB. Find approximately the HPBW of the antenna and its effective aperture at 10 GHz.

Solution:

The wavelength of radiation $\lambda = \frac{c}{f} = \frac{3\times10^8}{10\times10^9} = 0.03$ m. Since, the dish is circularly symmetric the antenna beam will be circular. If the HPBW is θ_{HPBW} the solid angle of the beam is $\approx \frac{\pi}{4}(\theta_{HPBW})^2$.

Directivity of the antenna $D = 30$ dB $= 10^3$

Using Eqn (8.109) we get,

$$10^3 = \frac{4\pi}{\frac{\pi}{4}(\theta_{HPBW})^2}$$

$$\Rightarrow \qquad \theta_{HPBW} = \frac{4}{\sqrt{1000}} = 0.1265 \text{ rad}$$

The effective aperture using (8.137) is

$$A_e = \frac{G\lambda^2}{4\pi} = \frac{1000 \times (0.03)^2}{4\pi} = 0.07165 \text{ m}^2$$

Note : Here we have assumed $G = D$, since, the efficiency is assumed to be 100%.

8.6 MONOPOLE ANTENNA

At low frequencies the wavelength becomes excessively large and consequently the length of the dipole antenna also increases. At medium waves where the wavelength is of the order of few hundred meters, realization of a dipole antenna becomes practically difficult. A monopole antenna above a ground plane is used in these situations. The radiation characteristics of a monopole antenna are identical to that of a dipole antenna of double the length except the power radiated by the monopole is half of that of the corresponding dipole. Since, the power can be increased by increasing the excitation current, a mono-pole antenna can appropriately replace a dipole antenna. Most of the radio broadcasting stations at medium waves use monopole antenna vertically erected above a conducting ground. Also, this antenna is commonly used in walkie-talkie hand sets.

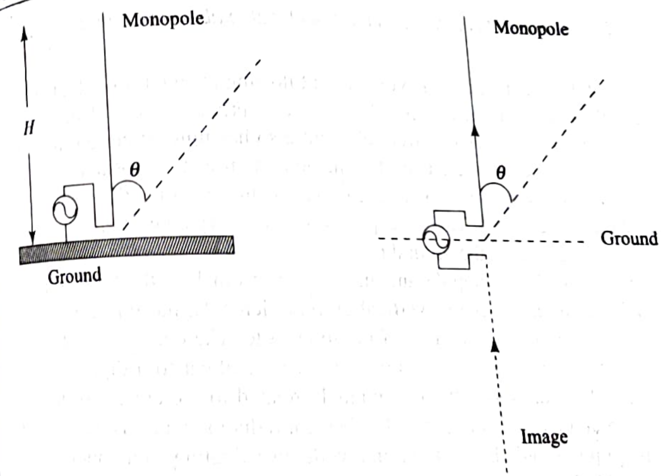


Fig. 8.30 *Vertical mono-pole antenna above a ground surface and its equivalent dipole antenna.*

A monopole antenna of length H is shown in Fig. 8.30. A voltage is applied between the bottom of the antenna and the ground.

Now using the concept of images, the ground plane is equivalent to having an image of the monopole below the ground surface. Every charge above a conducting ground has an image charge of opposite sign below the ground. The images of charges and currents above an ideal ground plane are as shown in Fig. 8.31.

It should be noted from Fig. 8.31 that at every instant of time a charge above the ground has an image charge of opposite sign below the ground. Since, the current is due to motion of charges, a current element above the ground will have identical current element below the ground. However, it should be noted that for

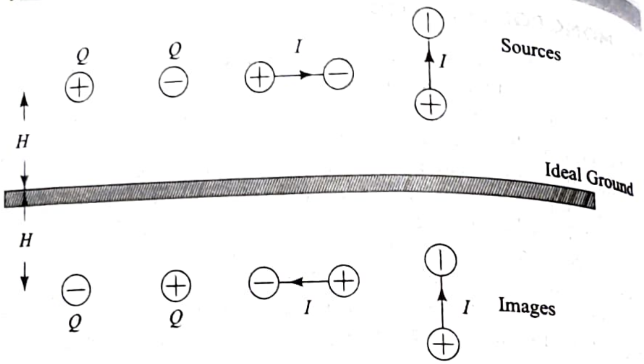


Fig. 8.31 *Images of charges and currents located above a ground surface.*

a horizontal current element having current flowing from left to right, the image current flows from right to left, whereas for a vertical current element having upward current, the image current element also has upward current flow. That is, for horizontal current element, the image current is in opposite direction but for vertical current element the image current is in same direction as that of the original current element. The concept of images is very convenient in analysing antennas located above the ground plane.

Returning to the monopole antenna, we see from Fig. 8.30 that since, the monopole is an enesemble of vertical current elements, the image has current flowing in same direction as that of the monopole. The current distribution on a monopole of length H and its image is then identical to a dipole of length $2H$. Since, the radiation pattern is uniquely related to the current distribution, a monopole and a dipole have exactly identical radiation patterns. However, the following points should be kept in mind while investigating a monopole antenna vis-a-vis a dipole antenna.

1. The radiation of a dipole is in infinite space, whereas, the radiation from a monopole is in semi-infinite space (above the ground only). Consequently, the range of θ is 0 to π for the dipole but 0 to $\frac{\pi}{2}$ for the monopole. There is no radiation for $\pi/2 < \theta < \pi$ from a monopole antenna.
2. For identical currents on a monopole and corresponding dipole, the power radiated by the monopole is only half of that of the corresponding dipole since the power is radiated only in the upper hemisphere of the space.
3. Since the power radiated by the monopole is half for a input current same as that of the corresponding dipole, the input impedance of the monopole is half of that of the dipole.

Figure 8.32 shows radiation pattern of a $\frac{3\lambda}{4}$ monopole above an ideal ground plane. The radiation pattern is identical to a $\frac{3\lambda}{2}$ dipole as shown in Fig. 8.32. (Thick

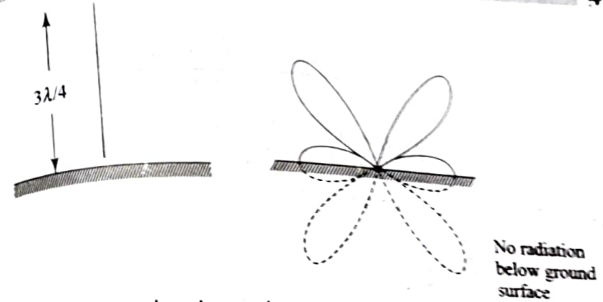


Fig. 8.32 *A mono-pole antenna above a ground surface and its radiation pattern.*

line shows the radiation pattern of the monopole antenna whereas, the thick and dotted lines together show the radiation pattern of a $\frac{3\lambda}{2}$ dipole.)

8.7 HALF-WAVELENGTH DIPOLE ANTENNA

A half-wavelength dipole is a commonly used antenna in high frequency communication systems. This antenna offers many advantages like reasonable size, single beam radiation pattern and more importantly comfortable input impedance.

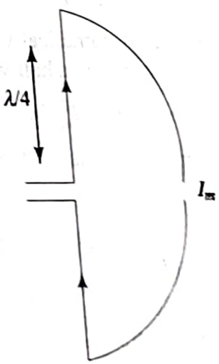


Fig. 8.33 *Current distribution on a half-wavelength dipole antenna.*

For a half wave dipole $H = \frac{\lambda}{4}$ (length of the dipole is $2H$) and $\beta H = \pi/2$. The current distribution is as shown in Fig. 8.33. From Eqns (8.83) and (8.85), the radiation electric field of a $\lambda/2$ dipole is given as

$$E_\theta = j60\frac{I_m e^{-j\beta R}}{R}\frac{\cos(\frac{\pi}{2}\cos\theta) - \cos(\frac{\pi}{2})}{\sin\theta} \tag{8.139}$$

$$= j60\frac{I_m e^{-j\beta R}}{R}\frac{\cos(\frac{\pi}{2}\cos\theta)}{\sin\theta} \tag{8.140}$$

The power density of the radiated wave is

$$P_r = \frac{1}{2}\frac{|E_\theta|^2}{\eta}$$

$$= \frac{1}{2}\left[\frac{1}{120\pi}\left(\frac{60 I_m \cos(\frac{\pi}{2}\cos\theta)}{R\sin\theta}\right)^2\right] \tag{8.141}$$

The total power radiated by the dipole is (on the lines similar to that for the Hertz dipole)

$$W = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} P_r R^2 \sin\theta\, d\theta\, d\phi \tag{8.142}$$

$$= 2\pi\int_{\theta=0}^{\pi}\frac{15}{\pi}I_m^2\frac{\cos^2(\frac{\pi}{2}\cos\theta)}{\sin^2\theta}\sin\theta\, d\theta \tag{8.143}$$

$$= 30 I_m^2 \int_0^{\pi}\frac{\cos^2(\frac{\pi}{2}\cos\theta)}{\sin\theta}d\theta \tag{8.144}$$

Integral in Eqn (8.144) is evaluated numerically and its value is 1.218. We, therefore, get the total radiated power from a half wavelength dipole as

$$\boxed{W = 36.54 I_m^2 \text{ W}} \tag{8.145}$$

Since, the peak current at the input terminal of the half-wave dipole is I_m, the input resistance of the dipole is

$$\boxed{R_{in} = \frac{W}{(I_m/\sqrt{2})^2} = 73.1\ \Omega} \tag{8.146}$$

The input resistance of a half wavelength dipole therefore is $\approx 73\ \Omega$. Generally, there is a small reactance at the input terminals of the dipole which depends upon the thickness of the dipole and the gap between the two halves of the dipole. In practice, the dipole length is slightly shortened to make the reactance zero. A practical dipole therefore is of length 0.47λ. Since, the antenna input resistance of a half wave dipole is 73 Ω, most of the antenna related cables are standardized to 75 Ω characteristic impedance rather than 50 Ω.

From Eqn (8.140) we can see that, the radiation pattern of a dipole is isotropic in the H-plane (since E_θ is not a function of ϕ) and it is a figure of infinity in the E plane as shown in Fig. 8.34.

In the E-plane, the radiation is maximum along $\theta = \frac{\pi}{2}$. The direction θ_{3dB} in which the field reduces to $\frac{1}{\sqrt{2}}$ of its maximum value, can be obtained making

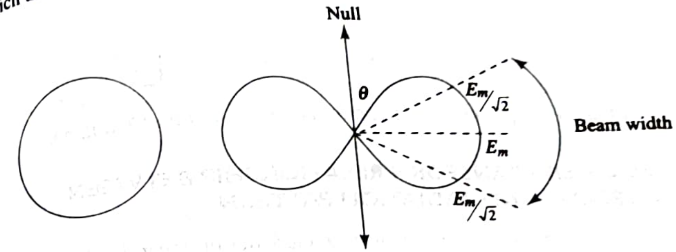


Fig. 8.34 *H-plane and E-plane radiation patterns of a half-wavelength dipole antenna.*

$$\frac{\cos(\frac{\pi}{2}\cos\theta_{3dB})}{\sin\theta_{3dB}} = \frac{1}{\sqrt{2}} \tag{8.147}$$

The HPBW of the antenna from Eqn (8.147) is obtained as 78°.

The directivity of the half wavelength dipole can be obtained from Eqn (8.106) as

$$D = \frac{4\pi}{\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\frac{\cos^2(\frac{\pi}{2}\cos\theta)}{\sin^2\theta}\sin\theta\, d\theta\, d\phi} \tag{8.148}$$

$$= \frac{4\pi}{2\pi\times 1.218} = 1.64 \tag{8.149}$$

In dB the directivity of $\lambda/2$ dipole is $10\log 1.64 = 2.15$ dB.

The effective aperture of a half wavelength dipole is

$$\boxed{A_e = \frac{\lambda^2 D}{4\pi} = 1.125\times 10^{-3}\lambda^2} \tag{8.150}$$

As mentioned above, the radiation characteristics of a $\frac{\lambda}{4}$ monopole above the ground can be obtained from the $\frac{\lambda}{2}$ dipole. The maximum radiation for the monopole is along the ground and it is vertically polarized (See Fig. 8.35).

The directivity of the monopole is 1.64 or 2.15 dB. It is important to note here that from a $\frac{\lambda}{4}$ monopole, the radiation is maximum and uniform in all directions on the ground plane. As we go above the plane, the radiation decreases and right above the monopole the radiation is zero. In radio broadcasting since the reception is primarily required on the surface of the earth, a vertical monopole antenna appears a natural choice for medium wave radio broadcasting.

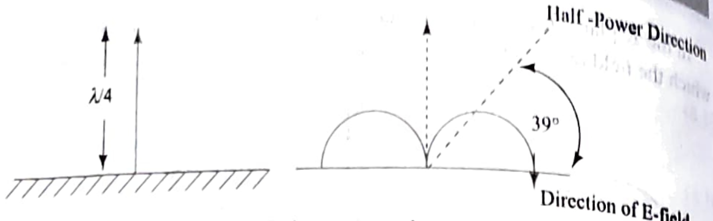


Fig. 8.35 *E-plane radiation pattern of a λ/4 mono-pole antenna.*

8.8 FOURIER TRANSFORM RELATIONSHIP BETWEEN CURRENT AND RADIATION PATTERN

In the previous sections we studied the characteristics of linear dipole/monopole antennas, by modeling them as collection of Hertz dipoles. The current distribution on these dipoles was sinusoidal. At this point it is worthwhile to ask a question - "what is the general relationship between the current distribution and the radiation pattern of an antenna?" If we can establish a general relationship between the current distribution and the radiation pattern, not only will we be able to easily visualize the radiation patterns of different type of antennas, but will also be able to predict the current distribution for a given radiation pattern. In other words, we will be able to design a current distribution to realize a required radiation pattern.

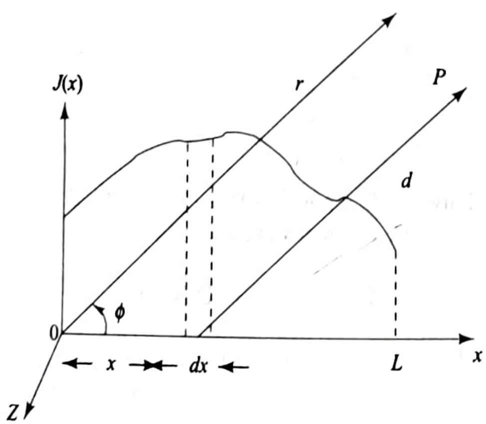


Fig. 8.36 *Continuous current distribution.*

Let us assume a continuous current distribution with linear current density $J(x)$ along x direction as shown in Fig. 8.36. The current distribution is complex in general, i.e. it has a magnitude and a phase.

$$J(x) = |J(x)|e^{j\delta(x)} \qquad 0 \le x \le L$$
$$= 0 \qquad \text{otherwise} \qquad (8.151)$$

In practice these types of current distributions are induced along metal sheets, dielectric sheets, metal wires, apertures, etc. when they are placed in an electromagnetic field. Alternatively, one may consider closely spaced antenna elements to appear almost like a continuous current sheet. For example, the current distribution in Fig. 8.36 can be due to large number of Hertz dipoles of length 'l' oriented perpendicular to the plane of the paper and placed along the x axis. The Hertz dipoles can be excited with appropriate currents to mimic the current density $J(x)$ along the x axis.

Let us now consider a small current element of width dx located at a distance x from the origin. Since, the direction of the current flow is perpendicular to the plane of the paper, the current element dx has isotropic radiation pattern in the plane of the paper. Now let us consider an observation point P at a far away distance r in the direction ϕ from the x axis. Here, we measure the distance r from the origin. The distance of point P from the current element dx then is

$$d = r - x\cos\phi \qquad (8.152)$$

The current in the current element is $J(x)dx$. The radiation field due to the current element of width dx and length l can be written from Eqn (8.77) as

$$dE = j\frac{\eta J(x)dx\, l\beta e^{-j\beta d}}{4\pi d} \qquad (8.153)$$

Since $r >> x$, we can approximately take $d = r$ in the denominator of Eqn (8.153). The approximation can not be applied to the term in the exponent. As mentioned earlier, this is due to the fact that for $r >> x$, the amplitude of E does not vary significantly with x but the phase (which is $\beta x\cos\phi$) has large variation if x is comparable to λ. Equation (8.153) therefore can be written as

$$dE = j\frac{\eta\beta l}{4\pi r}e^{-j\beta(r - x\cos\phi)}J(x)dx \qquad (8.154)$$

The total radiation field at point P due to whole current distribution can be obtained by integrating from $x = 0$ to $x = L$ as

$$E(\phi) = \int_0^L j\frac{\eta\beta l}{4\pi r}e^{-j\beta r}J(x)e^{j\beta x\cos\phi}dx \qquad (8.155)$$

$$= j\frac{\eta\beta l e^{-j\beta r}}{4\pi r}\int_0^L J(x)e^{j\beta x\cos\phi}dx \qquad (8.156)$$

For a given distance r, the quantity outside the integral sign is constant, say K. Since, $J(x)$ is zero for $x < 0$ and $x > L$, even if we change the limits of integration from $-\infty$ to $+\infty$ in Eqn (8.156), the value of integral does not change. We can, therefore, write

$$E(\phi) = K\int_{-\infty}^{\infty} J(x)e^{j\beta x\cos\phi}dx \qquad (8.157)$$

$$= K\int_{-\infty}^{\infty} J(x)e^{j2\pi\left(\frac{x}{\lambda}\right)\cos\phi}dx \qquad (8.158)$$

Now, if we define $(\frac{x}{\lambda}) \equiv x'$ as the normalized distance, and $\cos\phi = p$, the direction cosine, we can write Eqn (8.158) as

$$E(p) = K\lambda \int_{-\infty}^{\infty} J(x')e^{j2\pi x' p} dx' \tag{8.159}$$

One would recognize that the integral in Eqn (8.159) is the Fourier integral. We, therefore, find that the spatial current distribution and the radiation field distribution are related through the Fourier transform. Note that the Fourier pair is not spatial length x and angle ϕ, but is the normalized distance $(\frac{x}{\lambda})$ and the direction cosine $p(\equiv \cos\theta)$.

Since, the radiation pattern is always normalized with respect to its maximum, the constant $K\lambda$ does not play any role in computation of the radiation pattern.

It should be mentioned here, that the far field radiation pattern in the plane of the paper is just due to distribution of the current. Since, the current flows perpendicular to the plane of the paper, the intrinsic current element (the Hertz dipole) has $\sin\theta$ radiation pattern.

The total radiation pattern of the current strip then is

$$E(\theta, \phi) = E(\phi)\sin\theta \tag{8.160}$$

The theory of Fourier transforms is very well developed in signal analysis. Therefore, it becomes convenient to visualize various antenna characteristics through properties and theorems of Fouriers transform. In the following sections we obtain radiation patterns of an antenna with various current distrubutions using the Fourier transform.

8.8.1 Radiation Pattern of a Uniform Current Distribution

Let us obtain the radiation pattern for a uniform current distribution. Without losing generality let the linear current density be unity over a normalized length L as shown in Fig. 8.37. Let us also assume that there is no phase variation in the current density $J(x')$. The radiation pattern is given by

$$E(p) = K_0 \int_{-\frac{L}{2}}^{\frac{L}{2}} 1 e^{j2\pi x' p} dx' = K_0 \frac{\sin(\pi p L)}{(\pi p)} \tag{8.161}$$

where K_0 is a constant.

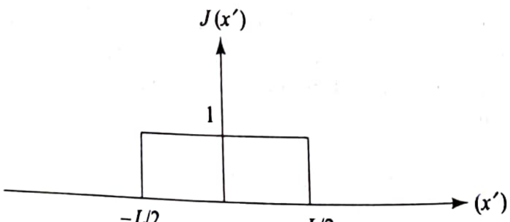


Fig. 8.37 *Uniform current distribution.*

The Maximum value of $E(p)$ will be $K_0 L$ and will correspond to $\phi = 0$. Normalizing the $E(p)$ with respect to the maximum value, we get the radiation pattern as

$$\frac{E(p)}{K_0 L} = \frac{\sin \pi p L}{\pi p L} = \text{sinc}(pL) \tag{8.162}$$

The uniform current distribution has a sinc function radiation pattern. However, remember, that p is the direction cosine and hence p lies between -1 and $+1$, whereas when we take Fourier transform of $J(x')$ we get $E(p)$ for all values of p. Therefore, only the portion of $E(p)$ corresponding to $-1 < p < +1$ gives the visible radiation pattern. $-1 < p < 1$ is called the visible range of p. Figure 8.38 shows the sinc(pL) function for the visible range of p. The function maximum at $p = 0$, i.e. $\phi = \frac{\pi}{2}$ represents the main beam. The function has zeros at $p = \pm\frac{1}{L}, \pm\frac{2}{L}$, and so on, and local maxima (i.e. side lobes) at $p = \pm\frac{3}{2}L, \pm\frac{5}{2}L$, and so on.

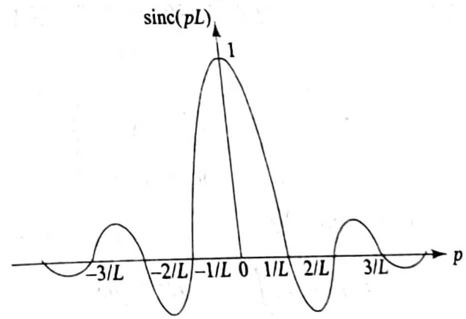


Fig. 8.38 *sinc(pL) for visible range of p.*

The zeros represent nulls of the radiation pattern and the local maxima represent the side lobes of the pattern. As can be seen from Fig. 8.38, the nulls and side lobes are equi-spaced along p.

Substitution of $p = \frac{3}{2}L, \frac{5}{2}L$, etc. in Eqn (8.162) we get the amplitudes of the side lobes as $2/3\pi$, $2/5\pi$, etc. respectively. That is, the amplitude of the first side lobe is about 21% of the peak and the second side lobe is about 13% of the peak. From Fig. 8.38, we may also note that there are approximately $2L$ nulls in the radiation pattern and width of the main beam is $\frac{2}{L}$. Larger the width of the current distribution narrower is the main beam and more are the nulls.

8.8.2 Tilting the Main Beam

Let us now consider a current distribution $J(x')$ which has uniform amplitude b linear phase. The phase of the current density $J(x')$ linearly increases from o end to other as shown in Fig. 8.39.

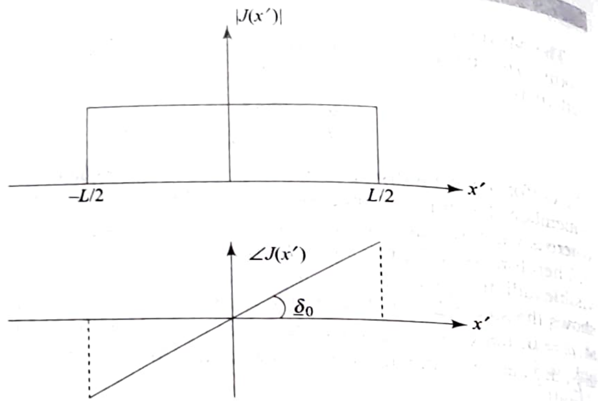


Fig. 8.39 *Uniform current distribution with a linear phase variation.*

$$J(x') = e^{j\delta_0 x'} \qquad \frac{-L}{2} < x' < \frac{L}{2}$$
$$= 0 \qquad \text{otherwise} \tag{8.163}$$

The radiation pattern is given by

$$E(p) = K_0 \int_{\frac{-L}{2}}^{\frac{L}{2}} e^{j\delta_0 x'} e^{j2\pi x' p} dx' \tag{8.164}$$

$$= K_0 \int_{\frac{-L}{2}}^{\frac{L}{2}} e^{j2\pi (p+\frac{\delta_0}{2\pi})x'} dx'$$

$$= K_0 \int_{\frac{-L}{2}}^{\frac{L}{2}} e^{j2\pi x'(p+p_0)} dx' \tag{8.165}$$

$$E(p) = E(p + p_0) \tag{8.166}$$

where $p_0 = \frac{\delta_0}{2\pi}$.

The radiation pattern $E(p)$ is shifted by $-p_0$ with respect to the original $E(p)$, that is, the direction of the main beam is shifted from 0 to $-p_0$. The shifted radiation pattern is shown in Fig. 8.40.

We, therefore, conclude that any linear phase gradient, δ_0, translates the radiation pattern along p axis by $-\delta_0/2\pi$. A tilt in the main beam therefore can be obtained by introducing a linear phase gradient in the current distribution. For a positive phase gradients, the beam tilts towards left and for a negative phase gradient the beam tilts towards right.

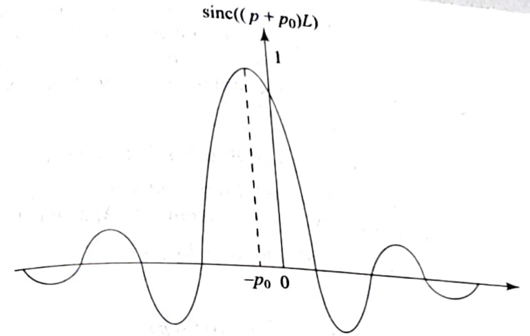


Fig. 8.40 *Shifted radiation pattern due to the current distribution in Fig. 8.39.*

8.8.3 Effect of Illumination Taper on Radiation Pattern

It is the fundamental property of the Fourier transform that any sudden change in the function or its derivatives produces large ripple in the transform domain. For a uniform current distribution, the current abruptly changes to zero at the ends of the distribution. This abrupt change appears as a ripple in the form of sidelobes in the radiation pattern. The reduction in the sidelobe level can therefore be achieved by reducing the abruptness of the change in the current at the ends of the distribution. In other words, tapering the current distribution gradually towards the ends of the distribution reduces the sidelobe level. Figure 8.41 shows the different current distributions and their respective radiation patterns.

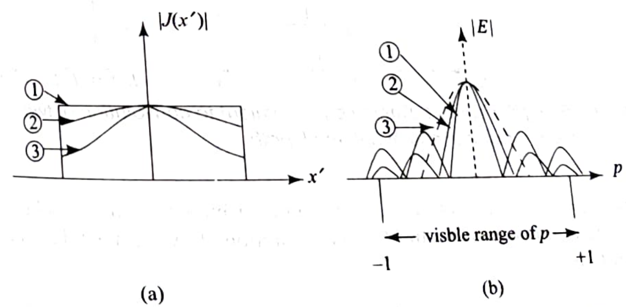


Fig. 8.41 *Tapered current distributions (a) and the corresponding radiation patterns (b).*

Where, on one side the current taper decreases the sidelobe level, on other side it widens the main beam. This is due to the fact that by tapering the current on the edges we provide higher weightage to the central region of the current distribution, or in other words, the effective width of the current distribution reduces and since the width of the main beam is inversely related to the width of the current distribution, the width of the main beam increases. It is important to note that for a fixed width of the current distribution, no sidelobe reduction can be achieved without broadening the main beam.

8.8.4 Use of convolution property of Fourier Transform

If $f_1(x)$ and $f_2(x)$ are two functions and $F_1(p)$ and $F_2(p)$ are their Fourier transforms respectively, then by convolution theorem (see Appendix F) we have

$$f_1(x)f_2(x) \leftrightarrow^{FT} F_1(p) * F_2(p) \tag{8.167}$$

and

$$f_1(x) * f_2(x) \leftrightarrow^{FT} F_1(p)F_2(p) \tag{8.168}$$

$\leftrightarrow^{FT}$ denotes the Fourier transform relationship between left and right hand sides of the equations and $*$ denotes the convolution operation defined as

$$f_1(x) * f_2(x) = \int_{-\infty}^{\infty} f_1(x')f_2(x-x')dx' \tag{8.169}$$

and

$$F_1(p) * F_2(p) = \int_{-\infty}^{\infty} F_1(p')F_2(p-p')dp' \tag{8.170}$$

The convolution property can be used to quickly obtain the radiation patterns due to non-uniform current distributions.

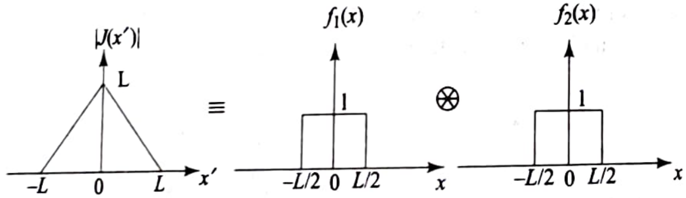


Fig. 8.42 *Triangular current distribution equivalent to convolution of two uniform current distributions of half the length.*

For example, consider a triangular current distribution as shown in Fig. 8.42. The triangular function of width $2L$ is a convolution of two rectangular functions of width L as

$$J(x) = f_1(x) * f_2(x) \tag{8.171}$$

where

$$f_1(x) = f_2(x) = \prod\left(\frac{x}{L}\right) \tag{8.172}$$

$\prod(\frac{x}{L})$ is the window function of width L centered at $x = 0$. The Fourier transforms of $f_1(x)$ and $f_2(x)$ are

$$F_1(p) = F_2(p) = \frac{\sin(\pi Lp)}{\pi Lp} \tag{8.173}$$

The radiation pattern of $J(x)$ is

$$E(p) = FT\{J(x)\} = FT\{f_1(x) * f_2(x)\} = F_1(p)F_2(p) = \left(\frac{\sin(\pi Lp)}{\pi Lp}\right)^2 \tag{8.174}$$

Using convolution property we could find the radiation pattern of a triangular current distribution without actual integration as in Eqn (8.159).

8.8.5 Radiation Pattern of Uniformly Spaced Discrete Current Elements

Let us now consider a current distribution consisting of discrete current sources distributed over length L as shown in Fig. 8.43. Without losing generality, let us assume that, the current sources are of unit amplitude and the spacing between them is 'd'.

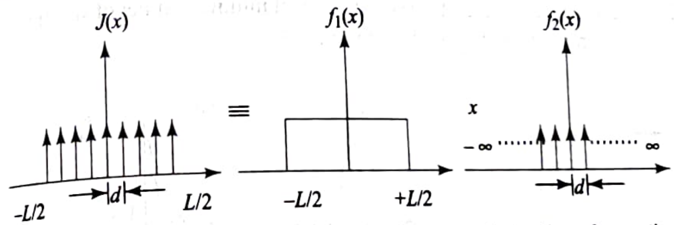


Fig. 8.43 *Discrete current distribution equivalent to multiplication of a continuous current distribution and a spatial sampling function.*

The current distribution is equivalent to a product of two functions $f_1(x)$ and $f_2(x)$ given as

$$f_1(x) = \prod\left(\frac{x}{L}\right) \tag{8.175}$$

$$f_2(x) = \sum_{K=-\infty}^{\infty} \delta(x - Kd) \tag{8.176}$$

where $\delta(x - Kd)$ is a Dirac δ function.

Since, $J(x)$ is the product of $f_1(x)$ and $f_2(x)$, the radiation pattern will be the convolution of $F_1(p)$ and $F_2(p)$.

The Fourier transform of $f_1(x)$ is

$$F_1(p) = FT\left\{\prod\left(\frac{x}{L}\right)\right\} = \frac{\sin(\pi Lp)}{\pi Lp} \tag{8.177}$$

and the Fourier transform of $f_2(x)$ is (See appendix F)

$$F_2(p) = FT\left\{\sum_{K=-\infty}^{\infty} \delta(x - Kd)\right\}$$

$$= \sum_{K=-\infty}^{\infty} \delta\left(p - \frac{K}{d}\right) \tag{8.178}$$

Since, the convolution of any function with the δ function gives translation of the function as

$$A(t) * \delta(t - u) = A(t - u) \tag{8.179}$$

The radiation pattern will be

$$E(p) = F_1(p) * F_2(p)$$

$$= \left\{\frac{\sin(\pi L p)}{\pi L p}\right\} * \left\{\sum_{K=-\infty}^{\infty} \delta(p - \frac{K}{d})\right\}$$

$$= \sum_{K=-\infty}^{\infty} \frac{\sin \pi L(p - \frac{K}{d})}{\pi L(p - \frac{K}{d})} \tag{8.180}$$

The radiation pattern therefore is collection of infinite number of sinc functions displaced by multiples of $\frac{K}{d}$ as shown in Fig. 8.44.

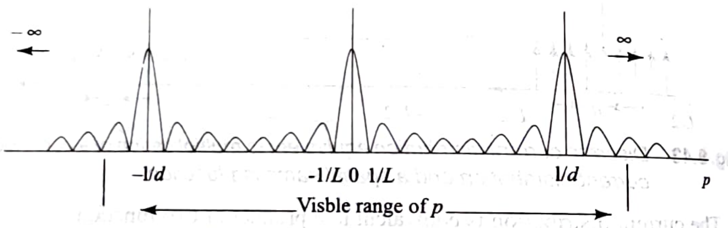


Fig. 8.44 *Radiation pattern of a discrete current distribution.*

As 'd' increases, the sinc functions move closer and their main lobes may even overlap if d is of the order of L. On the other hand if d is small, the sinc functions move apart and there is less overlap between them. If we make $d << 1$, the radiation pattern in the visible range almost approximates to a single sinc function. Since, the radiation pattern and current distribution have Fourier transform relationship, when the radiation pattern approximates to the sinc function the corresponding current distribution almost becomes a continuous function. It is then interesting to note that a discrete set of current elements behave like a continuous distribution if spacing between the elements $d << 1$. Remember that 'd' is the normalized distance with respect to the wavelength. In practice, the rule of thumb is d should be $< \frac{1}{20}$. In the reflector type antennas, like a parabolic dish, the continuous metallic reflectors are realized in the form of a mesh. The spacing of mesh is $< \frac{\lambda}{20}$, therefore, the mesh electromagnetically behaves almost like a solid surface but is transparent to the wind flow avoiding wind loading on the antenna (See Fig. 8.45).

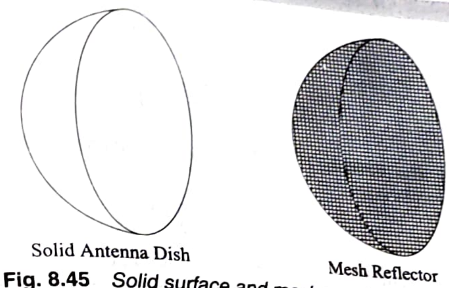


Fig. 8.45 *Solid surface and mesh parabolic reflectors.*

EXAMPLE 8.16 Two identical vertical Hertz dipole are separated by a distance of 5 m. The peak current in the dipole is 10 A and the wavelength of excitation is 20 m. Find the radiation pattern of the combined antennas.

Solution:

The current distribution consists of two δ-functions separated by a distance of 5 m. Without losing generality let us consider the origin at the mid-point of the spacing between the two antennas (see Fig. 8.46).

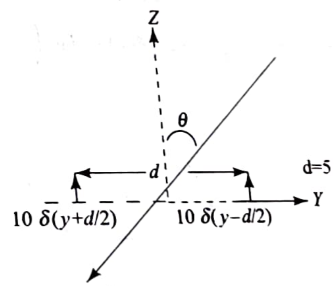


Fig. 8.46 *A two-element array.*

The current distribution can be written as

$$I(y) = 10\delta(y - d/2) + 10\delta(y + d/2)$$

Radiation pattern

$$F(\theta) = K \int_{-\infty}^{\infty} I(y) e^{j\frac{2\pi}{\lambda} y \sin\theta} dy$$

$$= 10K \int_{-\infty}^{\infty} \{\delta(y - d/2) + \delta(y + d/2)\} e^{j\frac{2\pi}{\lambda} y \sin\theta} dy$$

$$= 10K \left\{e^{+j\frac{2\pi}{\lambda}\cdot\frac{d}{2}\sin\theta} + e^{-j\frac{2\pi}{\lambda}\cdot\frac{d}{2}\sin\theta}\right\}$$

$$= 20K\cos\left(\frac{\pi d\sin\theta}{\lambda}\right)$$

$$= 20K\cos\left(\frac{5\pi\sin\theta}{20}\right)$$

Since, in the vertical plane the Hertz dipole has $\sin\theta$ pattern, the total radiation pattern in the yz-plane is

$$F(\theta) = \cos\left(\frac{\pi\sin\theta}{4}\right)\sin\theta$$

Here the radiation pattern has been normalized with $20\,K$.

EXAMPLE 8.17 Three current sources are oriented so as to individually give isotropic radiation pattern. The separation between the elements is $\lambda/3$. The current in the three elements are in the ratio of 1:2:1. Using Fourier transform properties find the radiation pattern of the combined antenna system.

Solution:

The current distribution is shown in Fig. 8.47(a). Without losing generality we can take the origin at the location of the middle source.
The current distribution then can be written as

$$I(x) = I_0\delta\left(x+\frac{\lambda}{3}\right) + 2I_0\delta(x) + I_0\delta\left(x-\frac{\lambda}{3}\right)$$

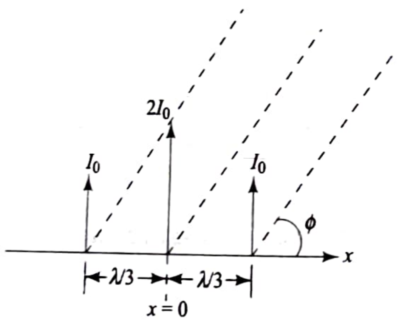


Fig. 8.47(a) *A 3-element array with binomial current distinction.*

The current distribution can be thought of as a convolution of two 2-element antennas as shown in Fig. 8.47(b).
We, therefore, have

$$I(x) = I_1(x) * I_1(x) = \left\{\delta\left(x-\frac{\lambda}{6}\right)+\delta\left(x+\frac{\lambda}{6}\right)\right\} * \left\{\delta\left(x-\frac{\lambda}{6}\right)+\delta\left(x+\frac{\lambda}{6}\right)\right\}$$

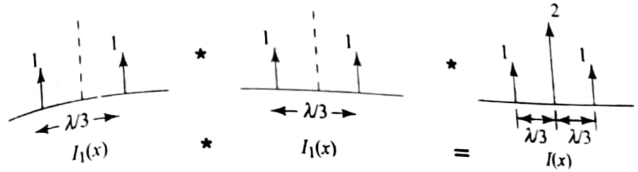


Fig. 8.47(b) *Current distribution $I(x)$ as a convolution of two current distributions $I_1(x)$ and $I_1(x)$.*

The radiation pattern

$$F(\phi) = FT\{I(x)\} = FT(I_1(x) * I_1(x))$$
$$= FT\{I_1(x)\}\cdot FT\{I_1(x)\}$$
$$= (F_1(\phi))^2$$

where $F_1(\phi)$ is the radiation pattern of current distribution $I_1(x)$.

$$F_1(\phi) = K\int_{-\infty}^{\infty}\left\{\delta\left(x-\frac{\lambda}{6}\right)+\delta\left(x+\frac{\lambda}{6}\right)\right\}e^{j\frac{2\pi}{\lambda}x\cos\phi}\,dx$$

$$= K\left\{e^{j\frac{2\pi}{\lambda}\frac{\lambda}{6}\cos\phi} + e^{-j\frac{2\pi}{\lambda}\frac{\lambda}{6}\cos\phi}\right\}$$

$$= 2K\cos\left(\frac{\pi\cos\phi}{3}\right)$$

The radiation pattern of the three element system is (after normalization)

$$F(\phi) = (F_1(\phi))^2 = \cos^2\left(\frac{\pi\cos\phi}{3}\right)$$

EXAMPLE 8.18 A vertical mono-pole antenna of 50 m height is used to broadcast an AM radio station at 200 m band. The AM station must provide a minimum electric field strength of 25 mV/m at a distance of 30 km on the ground. Find the base current and the total power radiated by the antenna.

Solution:

Height of the monopole is 50 m and the wavelength is 200 m. That is, the monopole is $\lambda/4$ long. The electric field due to a $\lambda/4$ monopole is

$$E_\theta = \frac{60I_m}{r}\frac{\cos(\pi/2\cos\theta)}{\sin\theta}$$

The maximum field is at $\theta = \pi/2$, i.e. on the ground. Hence, at $r = 30$ km,

$$\Rightarrow \quad E_{max} = \frac{60I_m}{r} = \frac{60I_m}{30\times10^3} = 25\times10^{-3} \quad \text{(given)}$$

$$\Rightarrow \quad I_m = 12.5\text{ A}.$$

For quarter wavelength monopole the base current is same as I_m. The input current to the antenna, therefore, is 12.5 A.

The monopole antenna has input resistance of 36.5 Ω. (Note: The input resistance of a $\lambda/4$ monopole is half of that of a $\lambda/2$ dipole). The power radiated by the monopole then is

$$P = \frac{1}{2} I_m^2 R_{in} = \frac{1}{2}(12.5)^2 \times 36.5$$
$$= 2.85 \text{ kW}$$

EXAMPLE 8.19 The current distribution on a linear aperture is given as

$$I(x) = 10\cos\left(\frac{\pi x}{L}\right) \qquad -L/2 \le x \le +L/2$$
$$= 0 \qquad \text{Otherwise}$$

Find the radiation pattern of the aperture. What is the radiation pattern for $L = 3\lambda$?

Solution:

The current distribution and the radiation pattern have Fourier transform relation. The electric field therefore is

$$E(p) = K \int_{-L/2}^{L/2} 10\cos\left(\frac{\pi x}{L}\right) e^{j\frac{2\pi}{\lambda}xp} dx$$

Where, $p \equiv \cos\phi$; ϕ is the angle measured from the $+x$ direction, and K is constant.

Expanding the cos function, we get

$$E(p) = 10K \int_{-L/2}^{L/2} \left(\frac{e^{j\frac{\pi x}{L}} + e^{-j\frac{\pi x}{L}}}{2}\right) e^{j\frac{2\pi}{\lambda}xp} dx$$
$$= \frac{20KL\lambda^2}{\pi} \frac{\cos(\pi Lp/\lambda)}{(\lambda^2 - 4p^2L^2)}$$

The normalized radiation pattern is

$$E_n(p) = \lambda^2 \frac{\cos(\pi Lp/\lambda)}{(\lambda^2 - 4p^2L^2)}$$

For $L = 3\lambda$ the radiation pattern is

$$E_n(p) = \frac{\cos(3\pi p)}{(1 - 36p^2)}$$

EXAMPLE 8.20 The signals from the two receceiving antennas are added to get the output. The antennas have 2λ linear apertures and uniform current distributions as shown in Fig. 8.48. The antennas are placed in a colinear geometry with a separation of 10λ. Find the effective radiation pattern of the receiving antenna system.

Solution:

The current distribution on the total system can be visualized as

$$I(x) = \sqcap\left(\frac{x}{2\lambda}\right) * \{\delta(x - 5\lambda) + \delta(x + 5\lambda)\}$$

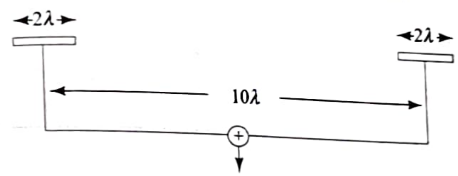


Fig. 8.48 *Interferometer of two aperture antennas.*

Taking Fourier transform of the current distribution $I(x)$, we get the radiation pattern

$$E(p) = FT\left\{\sqcap\left(\frac{x}{2\lambda}\right)\right\} \times FT\{\delta(x - 5\lambda) + \delta(x + 5\lambda)\}$$

Since, the normalized length of the antenna is 2, the radaition pattern is

$$E(p) = \frac{\sin(2\pi p)}{2\pi p} \times 2\cos(10\pi p)$$

EXAMPLE 8.21 Find the radiation pattern of a 2λ long thin travelling wave antenna. Also find the directions of the nulls for the antenna and approximate direction of the main beam.

Solution:

The current distribution on a travelling wave antenna is given as

$$I(z) = I_0 e^{-j\beta z}$$

Then using the same approach that used for a linear antenna, the electric field is given as

$$E(\theta) = K \int_0^{2\lambda} I_0 e^{-j\beta z} e^{j\beta z\cos\theta} \sin\theta dz$$

After some algebraic manipulations we get radiation pattern of the antenna as

$$|E(\theta)| = 2|KI_0| \sin\left(\frac{2\pi(\cos\theta - 1)}{\beta(\cos\theta - 1)}\right) \sin\theta$$

The radiation pattern has nulls when

$$\theta = \pm m\pi$$

and

$$2\pi(\cos\theta - 1) = \pm m\pi$$

$$\Rightarrow \qquad \cos\theta = 1 \pm \frac{m}{2}$$

where, m is an integer.

The nulls therefore are $\theta = 0^\circ, 60^\circ, 90^\circ, 120^\circ, 180^\circ$.

The main beam is approximately half way between first two nulls, 0° and 60°.

The main beam direction is approximately 30°.

8.9 SUMMARY

A very important component of a communication system, the antenna has been studied in this chapter. It has been argued that time varying currents are potentially capable of producing radiation. The basic mathematical frame work of antennas have been developed in this chapter. The simplest form of antenna, the Hertz dipole has been investigated. The Hertz dipole is an element in which any complex antenna structure can be segmented into. The analysis of the Hertz dipole therefore provides the foundation for any complex antenna structure. The characteristics of more practical dipoles have been investigated further. The most fundamental Fourier relation between the radiation pattern and spatial current distribution on an antenna has been established.

Antenna design is one of the most challenging tasks in the design of a communication link. The most passive looking structure, antenna, has made the communication over thousands of km possible with small power. Without proper antennas the systems like satellite communication links, TV and radio broadcasting, radars, microwave links, etc. would not be a reality.

Review Questions

8.1 What is radiation?

8.2 What are the essential conditions for radiation?

8.3 Can one have radiation from DC current? Why?

8.4 Why does a transmission line not radiate?

8.5 What is magnetic vector potential?

8.6 What is Lorentz gauge condition?

8.7 What is the reasoning used to choose the Lorentz gauge condition the way it has been chosen?

8.8 For analysing radiation what is the appropriate geometry and why?

8.9 What are the constant phase surfaces for a radiation from a point source? What is this wave called?

8.10 What is a Hertz dipole? What is the current distribution of the Hertz dipole?

8.11 How can current flow in a piece of wire when it is not connected on either ends?

8.12 What type of fields are generated due to Hertz dipole?

8.13 At what distance from the Hertz dipole do the electrostatic, induction and radiation fields become equal?

8.14 Why are the electrostatic, induction and radiation fields not in time phase?

8.15 Why does electrostatic fields increase as the frequency reduces?

8.16 What field components does the electrostatic field due to a Hertz dipole have?

8.17 For a radiation field what is the relation between electric and magnetic fields?

8.18 Draw the field patterns for the electrostatic, induction and radiation fields.

8.19 For a spherical wave with electric field E, what is the power density and the direction of power flow?

8.20 For a Hertz dipole if the wavelength is reduced by factor of 2 and the current is kept same, by what factor will the radiated power increase?

8.21 What is radiation resistance? How does it vary as a function of dipole length?

8.22 Why are radiation patterns of a dipole and monopole of half the dipole length identical?

8.23 How does the terminal impedance change as a function of dipole length?

8.24 Explain physically why we get more nulls when the dipole length increases? Visualize 3-D radiation pattern for a 2λ dipole.

8.25 What is radiation intensity and what is its unit?

8.26 What is directivity of an antenna?

8.27 What is the difference between directivity and directive gain of an antenna?

8.28 What is side lobe level of an antenna?

8.29 What is effective aperture of an antenna?

8.30 What is circuit model of a receiving antenna?

8.31 What is the relation between directivity and effective aperture of an antenna?

8.32 What is the phase difference between the currents of a current-element and its image when (a) the current is vertically paced above an ideal ground (b) the current element is horizontally placed above an ideal ground?

8.33 What is the radiation resistance of a $\frac{\lambda}{2}$ dipole?

8.34 What is HPBW of a $\frac{\lambda}{2}$ dipole?

8.35 What is the effective aperture of a $\frac{\lambda}{2}$ dipole?

8.36 What is the relationship between spatial current distribution and the radiation field pattern?

8.37 What is the radiation pattern of a uniform current distribution?

8.38 What happens to the radiation pattern if a phase gradient is introduced in the current distribution?

8.39 What is the effect of tapering of the current distribution on the radiation pattern?

8.40 If the signals from two antennas are multiplied what will be the effective radiation pattern?

8.41 How does a discrete current distribution behave almost like a continuous current distribution?

Problems

8.1 A z-oriented Hertz dipole of 0.1λ length at 10 MHz is located at the origin and excited with 100 A peak current. Find the vector electric and magnetic fields at a distance of 1 m, 100 m and 10 km along the z-axis, and in the xy-plane.

8.2 For a 1 MHz Hertz dipole find the distance beyond which the reactive fields are less than 10% of the radiation fields.

8.3 A Hertz dipole of length 5cm is excited with 5MHz signal. The rms value of the signal is 2 A. Find the maximum accumulation of the electric charge on the tips of the dipole. Hence or otherwise find the electrostatic field due to the dipole at a distance of 1m from the dipole.

8.4 The electric field at some distance r due to an antenna is given as

$$\mathbf{E} = 10e^{-j\beta r}\left\{\frac{\cos\theta}{r^2}\hat{\mathbf{r}} + 0.2\frac{\sin\theta}{r}\hat{\theta}\right\}$$

Find the vector magnetic field at that location.

8.5 A Hertz dipole of 10 cm length has 5 A (rms) current at 100 MHz. What is the total power radiated by the dipole? What is the power density at a distance of 10km from the dipole along a line perpendicular to the dipole.

8.6 A vertical dipole of length 1 m is excited at 1 MHz. The elctric field at a distance of 20 km from the antenna in the horizontal plane passing through it, is 10 μV/m. Find the peak excitation current of the antenna.

8.7 For vertical Hertz dipole the Poynting vector at a distance of 10 km along 30° elevation line is 10 mW/m^2. Find the total power radiated by the antenna.

8.8 Two identical Hertz dipoles, one vertical and other at an angle of 45° from the vertical are co-located. The excitation currents for the two antennas are $10\angle 0$ and $10\angle 45^\circ$ A respectively. Find the parameters of the ellipse of polarization for the combined wave along a direction perpendicular to the plane containing the antennas.

8.9 Two Hertz dipoles, one along $\theta = 0$ and other along $\theta = \pi/2, \phi = 0$, are excited with equi-amplitude currents with 90° phase difference. Find the radiation pattern in the E-plane which is common to both the antennas. What is state of polarization of the wave in directions (i) $\theta = 0$ (ii) $\theta = \phi = \pi/2$ (iii) $\theta = \pi/2, \phi = 0$.

8.10 Find the maximum electric and the magnetic field strengths which one would get from a $\lambda/2$ dipole at 1 GHz at a distance of 10 km from the dipole. The dipole is excited with 10 A rms current.

8.11 For antenna in Problem 8.10, find the total power radiated by the antenna.

8.12 Assume that a $\lambda/3$ dipole has triangular current distribution with maximum current

at the center and zero current at the ends of the dipole. Derive an expression for the radiation field of the antenna.

8.13 A linear 20 m long travelling wave antenna has current along its length as

$$I(z) = I_0 e^{-j2\pi z}$$

The frequency of the current is 100 MHz. Find the radiation pattern of the antenna.

8.14 A 10 m long antenna is excited at its center at 100 MHz. If the input current is 20 A, find the maximum current seen on the antenna. Also find the directions of the maximum radiation and nulls for the antenna.

8.15 At 10 MHz, a z-oriented half-wavlength dipole located at the origin radiates 100 W power. At a distance of 10 km along the x-axis a z-oriented dipole of 2 m length is placed. What is the open circuit voltage between the terminals of the dipole?

8.16 In Problem 8.15, if the receiving dipole is connected to a load of 50—j10 Ω, How much power will be delivered to the load?

8.17 Find directivity of a one-wavelength long dipole.

8.18 An antenna has $\cos^3\theta$ radiation pattern, where θ is the angle measured from the z-axis in the spherical coordinate system. Find the HPBW and directivity of the antenna.

8.19 If the antenna in Problem 8.18 radiates 100 W power, find the radiation intensity along $\theta = 0$ direction. Also find the electric field intensity at a distance of 5 km from the antenna along $\theta = 0$ direction.

8.20 The radiation pattern of an antenna is $\cos 2\theta$. Find the HPBW in the θ-plane and the directivity of the antenna.

8.21 A satellite antenna has elliptical conical beam. The HPBWs along the two principal axes of the ellipse are 2° and 3° respectively. The antenna is supplied with 10 W power. Find the power density of the radiated wave on the earth. The height of the satellite is 36000 km.

8.22 An antenna has 0.5° wide circular beam. If the efficiency of the antenna is 80%, find the power gain of the antenna.

8.23 An antenna has the directive gain of 50 dB. Find the effective aperture of the antenna.

8.24 An antenna has the directivity of 15 dB. The antenna when excited with 100 V(rms) source, produces radiation intensity of 50 W/Sr in the direction of the maximum radiation. Find the input resistance of the antenna.

8.25 Two parabolic dish antennas having radiation patterns $\cos^2\theta$ and $\cos^3\theta$ are mounted on microwave towers to look into each other. θ is the angle measured from the line joining the two antennas. If 25 W of power at 2.4 GHz is supplied to one of the antennas, how much power will be received by the other antenna? Assume the antennas are perfectly matched and the separation between them is 50 km.

8.26 Find the radiation pattern of an antenna having uniform current distribution over 2λ length. Find the directions of the nulls with respect to the direction of maximum radiation.

8.27 A monopole antenna of 50 cm height above a ground plane is used for radio receiption in a car at 500 kHz AM band. Find the input resistance of the antenna.

8.28 Find the radiation pattern for the current distribution shown in Fig. 8.50. Use Fourier transform properties.

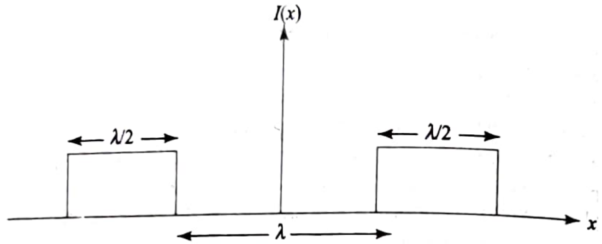


Fig. 8.50 *Spatial current distribution.*

8.29 The current distribution on an aperture antenna of length L is

$$I(x) = \cos^2(\pi x/L) \qquad -L/2 \le x \le L/2$$
$$= 0 \qquad \text{otherwise}$$

Find the radiation pattern, HPBW and the nulls of the antenna. Compare these parameters with that of uniform aperture antenna of same length.

CHAPTER 9

Antenna Arrays

In the previous chapter, we investigated the radiation characteristics of monopole and dipole antennas. These antennas generally have broad radiating beams and consequently low directivity. The directivity of these antennas can be increased by increasing their physical dimensions. However, an increase in the size of an antenna not only increases the directivity but the radiation pattern as a whole gets modified in undesirable fashion. Also, while controlling the radiation pattern, the terminal characteristics of the antenna might get altered in an unpleasant fashion. In fact, one does not know precisely how a radiation pattern is modified by shaping a particular type of antenna. We can only find the radiation pattern for a given antenna structure, (this too is not very simple for complex antenna shapes) but finding an antenna shape to realize a given radiation pattern is rather an impossible task. In practice, depending upon the application, the radiation pattern for an antenna is specified and the engineer has to design the radiating structure. This inverse problem is rather impossible to handle with single antenna.

Referring to the fundamental Fourier relationship between the current distribution on an antenna and the radiation pattern, we can say that, if, we control the current distribution on the antenna surface, we will be able to realize the desired radiation characteristic. Unfortunately, for the antennas discussed earlier, one does not have any control over current distribution. Once the antenna is excited by a voltage or current source at its feed point, the current distribution is automatically adjusted to satisfy the basic laws of electromagnetics. It is then clear that for better control of radiation pattern, one must find a mechanism of controlling spatial distribution of the current. Antenna array is the mechanism which provides precise control of the spatial current distribution. Using the spatial configuration of basic antenna elements one can realize arbitrarily complex radiation patterns with the help of arrays. Also while manipulating the radiation pattern, the terminal characteristics of the basic antenna elements alter only marginally. The input impedance of the antenna, therefore, gets decoupled from the radiation pattern. One can then choose a basic radiating element which gives proper impedance characteristics without worrying about its radiation pattern. The radiation pattern can be controlled independently by adjusting the array parameters. The arrays, therefore, provide tremendous flexibility in shaping the radiation characteristics of a radiating system.

An antenna array is a spatial distribution of basic antenna elements excited with voltages or currents having definite amplitude and phase relationship. For an array, in principle, it is neither necessary to have identical antenna elements nor their uniform distribution. However, in practice, not having identical antenna elements, does not provide any advantage. On the contrary, identical elements guarantee same impedance characteristics for all antennas and also same radiation characteristics like polarization, individual radiation patterns, etc. Hence, the antenna arrays generally consist of identical antenna elements.

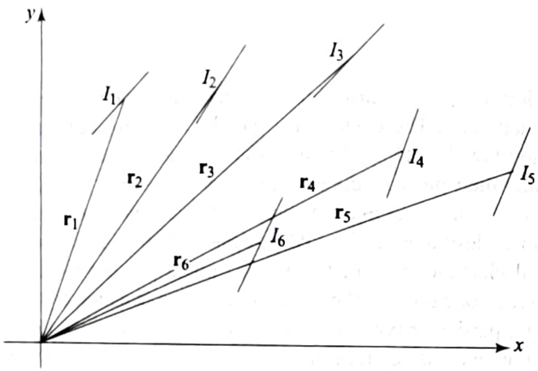


Fig. 9.1 *General array configuration.*

A general array of dipoles is shown in Fig. 9.1. The terminal currents for the elements are $I_1, I_2, I_3, \ldots$ etc. and their special location are $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots$ etc. The terminal currents $I_1, I_2, \ldots$ are in general complex, i.e. different antennas have different current amplitudes and different relative phases. Since, the radiation field is directly proportional to the antenna current, the fields due to different elements also have relative amplitudes and phases same as that of the antenna currents. The total field at any point in space can then be obtained by superposition of the fields of the individual antennas. It should be noted that while applying superposition we assume that the characteristics of the individual antennas do not get affected by the presence of other antenna elements. This assumption is valid only if the spacing between the elements is not very small. For small spacing between the elements, the phenomenon called 'mutual coupling' affects the behavior of the neighbouring antennas. The effect of mutual coupling is manageably small if the element spacing is larger than about $\frac{\lambda}{2}$. The radiation pattern of an array can be controlled by controlling either one or some or all of the following parameters.

1. Spatial configuration of antenna elements
2. Relative amplitudes of the antenna currents
3. Relative phases of the antenna currents
4. Radiation pattern of the basic antenna element
5. Location of antenna elements.

Now, for a given basic antenna element and given spatial configuration, for an N-element antenna array, we have $(N - 1)$ relative amplitudes, $(N - 1)$ relative phases, and $(N - 1)$ relative positions of the antenna elements. That is to say that we have $3(N - 1)$ control parameters and hence we can put $3(N - 1)$ constraints on the radiation pattern. The constraints could be like, the direction of maximum radiation, directions of nulls, directivity, maximum side lobe level etc. It is then clear that by increasing the number of elements, N, in an array, we can obtain better realization of the desired radiation pattern.

In the following sections, we investigate the effect of each of the above parameters on the radiation pattern. First, we investigate the simplest form of an array, the two element array, and subsequently we investigate the linear arrays. Other special array configurations like planar or volume arrays are beyond the scope of this book. However, once the principle of arrays is understood for the linear arrays, it can be extended to the other configuration in a rather straight-forward manner.

9.1 TWO ELEMENT ARRAY

Let us consider a two element array of isotropic antennas. Although an isotropic antenna does not exist in practice, its use helps in understanding the principle of arrays. Let us assume that the two antennas are separated by a distance 'd' and are excited with currents $I_1\angle\delta_1$ and $I_2\angle\delta_2$. I_1, I_2 and δ_1, δ_2 are the amplitudes and phases of the currents in the two antennas respectively (see Fig. 9.2).

The line joining the two elements is called the axis of the array. The line AB, therefore, defines the axis of the array. All the angles ϕ are measured from the axis of the array. Note that the measurement of the angle ϕ from the axis of the array is arbitrary, and it has nothing to do with the standard spherical coordinate system (r, θ, ϕ).

It is obvious from Fig. 9.2 that the array has a revolutional symmetry around the axis AB. That is to say that the radiation pattern is identical in all planes passing through the axis AB. Without losing generality, we can first obtain the radiation pattern in the plane of the paper and the total radiation pattern can be obtained by revolving the planar radiation pattern about the array axis AB. The range of ϕ is therefore from 0 to π only.

As discussed earlier, the radiation pattern is a plot of magnitude of the electric field measured at a far away distance from the antenna as a function of direction. Let us find the total electric field at a far away point P. The distance r of the

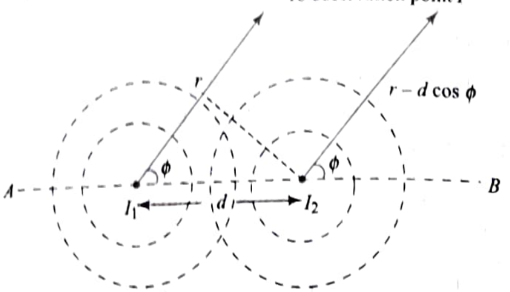


Fig. 9.2 *A two element array.*

point P from the array is much larger compared to the size of the array. The field due to the radiating element is proportional to the current. Hence, the field due to radiator 1 is

$$E_1 = \frac{K I_1 e^{j\delta_1}}{r} e^{-j\beta r} \tag{9.1}$$

where K is a constant which depends upon the medium parameters, etc. Since point P is very far away from the array, the distance of point P from radiator 2 is $(r - d\cos\phi)$, and consequently the field due to radiator 2 is

$$E_2 = \frac{K I_2 e^{j\delta_2}}{(r - d\cos\phi)} e^{-j\beta(r - d\cos\phi)} \tag{9.2}$$

Since, $d << r$, the amplitude of E_2 can be approximated by $\frac{K I_2}{r}$, i.e. $d\cos\phi$ can be neglected compared to r in the amplitude term. However, same approximation cannot be applied in the phase term. This is due to the fact that $\beta d\cos\phi = 2\pi\frac{d}{\lambda}\cos\phi$. The phase is significant if d is comparable to λ. If $d << \lambda$ then the phase is negligible. In an array, however, d is comparable to λ, hence, $d\cos\phi$ cannot be neglected in the phase term. This approximation is invariably done in all antenna and array problems. The electric field due to radiator 2 can then be written as

$$E_2 = \frac{K I_2 e^{j\delta_2}}{r} e^{-j\beta r} e^{j\beta d\cos\phi} \tag{9.3}$$

Since, the two radiating elements are identical, the direction of the two electric fields at point P is the same. The total electric field at point P therefore is

$$\begin{aligned} E &= E_1 + E_2 \\ &= \frac{K I_1 e^{j\delta_1}}{r} e^{-j\beta r} + \frac{K I_2 e^{j\delta_2}}{r} e^{-j\beta r} e^{j\beta d\cos\phi} \\ &= \frac{K e^{-j\beta r}}{r} \{ I_1 e^{j\delta_1} + I_2 e^{j\delta_2} e^{j\beta d\cos\phi} \} \end{aligned} \tag{9.4}$$

For a given distance r,

$$\frac{K e^{-j\beta r}}{r} = \text{constant } K_0 \ (say) \tag{9.5}$$

and we have the total electric field

$$E = K_0 \left\{ I_1 e^{j\delta_1} + I_2 e^{j\delta_2} e^{j\left(\frac{2\pi d}{\lambda}\right)\cos\phi} \right\} \tag{9.6}$$

$$\boxed{E = K_0 I_1 e^{j\delta_1} \left\{ 1 + \frac{I_2}{I_1} e^{j(\delta_2 - \delta_1)} e^{j2\pi\left(\frac{d}{\lambda}\right)\cos\phi} \right\}} \tag{9.7}$$

A plot of $|E|$ as a function ϕ gives the radiation pattern of the array in the plane containing the axis of the array AB. Now, since the radiation pattern is relative variation of $|E|$ as a function of ϕ, the term outside the bracket in Eqn (9.7) does not alter the radiation pattern (it only scales the radiation pattern). It is then clear that, the radiation pattern can be controlled by controlling any or all of the following parameters,

(i) d/λ (spacing between elements with respect to the wavelength),
(ii) $(\delta_2 - \delta_1)$ (phase difference between two elements),
(iii) I_2/I_1 (ratio of the amplitudes of the currents).

Let us investigate the effect of each of the parameters on the radiation pattern.

9.1.1 Effect of Phase Difference on Radiation Pattern

Let us assume that $I_1 = I_2 = I$ giving $\frac{I_2}{I_1} = 1$. The radiation pattern from Eqn (9.7) can be written as

$$E = K_0 I e^{j\delta_1} \left\{ 1 + e^{j(\delta_2 - \delta_2)} e^{j\frac{2\pi d}{\lambda}\cos\phi} \right\} \tag{9.8}$$

Since, the radiation pattern is a plot of $|E|$ vs angle ϕ, we have,

$$|E| = |K_0 I| \left| \left\{ 1 + e^{j(\delta_2 - \delta_1)} e^{j\frac{2\pi d}{\lambda}\cos\phi} \right\} \right| \tag{9.9}$$

Taking element 1 as the reference element (i.e. $\delta_1 \equiv 0$), and phase difference $(\delta_2 - \delta_1) = \delta$, we get

$$|E| = |K_0 I| \left| \left\{ 1 + e^{j(\delta + \frac{2\pi d}{\lambda}\cos\phi)} \right\} \right| \tag{9.10}$$

Using Euler's identity, $\frac{e^{jx} + e^{-jx}}{2} = \cos x$, Eqn (9.10) can be written as

$$|E| = 2|K_0 I| \left| \cos\left(\frac{1}{2}\{\delta + \frac{2\pi}{\lambda} d\cos\phi\} \right) \right| \tag{9.11}$$

From Eqn (9.11), we note that, $|E|$ becomes maximum when

$$\frac{1}{2}\left(\delta + \frac{2\pi}{\lambda} d\cos\phi \right) = m\pi \qquad m = 0, 1, 2, \cdots \tag{9.12}$$

Taking the principle value of m, we get for the direction of maximum radiation ϕ_{max}

$$\delta + \frac{2\pi}{\lambda} d \cos\phi_{max} = 0 \tag{9.13}$$

$$\Rightarrow \qquad \phi_{max} = \cos^{-1}\left(\frac{-\delta\lambda}{2\pi d}\right) \tag{9.14}$$

For a given wavelength of operation and given spacing between the elements, the direction of maximum radiation varies with δ. In other words, by varying the phase difference, the direction of maximum radiation can be varied. For $\delta = 0$, i.e. if the two elements are excited with equal phases, we get $\phi_{max} = \cos^{-1}(0) = \frac{\pi}{2}$, whereas for $\delta = -\beta d = -\frac{2\pi d}{\lambda}$, we get $\phi_{max} = \cos^{-1}(1) = 0$

From Eqn (9.14) we can see that as δ varies from $-\beta d$ to $+\beta d$, the direction of maximum radiation varies from $\phi_{max} = 0$ to $\phi_{max} = 180°$. The inter-element phase difference therefore is one of the crucial parameters as it primarily decides the direction of maximum radiation. Figure 9.3 shows the radiation patterns of a two element array for various values of δ.

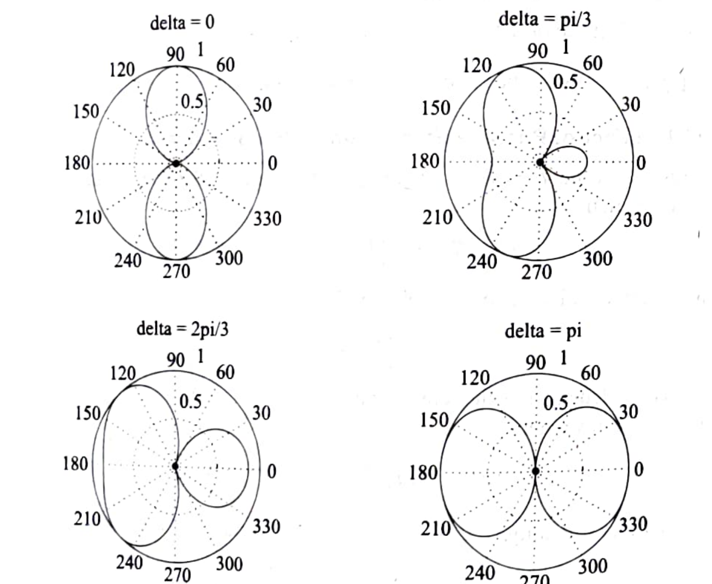


Fig. 9.3 *Radiation pattern of a two element array for various values of δ. The spacing between the elements is $\lambda/2$.*

9.1.2 Effect of Inter-element Spacing

Let us assume in this case $I_1 = I_2 = I$ and $\delta_1 = \delta_2$ (i.e. $\delta = 0$). The radiation pattern can be written from Eqn (9.11) as

$$|E| = 2|K_0 I| \left|\cos\left(\frac{\pi d}{\lambda}\cos\phi\right)\right| \tag{9.15}$$

First thing we note from Eqn (9.15) is that we always have quantity $(\frac{d}{\lambda})$ in the expression of the array pattern meaning there is no significance of the absolute distance d. It is the ratio $\frac{d}{\lambda}$ (the distance normalized with respect to the wavelength) which matters for the radiation pattern.

The maximum radiation occurs when

$$\pi\frac{d}{\lambda}\cos\phi_{max} = m\pi, \qquad m = 0, \text{ integer} \tag{9.16}$$

$$\Rightarrow \qquad \phi_{max} = \cos^{-1}\frac{m\lambda}{d} \tag{9.17}$$

All those values of ϕ_{max} are physically realizable for which $m\frac{\lambda}{d} \le 1$. That is, for all values of $m \le \frac{d}{\lambda}$, the radiation is maximum. The number of directions N in which the radiation is maximum is

$$N = \text{Maximum allowed value of } m = floor\left(\frac{d}{\lambda}\right) + 1 \tag{9.18}$$

where $floor(x)$ represents highest integer smaller than x.

From Eqn (9.18) it is clear that as $\frac{d}{\lambda}$ increases, number of directions of maximum radiation also increases.

Now the cosine function in Eqn (9.15) goes to zero for angles ϕ_{null} given by

$$\frac{\pi d}{\lambda}\cos\phi_{null} = \left(m + \frac{1}{2}\right)\pi, \qquad m = 0, \text{ integer} \tag{9.19}$$

Hence, as the ratio $\frac{d}{\lambda}$ increases, maximum allowed value of m increases and consequently number of directions in which the radiation is zero also increases. The directions in which the radiation becomes zero are called the 'nulls' of the radiation pattern. It is clear from Eqns (9.16) and (9.19) that the directions of nulls and the directions of maximum radiation appear alternately as shown in Fig. 9.4.

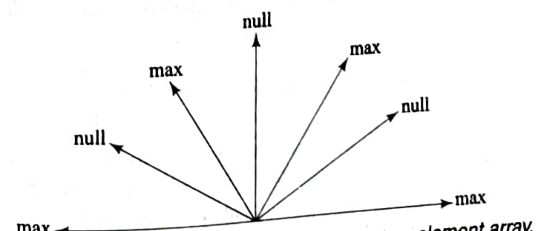


Fig. 9.4 *Maxima and nulls in a radiation pattern of a two-element array.*

The effect of inter-element spacing, $\frac{d}{\lambda}$, is therefore to split the radiation pattern in angular zones. Larger the spacing, more is the number of zones.

From Eqn (9.18) we can also note that to keep single direction for maximum radiation, N should be equal to 1 i.e. $\frac{d}{\lambda}$ should be less than 1. We can then conclude that for $\frac{d}{\lambda} < 1$, there is only one direction in which the radiation is maximum. Figure 9.5 shows radiation patterns of a two element array for various values of inter element spacings.

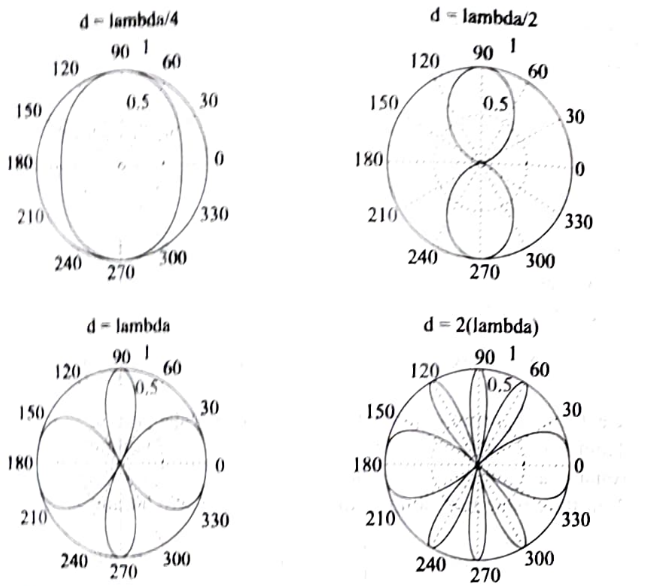


Fig. 9.5 *Radiation pattern of a two-element array for various values of the inter-element spacing, d.*

9.1.3 Effect of Amplitude Ratio

Let us take a general case and investigate the effect of $\frac{I_2}{I_1}$ on the radiation pattern. The radiation pattern is given by Eqn (9.7) as

$$|E| = |K_0 I_1| \left|\left\{1 + \frac{I_2}{I_1} e^{j(\delta + \frac{2\pi}{\lambda} d \cos\phi)}\right\}\right| \qquad (9.20)$$

It is clear that the ratio $R = \frac{I_2}{I_1}$ can vary from 0 to ∞. For $R = 0$, we have $I_2 = 0$ meaning no excitation to the second element, and similarly $R = \infty$ meaning no excitation to the first element. In either of the cases, there is no array, and the radiation pattern is that of the individual elements (isotropic in this case). The maximum array effect is therefore realized when $I_1 = I_2$, i.e. equal

excitation to both the array elements. Generally, the array elements are excited with equal magnitudes but with different phases. This however reduces the degrees of freedom in the design of arrays. Indeed, there are special applications where the desired radiation pattern cannot be realized without varying the amplitude of the current across the array elements. Figure 9.6 shows radiation pattern of a two element array for different current ratios.

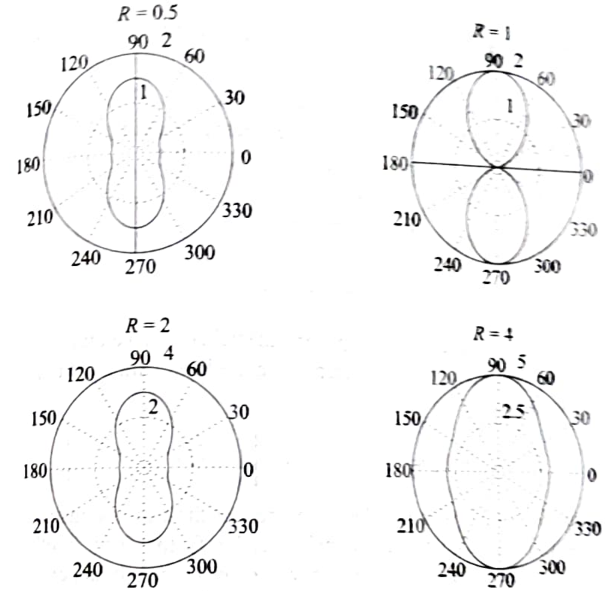


Fig. 9.6 *Radiation pattern of a two element array for various values of the current ratio, R.*

From the above discussion we find that the three parameters of the array namely δ, $\frac{d}{\lambda}$ and $\frac{I_1}{I_2}$ have distinct effects on the radiation pattern. By controlling all three parameters one can realize a desired radiation pattern of any higher complexity.

In the above discussion, we assumed the basic radiating element to be isotropic. In practice, however, we can never find an isotropic radiator. It may be recalled that even the smallest possible antenna, the Hertz dipole, is not isotropic as it has no radiation along its axis and maximum radiation perpendicular to it. One would then wonder, how the array analysis would proceed if the array elements are not isotropic. Obviously, if the elements have non-identical patterns, there is no choice but to compute the radiation pattern in each plane passing through the array axis. The three dimensional radiation pattern would be collection of all planar patterns. If, the elements are identical and are placed with identical orientations, the array analysis simplifies considerably.

Let us consider a two element array of non-isotropic but identical antennas as shown in Fig. 9.7. Let the radiation pattern for the individual element be given as

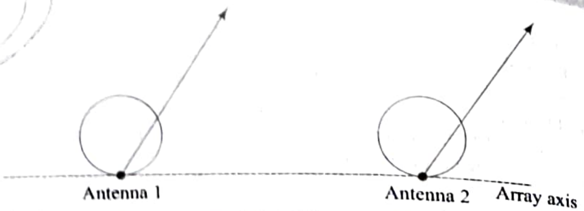


Fig. 9.7 *Array of antennas with directional primary radiation pattern.*

$f(\phi)$ in the plane of the paper. If both elements have identical orientations and are excited with currents, $I_1\angle 0$ and $I_2\angle\delta$, the electric fields at a far-away point P due to two sources can be written as (Eqns (9.1) and (9.2))

$$E_1 = \frac{K I_1 e^{-j\beta r}}{r} f(\phi) \tag{9.21}$$

$$E_2 = \frac{K I_2 e^{j\delta} e^{-j\beta r}}{r} f(\phi) \tag{9.22}$$

Note that due to individual radiation pattern, the field of each element is weighted by $f(\phi)$. The total field at point P then is (on the lines similar to that followed for isotropic elements)

$$E = E_1 + E_2 = \frac{K e^{-j\beta r} I_1}{r} f(\phi) \left\{ 1 + \frac{I_2}{I_1} e^{j\delta} e^{j\beta d \cos\phi} \right\} \tag{9.23}$$

Comparing Eqn (9.23) with Eqn (9.7) we find that the expressions are identical except that in Eqn (9.23) we have a multiplication factor $f(\phi)$. In other words, the radiation pattern of a non-isotropic array is equal to the product of the element pattern (also called the primary pattern) and the radiation pattern of the array of isotropic elements. Since the radiation pattern of an array of isotropic elements is purely defined by the array parameters, it is generally called the 'array factor' (AF). We then obtain

Radiation pattern of an array of non-isotropic elements = Element radiation pattern × array factor	(9.24)

The analysis of an array of non-isotropic but identical elements is therefore quite straight-forward. First, replace the actual elements by point (isotropic) sources and compute the radiation pattern of the array, the array factor. Finally, multiply the array factor by the radiation pattern of the element.

Before we proceed further, it should be mentioned that since the radiation pattern is always normalized with respect to the maximum field amplitude, the constant K_0, etc. do not have any importance in calculation of the array patterns.

EXAMPLE 9.1 A two element array of isotropic antennas is to be used to get maximum radiation and a null along the array axis in the opposite directions. Find the array parameters.

Solution:

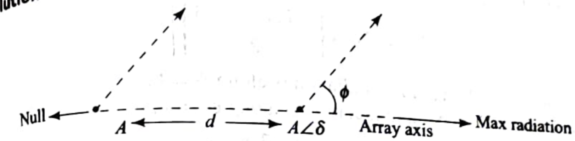


Fig. 9.8 *Two element array for Example 9.1.*

Since we need a null in the radiation pattern, the amplitudes of two antenna currents must be equal.

Let the currents in the two elements be A and $Ae^{j\delta}$ respectively, and let the element spacing be d. The radiation pattern then is

$$|E| = |K\{A + Ae^{j\delta}.e^{j\beta d\cos\phi}\}|$$

$$= \left|2AK\cos\left(\frac{\beta d\cos\phi + \delta}{2}\right)\right|$$

Let the maximum radiation be along $\phi = 0$. Then the null is along $\phi = \pi$.

$\Rightarrow$ At $\phi = 0$, we should have $\dfrac{\beta d\cos\phi + \delta}{2} = 0$

$\Rightarrow$ $\beta d + \delta = 0$ ... (1)

At $\phi = \pi$, we should have $\dfrac{\beta d\cos\phi + \delta}{2} = \pm\dfrac{\pi}{2}$

$\Rightarrow$ $-\beta d + \delta = \pm\pi$ ... (2)

Solving the Eqns (1) and (2), we get

$$\delta = -\pi/2$$

and $\beta d = +\pi/2 \Rightarrow \dfrac{2\pi}{\lambda} d = \pm\pi/2$

$\Rightarrow$ $d = \lambda/4$

The radiation pattern is

$$|E| = \left|2AK\cos\left(\frac{\pi\cos\phi - \pi}{4}\right)\right|$$

EXAMPLE 9.2 A two element array of isotropic radiators is mounted in horizontal plane. The direction of the maximum radiation is to be scanned from horizon to horizon by varying the phase of the antenna elements. If the maximum phase variation possible with the phase shifter is $\pm\pi/2$, find the spacing between the elements.

Solution:

Assume that the elements have equal amplitudes. The radiation pattern is

$$|E| = \left|2\cos\left\{\frac{\beta d\cos\phi + \delta}{2}\right\}\right|$$

Now for $\delta = -\pi/2$, maximum radiation is along $\phi = 0$,
and for $\delta = +\pi/2$, maximum radiation is along $\phi = \pi$,

$$\Rightarrow \quad \beta d + \left(-\frac{\pi}{2}\right) = 0$$

$$\text{and} \quad -\beta d + \left(\frac{\pi}{2}\right) = 0$$

Both are in fact the same equations, i.e. if $\delta = -\pi/2$ scans the maximum to $\phi = 0$, $\delta = \pi/2$ will automatically scan the maximum to $\phi = -\pi$. There is no independent control of scanning angles of the two maxima. We then get

$$d = \frac{\pi}{2\beta} = \frac{\pi}{2 \times 2\pi/\lambda} = \lambda/4$$

Therefore for a two element array with inter-element spacing of $\lambda/4$, the direction of maximum radiation can be scanned from $\phi = 0$ to $\phi = \pi$ by varying δ from $-\pi/2$ to $\pi/2$.

EXAMPLE 9.3 A Hertz dipole is vertically mounted at a height of λ above an ideal conducting ground. Find the directions of maxima and the nulls in the vertical plane.

Solution:

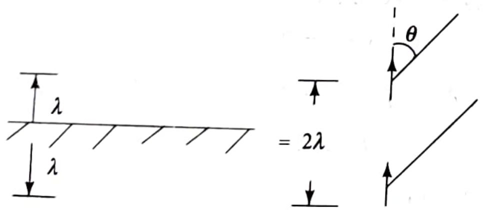


Fig. 9.9 *Hertz dipole located above the ground plane.*

Since, the ground is ideal, we can replace the ground by an image of the Hertz dipole. The configuration is equivalent to a two element array with element spacing $d = 2\lambda$, phase difference between elements, $\delta = 0$, and equal amplitudes.

The array factor is

$$AF = \cos\left\{\frac{\frac{2\pi}{\lambda}(2\lambda)\cos\theta}{2}\right\} = \cos(2\pi\cos\theta)$$

where θ is the angle measured from the array axis.
Now, the element pattern for the Hertz dipole is $\sin\theta$
The total radiation pattern is

$$|E| = |\sin\theta\cos(2\pi\cos\theta)|$$

In this problem θ varies from 0 to $\pi/2$ only.
For Nulls either $\sin\theta = 0 \quad \Rightarrow \quad \theta = 0$

$$\text{or} \quad \cos(2\pi\cos\theta) = 0 \quad \Rightarrow \quad 2\pi\cos\theta = \pm(2m+1)\frac{\pi}{2}$$

$$\Rightarrow \quad \cos\theta = \pm(2m+1)/4$$

The nulls are: $\theta_{null} = 0, \cos^{-1}(1/4) = 75.52°, \cos^{-1}(3/4) = 41.4°$
The direction of the maximum radiation is $\theta_{max} = \pi/2$.

EXAMPLE 9.4 Two elements of an array have current amplitudes 1.0 and 0.5, and the spacing between the elements is $\lambda/2$. Find the directions of maximum and minimum fields when the array is excited with (i) zero phase difference (ii) $\pi/2$ phase difference.

Solution:

The radiation pattern is

(i) For $\delta = 0$

$$|E| = \left|K\left(1 + 0.5e^{j\left(\frac{2\pi}{\lambda}\frac{\lambda}{2}\cos\phi + \delta\right)}\right)\right|$$

$$|E| = \left|K\left(1 + 0.5e^{j\pi\cos\phi}\right)\right|$$

The field becomes maximum when the two terms add in phase giving

$$e^{j\pi\cos\phi_{max}} = +1$$

$$\Rightarrow \quad \pi\cos\phi_{max} = 0 \quad \Rightarrow \quad \phi_{max} = \pi/2$$

For minimum field, the two terms should substract each other making

$$e^{j\pi\cos\phi_{min}} = -1 \quad \Rightarrow \quad \pi\cos\phi_{min} = \pm\pi,$$

$$\Rightarrow \quad \phi_{min} = 0, \pi$$

(ii) For $\delta = \pi/2$, we have

$$|E| = |K(1 + 0.5e^{j(\pi\cos\phi + \pi/2)})|$$

For maximum field,

$$\pi\cos\phi_{max} + \pi/2 = \pm 2m\pi \qquad m = 0, 1, 2, \ldots$$

$$\Rightarrow \quad \cos\phi_{max} = \pm 2m - 1/2$$

$$\Rightarrow \quad \phi_{max} = \cos^{-1}(-1/2) = 120^\circ$$

Note: Only those values of m are chosen for which $\cos\phi_{max} \leq 1$, and ϕ lies between 0 and 180°.

For minimum field,

$$\pi\cos\phi_{min} + \pi/2 = \pm(2m+1)\pi$$

$$\Rightarrow \quad \cos\phi_{min} = \pm(2m+1) - 1/2$$

$$\Rightarrow \quad \phi_{min} = \cos^{-1}(1/2) = 60^\circ$$

9.2 UNIFORM LINEAR ARRAY

We have seen in the earlier sections as to, how the radiation pattern of a two element array changes with spacing between the elements, the ratio of the amplitudes of the currents of the elements, and the phase difference between the two elements. Since, we get multiple directions for maximum radiation for large spacing between elements, generally, the spacing is kept between $\frac{\lambda}{2}$ and λ. That is, the spacing cannot be varied arbitrarily over a wide range. We then have only two parameters, the amplitude ratio and the phase difference between the elements at our disposal, for controlling the radiation pattern. Obviously, with these two parameters we can obtain only a limited control over the radiation pattern. To realize complex radiation patterns we need more controlling parameters in an array. We can increase the controlling parameters by increasing the number of elements in an array.

Let us consider an array of N elements. For an N-element array we can control $(N-1)$ current amplitude ratios, $(N-1)$ phase differences, and $(N-1)$ spacings between the elements. This array, naturally provides maximum control over the radiation pattern. In many applications, however, we do not need such complex radiation patterns which would need so many control parameters. Besides, more the control parameters, higher the complexity of the electronic circuitry in the array feeding network. In large number of applications, therefore, the spacing between any two adjacent elements (called 'inter-element spacing') is same, i.e. the elements are equispaced. Also, these elements are excited with equal currents making the current ratio between any two elements unity. We also have same phase difference between any two adjacent elements. The array then has uniform spacing, uniform amplitude and uniform phase difference. This array is called the 'Uniform Array'. If the elements are arranged along a straight line, the array is called a 'Uniform Linear Array'.

Uniform linear arrays are quite popular in practice as they have only three parameters to control namely, number of elements, inter-element spacing and

inter-element phase difference. In the following sections we develop the theory of the uniform linear array. Here also, we assume that the basic array elements are isotropic in nature.

Let there be N isotropic equi-spaced elements along a straight line as shown in Fig. 9.10. Let the inter element spacing be denoted by 'd', and the inter-element phase difference be denoted by 'δ'. δ is also called the progressive phase shift progressive phase shift of the array. Since, the current is uniform along the array, without losing generality, let us assume all currents to be of unity amplitude. Let us also assume that the left most current element is the reference element, i.e. all the phases are measured from that element. Then by assumption, the first element has zero phase. Since, the phase difference between any two adjacent elements is δ, the phases of 2, 3, 4,..., N elements would be $\delta, 2\delta, 3\delta, \ldots, (N-1)\delta$, respectively.

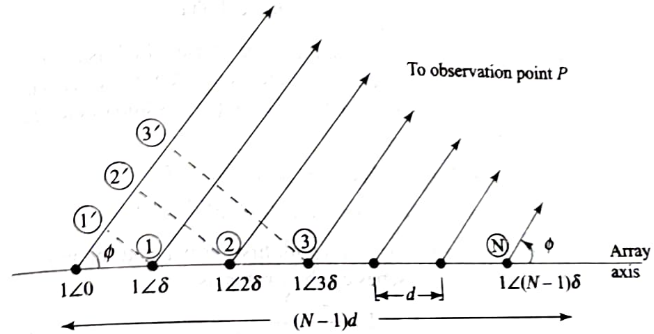


Fig. 9.10 *A N-element uniform array.*

Let us now consider an observation point P far away from the array in the direction making an angle ϕ with respect to the axis of the array. Since, the point P is at a very large distance, the radiation paths from the array elements to the observation points are almost parallel. The radiation from element 2 travels a shorter distance compared to element 1 by an amount $1'-2' = d\cos\phi$, where $2-2'$ is the perpendicular from 2 on the line joining 1 and the point of observation (P). If the radiation has a phase constant $\beta =$(phase change per unit length $= \frac{2\pi}{\lambda}$), the distance $1'-2'$ would correspond to a phase difference of $\beta d\cos\phi$. As the radiation from element 2 travels a shorter distance, compared to that from 1, the field at point P due to element 2 would lead by a phase, $\beta d\cos\phi$ with respect to the field due to element 1. Both fields, however, would have equal amplitudes, say, E_0, as the two elements have equal currents. Since, the current of element 2 itself leads by a phase δ with respect to the current of element 1, and since, the radiation field is proportional to the current of the element, the field at point P due to element 2, leads by a total phase of $\beta d\cos\phi + \delta$. Denoting the fields at point P due to elements 1, 2,... ,N by $E_1, E_2, \ldots, E_N$ respectively, we have

$$E_1 = E_0\angle 0 \tag{9.25}$$

$$E_2 = E_0\angle(\beta d\cos\phi + \delta) \tag{9.26}$$

and so on.

From Fig. 9.10, we can observe that the field due to some k^{th} element would be

$$E_k = E_0\angle\{(k-1)(\beta d\cos\phi + \delta)\}, \qquad k = 1, 2, \cdots, N \tag{9.27}$$

For the sake of brevity, let us denote the total phase difference between the fields due to any two adjacent elements by ψ, giving

$$\psi = \underbrace{\beta d\cos\phi}_{\text{Space phase}} + \underbrace{\delta}_{\text{Electrical phase}} \tag{9.28}$$

The total phase ψ consists of two parts, the space phase which is a function of direction ϕ, and the electrical phase which is independent of the direction ϕ.

The total electric field at point P is obtained by superposition of the fields due to the individual elements as

$$E = E_0 e^{j0} + E_0 e^{j\psi} + E_o e^{j2\psi} + \ldots\ldots\ldots\ldots + E_0 e^{j(N-1)\psi} \tag{9.29}$$

$$= E_0\left\{1 + e^{j\psi} + e^{j2\psi} + \ldots\ldots\ldots\ldots\ldots + e^{j(N-1)\psi}\right\} \tag{9.30}$$

The series inside { } is a geometric series with first term 1 and the progression ratio $e^{j\psi}$. The summation of the series can be written as

$$E = E_0\frac{1 - e^{jN\psi}}{1 - e^{j\psi}} \tag{9.31}$$

[Note: The summation of a geometric series $a + ar + ar^2 + \ldots\ldots\ldots\ldots + ar^{N-1} = \frac{a(1-r^N)}{(1-r)}$].

Equation (9.31) can be rewritten as

$$E = E_0\frac{e^{\frac{N\psi}{2}}}{e^{\frac{j\psi}{2}}}\left(\frac{e^{\frac{-jN\psi}{2}} - e^{\frac{jN\psi}{2}}}{e^{\frac{-j\psi}{2}} - e^{\frac{j\psi}{2}}}\right) \tag{9.32}$$

$$= E_0 e^{\frac{j(N-1)\psi}{2}}\left(\frac{\sin(\frac{N\psi}{2})}{\sin(\frac{\psi}{2})}\right) \tag{9.33}$$

The magnitude of the electric field (the radiation pattern) is

$$|E| = |E_0|\left|\frac{\sin(\frac{N\psi}{2})}{\sin(\frac{\psi}{2})}\right| \tag{9.34}$$

For $\psi = 0$, the expressions in Eqns (9.31) and (9.34) are to be evaluated in the limit $\psi \to 0$. This is rather simple, as we can obtain $|E|$ for $\psi = 0$ by substituting $\psi = 0$ in Eqn (9.30). Obviously, $\psi = 0$ gives the maximum field equal to NE_0.

Normalizing the electric field $|E|$ with respect to the maximum field NE_0, we get the normalized radiation pattern of the array (the Array Factor (AF)) as

$$\text{AF} \equiv \frac{E}{NE_0} = \frac{1}{N}\frac{\sin(\frac{N\psi}{2})}{\sin(\frac{\psi}{2})} \tag{9.35}$$

Equation (9.35) gives the general expression for the array factor of a linear uniform array.

Let us now investigate the characteristics of the array pattern for various array parameters. From Eqn (9.35) we can study the following characteristics.

9.2.1 Direction of Maximum Radiation

As seen above, the maximum radiation corresponds to $\psi = 0$. If the direction of maximum radiation is denoted by ϕ_{max} we have

$$\psi = \beta d\cos\phi_{max} + \delta = 0 \tag{9.36}$$

$$\Rightarrow \quad \cos\phi_{max} = -\frac{\delta}{\beta d} \tag{9.37}$$

$$\Rightarrow \quad \phi_{max} = \cos^{-1}\left(-\frac{\delta}{\beta d}\right) = \cos^{-1}\left(\frac{\delta\lambda}{2\pi d}\right) \tag{9.38}$$

Equation (9.38) is identical to Eqn (9.14). The direction of maximum radiation is independent of the number of elements in the uniform array. Therefore, as discussed in the case of the two element array, the direction of maximum radiation can be varied from $\phi = 0$ to $\phi = \pi$ by varying δ from $-\beta d$ to $+\beta d$.

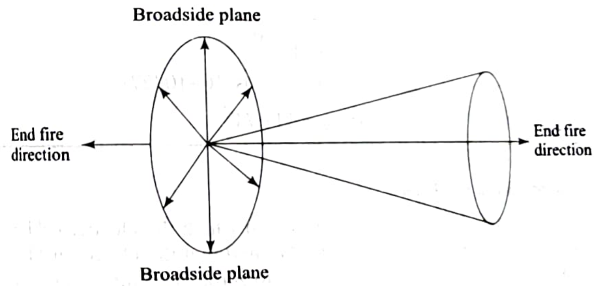


Fig. 9.11 *Broadside and end-fire directions of an antenna array.*

If the maximum radiation appears along the axis of the array (i.e. $\phi = 0$ or $\phi = \pi$) the array is called the 'End-fire Array', and if the maximum radiation appears along the direction perpendicular to the axis of the array (i.e. $\phi = \frac{\pi}{2}$), the array is called the 'Broad-side Array'. The two directions are shown in Fig. 9.11. Visualizing the directions in three dimensions, we would note that there are only two end fire directions, but infinite broadside directions. In fact, a plane which is perpendicular to the array axis is the broad-side plane and any direction in that plane is a broadside direction. If the array is neither broad-side nor end-fire, the maximum radiation goes along the surface of a cone as shown in Fig. 9.11.

Substituting for δ from Eqn (9.36) into Eqn (9.28) we get

$$\psi = \beta d \cos\phi - \beta d \cos\phi_{max} = \beta d(\cos\phi - \cos\phi_{max}) \tag{9.39}$$

EXAMPLE 9.5 A 5 element uniform array has inter element spacing 0.3λ and progressive phase shift of 40°. Find the radiation pattern and direction of maximum radiation for the array.

Solution:

For the array.

$$\psi = \frac{2\pi}{\lambda}(0.3\lambda)\cos\phi + 40^\circ$$

$$= 0.6\pi \cos\phi + \frac{40 \times \pi}{180} \text{ rad.}$$

Radiation pattern is

$$|E| = \left|\frac{\sin(N\psi/2)}{N\sin(\psi/2)}\right| = \left|\frac{\sin(5\psi/2)}{5\sin(\psi/2)}\right|$$

$$= \left|\frac{\sin(1.5\pi\cos\phi + 5\pi/9)}{5\sin(0.3\pi\cos\phi + \pi/9)}\right|$$

For direction of maximum radiation we put $\psi = 0$ giving

$$0.6\pi\cos\phi_{max} + \frac{40\pi}{180} = 0$$

$$\Rightarrow \quad \phi_{max} = \cos^{-1}(-10/27)$$

Noting that $0 < \phi < \pi$, we get $\phi_{max} = 111.74^\circ$.

9.2.2 Directions of Nulls

The direction in which no radiation goes, i.e. in which the electric field is zero is called a 'Null' of the radiation pattern. We find that the electric field is zero whenever, the numerator of Eqn (9.35) goes to zero. The nulls of the radiation pattern, therefore, correspond to

$$\sin\left(\frac{N\psi}{2}\right) = 0 \tag{9.40}$$

$$\Rightarrow \quad \frac{N\psi}{2} = \pm m\pi, \qquad m = 1, 2, 3, \cdots \tag{9.41}$$

$$\Rightarrow \quad \psi = \pm\frac{2m\pi}{N}, \qquad m = 1, 2, 3, \cdots \tag{9.42}$$

If ϕ_{null} denotes the direction of a null, we have

$$\psi = \beta d\cos\phi_{null} + \delta = \pm\frac{2m\pi}{N} \tag{9.43}$$

Substituting for δ from Eqn (9.36) into Eqn (9.43) we get

$$\beta d\{\cos\phi_{null} - \cos\phi_{max}\} = \pm\frac{2m\pi}{N} \tag{9.44}$$

$$\Rightarrow \quad \cos\phi_{null} = \cos\phi_{max} \pm \frac{2m\pi}{\beta d N} \tag{9.45}$$

$$\cos\phi_{null} = \cos\phi_{max} \pm \frac{m\lambda}{dN} \tag{9.46}$$

It should be remembered that in Eqn (9.46), m has to take a non-zero integer value. $m = 0$ corresponds to $\psi = 0$ which represents the maximum radiation and not a null. To obtain the directions of the nulls, ϕ_{null}, one has to choose all values of m for which the right hand side of Eqn (9.46) lies within ± 1. For every array, therefore, there is a maximum permissible value of m meaning there are finite number of nulls for an array. It may be pointed out however that depending upon the value of ϕ_{max}, the maximum value of m may be different with + and − sign, in Eqn (9.46). For example, if $\cos\phi_{max} = 0.8$, $d = \lambda$ and $N = 5$, we have

$$\cos\phi_{null} = 0.8 \pm \frac{m\lambda}{\lambda 5}$$

$$= 0.8 \pm 0.2m \tag{9.47}$$

For +ve sign, the maximum value of m is 1 whereas, for −ve sign, the maximum value of m is 9. While obtaining the directions of the nulls, it is therefore important to test +ve and −ve values of m separately.

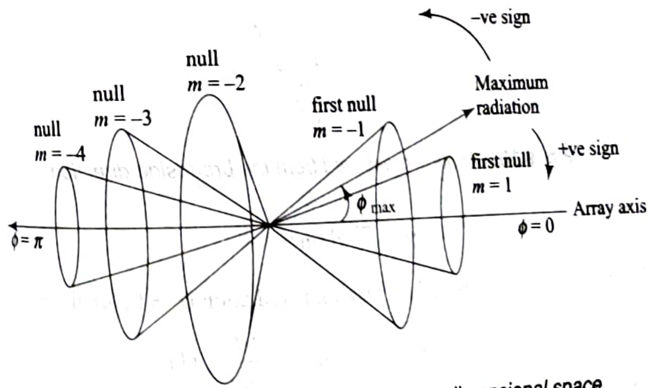


Fig. 9.12 *Nulls of a linear array in three dimensional space.*

Figure 9.12 shows a typical pattern of nulls along with the direction of maximum radiation. The asymmetry of maximum permissible value of m with

+ve or −ve sign, is clearly due to the tilt in the direction of the maximum radiation with respect to the broad-side direction. Since, the range of ϕ is from 0 to π, as ϕ_{max} tilts towards $\phi = 0$, the range of m for +ve sign, reduces, and the range of m for −ve sign, increases. The reverse occurs when the direction of maximum tilts towards $\phi = \pi$. For $\phi_{max} = \frac{\pi}{2}$, the nulls are symmetrically placed around the broad-side direction.

As mentioned earlier, the three dimensional radiation pattern is obtained by revolving the planar radiation pattern round the axis of the array. Each direction of null ϕ_{null}, therefore corresponds to the surface of a cone as shown in Fig. 9.12.

The angular zone between two nulls around the direction of maximum radiation is called the 'Main beam' of the array, and the angular width between the first nulls is called the 'Beam width between First Nulls' (BWFN). This, of course is meaningful if both the first nulls are visible, that is, they lie within $\phi = 0$ and $\phi = \pi$. If only one of the nulls lie in this range and the other does not exist, we can not define BWFN for the array.

EXAMPLE 9.6 A 4-element uniform array has maximum radiation at an angle of 30° from the broadside direction. If the array has BWFN of 90°, find the progressive phase shift and inter element spacing of the array. Also find the directions of all the nulls.

Solution:

In this example, let us define the angle from the broadside direction instead of from the array axis. We then get

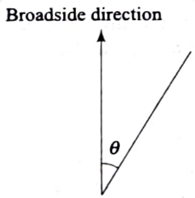


Fig. 9.13 *Angle θ measured from the broadside direction.*

$$\psi = \frac{2\pi}{\lambda} d \sin\theta + \delta$$

Now $\theta_{max} = 30° = \pi/6$, and for maximum radiation $\psi = 0$. We therefore get

$$\delta = -\frac{2\pi}{\lambda} d \sin(\pi/6) = -\frac{\pi d}{\lambda} \qquad (1)$$

The nulls of the array are given by $N\psi/2 = \pm m\pi$.

The first nulls therefore correspond to $N\psi/2 = 4\psi/2 = \pm\pi$

$\Rightarrow \quad \psi = \pm\pi/2$.

The directions of the first nulls give

$$\frac{2\pi}{\lambda} d \sin\theta_{n+} - \frac{\pi d}{\lambda} = \pi/2 \qquad (2)$$

and
$$\frac{2\pi d}{\lambda} \sin\theta_{n-} - \frac{\pi d}{\lambda} = -\pi/2 \qquad (3)$$

Adding Eqn (2) and (3) we get

$$\sin\theta_+ + \sin\theta_- = 1 \qquad (4)$$

It is given that $\quad \text{BWFN} = \theta_+ - \theta_- = \pi/2$

$\Rightarrow \quad \theta_+ = \pi/2 + \theta_-$

$\Rightarrow \quad \sin\theta_+ = \cos\theta_-$

Substituting in Eqn (4) we get

$$\cos\theta_- + \sin\theta_- = 1$$

$\Rightarrow \quad \sqrt{1 - \sin^2\theta_-} + \sin\theta_- = 1$

$\Rightarrow \quad \sin\theta_- = 0$ i.e. $\quad \theta_- = 0$

Therefore, $\quad \theta_+ = \pi/2$

The two first nulls are in the directions $\theta = 0$ and $\pi/2$.

(Note that the nulls are not symetrically placed around the direction of the maximum radiation).

Substituting $\theta_- = 0$ in (3) we get $d = \lambda/2$, and

$$\psi = \frac{2\pi}{\lambda} \cdot \frac{\lambda}{2} \cdot \sin\theta - \frac{\pi\lambda}{2\lambda} = \pi\sin\theta - \pi/2$$

For directions of other nulls, putting

$$\frac{N\psi}{2} = \frac{4\psi}{2} = 2\psi = \pm m\pi \quad \text{where } m = 2, 3$$

$\Rightarrow \quad 2(\pi\sin\theta_n - \pi/2) = \pm m\pi, \quad m = 2, 3$

$\Rightarrow \quad \sin\theta_n = \frac{1 \pm m}{2}, \quad m = 2, 3$

$$\theta_n = \sin^{-1}(-1/2) = -\pi/6, \quad \sin^{-1}(-1) = -\pi/2$$

The nulls of the array are located at

$$\theta = -\pi/2, -\pi/6, 0, \pi/2$$

9.2.3 Directions of Sidelobes

A local maximum in the radiation pattern is called the 'side lobe'. One can argue that between two adjacent nulls, there must be one local maximum of the radiation pattern and hence there exists a sidelobe between every adjacent pair of nulls.

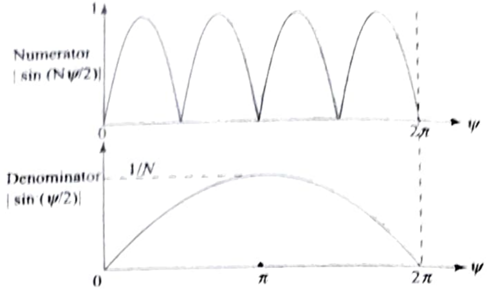


Fig. 9.14 *Plot of numerator and denominator of the array factor as a function of ψ for a uniform array.*

Figure 9.14 shows plots of the numerator and the denominator of Eqn (9.35) as a function of ψ. For large N, the numerator is a rapidly varying function of ψ compared to the denominator. We can then say that approximately whenever the numerator is maximum the radiation pattern has a local maximum that is, it has a side lobe. Approximately, then the direction of a sidelobe lies midway between two adjacent nulls. For the numerator to be maximum, we have

$$\frac{N\psi}{2} = \pm(m + \frac{1}{2})\pi \qquad (9.48)$$

$$\Rightarrow \quad \psi = \beta d\,\{\cos\phi_{SL} - \cos\phi_{max}\} = \pm\left(m + \frac{1}{2}\right)\pi, \qquad m = 1, 2, \ldots \qquad (9.49)$$

where ϕ_{SL} is the direction of a side lobe. Rearranging Eqn (9.49) and substituting $\beta = \frac{2\pi}{\lambda}$, we get

$$\cos\phi_{SL} = \cos\phi_{max} \pm \left(m + \frac{1}{2}\right)\frac{\lambda}{2d} \qquad (9.50)$$

The first sidelobes on either side of the main beam correspond to $m = 1$.

The amplitude of a side lobe can be obtained by substituting the value of ψ from Eqn (9.48) into Eqn (9.35). The amplitude of the m^{th} side lobe is

$$\frac{1}{N}\left|\frac{1}{\sin\left(\pm(m + \frac{1}{2})\frac{\pi}{N}\right)}\right| \qquad (9.51)$$

(Note: For a sidelobe the numerator $|\sin((m + \frac{1}{2})\pi)|$ is unity.) For large N, Eqn (9.51) can be approximated (through $\sin x \approx x$ for $x << 1$) to give amplitude of a side lobe as

$$\approx \frac{1}{N}\frac{1}{(m + \frac{1}{2})\frac{\pi}{N}} = \frac{2}{(2m + 1)\pi} \qquad (9.52)$$

The amplitudes of the 1^{st}, 2^{nd} and 3^{rd} side lobes are therefore $\frac{2}{3\pi}$, $\frac{2}{5\pi}$ and $\frac{2}{7\pi}$ respectively. The first sidelobe, therefore, has an amplitude of about 21% of the peak of the radiation pattern, the second sidelobe is about 13% of the peak and so on.

It is interesting to note that for large N, the sidelobe amplitudes are practically independent of any of the array parameters like the inter element spacing, progressive phase shift and the number of elements. In other words, for large uniform arrays the relative amplitudes in the radiation pattern are fixed. A typical radiation pattern of a uniform array is shown in Fig. 9.15.

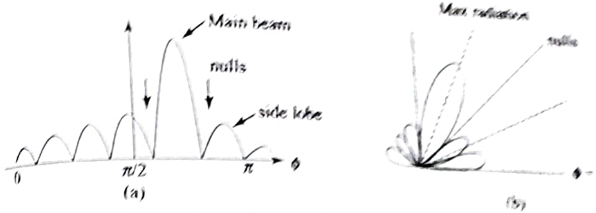


Fig. 9.15 *Radiation pattern drawn as Cartesian plot (a) and Polar plot (b).*

EXAMPLE 9.7 A three element uniform array is to be designed to get maximum radiation in the broadside direction. There is, however, an interference source located in a direction of 45° from the array axis. The interference is to be completely suppressed by placing a null in the direction of the interference source. Design the array.

Solution:

The array is uniform, so design parameters are only d and δ. The maximum radiation is in broadside direction giving $\delta = 0$. The array phase ψ then is

$$\psi = \frac{2\pi}{\lambda} d\cos\phi \qquad (\text{since } \delta = 0)$$

The nulls correspond to $\frac{N\psi}{2} = \pm m\pi$

$$\Rightarrow \quad \frac{3\psi}{2} = \frac{3\pi d}{\lambda}\cos\phi_n = \pm m\pi, \qquad m = 1, 2$$

For killing the interference, one of the nulls is to be placed in the direction of the interference. That is $\phi_n = 45° = \pi/4$

$$\Rightarrow \qquad d = \pm \frac{m\lambda}{3\cos\phi_n} = \pm \frac{m\lambda\sqrt{2}}{3}$$

Smallest array will correspond to smallest d i.e. smallest m. Taking $m = 1$, the array spacing is $d = \frac{\sqrt{2}}{3}\lambda$.

EXAMPLE 9.8 An end fire array with inter element spacing of 0.3λ, has 10 elements. Find the directions of the nulls and approximate directions of the sidelobes. Sketch the radiation pattern of the array.

Solution:

Let the end-fire array have maximum radiation along $\phi = 0$. We then have

$$\delta = -\beta d = -\frac{2\pi}{\lambda} 0.3\lambda = -0.6\pi$$

$$\text{and} \quad \psi = \frac{2\pi}{\lambda} 0.3\lambda \cos\phi - 0.6\pi = 0.6\pi(\cos\phi - 1)$$

The directions of nulls correspond to

$$\frac{N\psi}{2} = \frac{10\psi}{2} = 3\pi(\cos\phi - 1) = \pm m\pi, \qquad m = 1, 2, \ldots$$

$$\Rightarrow \qquad \cos\phi = 1 \pm \frac{m}{3}$$

For +ve sign, the RHS is > 1, hence, only the −ve sign is to be considered.

$$\Rightarrow \qquad \cos\phi = 1 - \frac{m}{3} \qquad\qquad m = 1, 2, 3, 4, 5, 6$$

There are six nulls at (Note: $0 < \phi < \pi$)

$$\phi_n = \cos^{-1}(2/3), \cos^{-1}(1/3), \cos^{-1}(0), \cos^{-1}(-1/3), \cos^{-1}(-2/3), \cos^{-1}(-1)$$
$$= 48.19^\circ, 70.53^\circ, 90^\circ, 109.47^\circ, 131.81^\circ, 180^\circ$$

The sidelobes are approximately half way between the nulls.
The directions of side lobes therefore are

$$\phi_{SL} = \cos^{-1}(1/2), \cos^{-1}(1/6), \cos^{-1}(-1/6), \cos^{-1}(-1/2), \cos^{-1}(-5/6)$$
$$= 60^\circ, 80.4^\circ, 99.59^\circ, 120^\circ, 146.44^\circ$$

Note: The directions of the sidelobes are midway between the nulls in the ψ-domain and not in the ϕ-domain.

9.2.4 Half-Power Beam Width (HPBW)

The half-power beam-width, as defined in Chapter 8, is the angular separation between two directions, one on either side of the direction of maximum radiation, along which the electric field reduces to $\frac{1}{\sqrt{2}}$ of its maximum value (see Fig. 9.16).

Since, the Poynting vector, i.e. the power density is proportional to the square of the electric field, the power density along ϕ_1 and ϕ_2 is half of that along ϕ_{max}. Hence the angular width $(\phi_2 - \phi_1)$ is called the 'half-power beamwidth' of the array. The half-power angles ϕ_1 and ϕ_2 can be calculated by equating $|E|$ to $\frac{1}{\sqrt{2}}$ in Eqn (9.35) as

$$\frac{1}{N}\left|\frac{\sin\left(\frac{N\psi}{2}\right)}{\sin\left(\frac{\psi}{2}\right)}\right| = \frac{1}{\sqrt{2}} \tag{9.53}$$

Equation (9.53) has to be solved numerically to obtain the half-power angles ϕ_1 and ϕ_2, which can then be used to get the exact value for the HPBW.

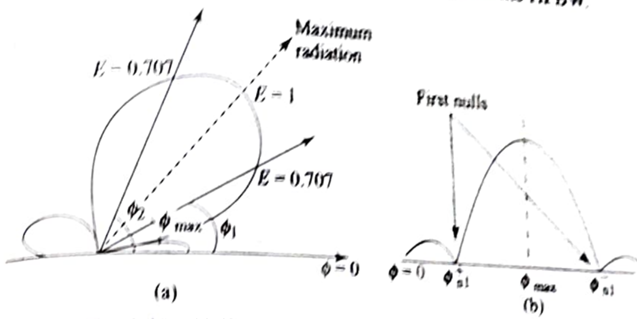


Fig. 9.16 *Half-power beam width of a radiation pattern.*

Although, the above mentioned procedure provides an accurate estimate of HPBW, it involves numerical solution of Eqn (9.53). In practice where a very accurate estimate of the HPBW is not needed, one can use some approximate approach to quickly estimate the HPBW of a large array.

If we look at Fig. 9.16 we notice that the width between the first nulls (BWFN) is approximately twice the HPBW. Since, finding the directions of the nulls is much easier, we can first obtain BWFN and then HPBW.

Assuming that the first nulls are more or less symmetric with respect to the direction of the maximum radiation, the HPBW is approximately $\phi_{n1}^+ - \phi_{max}$ or $\phi_{max} - \phi_{n1}^-$. ϕ_{n1}^- and ϕ_{n1}^+ denote directions of the first nulls (see Fig. 9.16).

$$\phi_{HPBW} = \phi_2 - \phi_1 \approx \frac{1}{2}(\phi_{n1}^+ - \phi_{n1}^-)$$
$$\approx \phi_{n1}^+ - \phi_{max} \approx \phi_{max} - \phi_{n1}^- \tag{9.54}$$

We have assumed here that $\phi_{max} \approx (\phi_{n1}^+ + \phi_{n1}^-)/2$. It should be noted that as ϕ_{max} approaches zero, the first null corresponding to ϕ_{n1}^- may not be visible, and when ϕ_{max} approaches π, the first null corresponding to ϕ_{n1}^+ may not be visible (see Fig. 9.17). Hence, depending upon the direction of the maximum radiation, appropriate expression in Eqn (9.54) should be used.

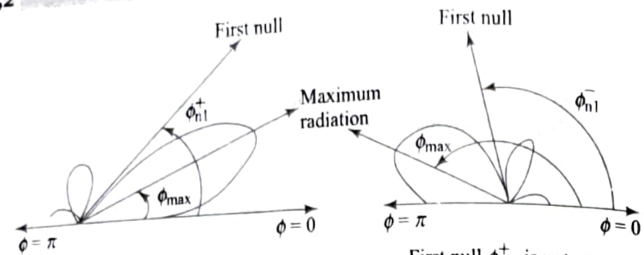


Fig. 9.17 *Radiation patterns showing directions of maximum radiation and nulls.*

Without losing generality let us consider the case in Fig. 9.18(a). For this case

$$\phi_{\text{HPBW}} \approx \phi_{n1}^+ - \phi_{max} \tag{9.55}$$

From Eqn. (9.46) we have, (for $m = 1$)

$$\cos\phi_{n1}^+ = \cos\phi_{max} - \frac{\lambda}{dN} \tag{9.56}$$

$$\Rightarrow \quad \cos\phi_{n1}^+ - \cos\phi_{max} = -\frac{\lambda}{dN} \tag{9.57}$$

$$\Rightarrow \quad 2\sin\left(\frac{\phi_{n1}^+ - \phi_{max}}{2}\right)\sin\left(\frac{\phi_{n1}^+ + \phi_{max}}{2}\right) = \frac{\lambda}{dN} \tag{9.58}$$

Taking $\phi_{\text{HPBW}} \simeq \phi_{n1}^+ - \phi_{max}$ from Eqn (9.54), we can rewrite Eqn. (9.58) as

$$2\sin\left(\frac{\phi_{\text{HPBW}}}{2}\right)\sin\left(\frac{2\phi_{max} + \phi_{\text{HPBW}}}{2}\right) = \frac{\lambda}{dN} \tag{9.59}$$

$$\Rightarrow \quad 2\sin\left(\frac{\phi_{\text{HPBW}}}{2}\right)\left\{\sin\phi_{max}\cos\left(\frac{\phi_{\text{HPBW}}}{2}\right) + \cos\phi_{max}\sin\left(\frac{\phi_{\text{HPBW}}}{2}\right)\right\} = \frac{\lambda}{dN} \tag{9.60}$$

For large array (i.e. for $N >> 1$), the $\phi_{\text{HPBW}} << 1$ and we can approximate Eqn (9.60) (through approximations $\sin x \approx x$ and $\cos x \approx 1$ for $x << 1$) as

$$2\frac{\phi_{\text{HPBW}}}{2}\left\{\sin\phi_{max} + \frac{\phi_{\text{HPBW}}}{2}\cos\phi_{max}\right\} = \frac{\lambda}{dN} \tag{9.61}$$

$$\Rightarrow \quad \phi_{\text{HPBW}}^2\cos\phi_{max} + 2\sin\phi_{max}\phi_{\text{HPBW}} - \frac{2\lambda}{dN} = 0 \tag{9.62}$$

The approximate value of ϕ_{HPBW} can be obtained by solving the quadratic Eqn (9.62) as

$$\phi_{\text{HPBW}} = \frac{-2\sin\phi_{max} + \sqrt{4\sin^2\phi_{max} + \frac{8\lambda}{dN}\cos\phi_{max}}}{2\cos\phi_{max}} = \frac{-\sin\phi_{max} + \sqrt{\sin^2\phi_{max} + \frac{2\lambda}{dN}\cos\phi_{max}}}{\cos\phi_{max}} \tag{9.63}$$

From Eqn. (9.63) we can note that as ϕ_{max} increases from 0 to $\frac{\pi}{2}$, the beam width ϕ_{HPBW} decreases monotonically. That is, when $\phi_{max} = 0$, (end-fire array) the HPBW is maximum and when $\phi_{max} = \frac{\pi}{2}$ (broad-side array) the HPBW is minimum. Using Eqn (9.62) we can get the HPBW for two extreme cases. For a broadside array ($\phi_{max} = \frac{\pi}{2}$), the HPBW is

$$\phi_{\text{HPBW}} = \frac{\lambda}{dN} \simeq \frac{\lambda}{\text{Length of the array}} \tag{9.64}$$

For an end-fire array ($\phi_{max} = 0$), the HPBW is

$$\phi_{\text{HPBW}} = \sqrt{\frac{2\lambda}{dN}} = \sqrt{\frac{2\lambda}{\text{Length of the array}}} \tag{9.65}$$

The length of the array is $(N-1)d$. However for $N >> 1$ the length of the array is approximated to dN.

From Eqns (9.64) and (9.65) it is important to note that, for a broadside array, the HPBW is inversely proportional to the length of the array, whereas, it is inversely proportional to the square root of the length of the array for an end-fire array. We, therefore, find that for a given array length, the broadside array has much smaller HPBW compared to that for the end-fire array.

EXAMPLE 9.9 A large uniform beam scanning array with full angular coverge is to be designed to get maximum half power beam-width of 10°. If the maximum allowed array spacing is $\lambda/2$, find the minimum number of elements needed for the array. How much is the variation in the HPBW during the scanning if the beam is scanned over the full range?

Solution:

The array has largest HPBW when it is in the end-fire direction. So we get from Eqn (9.65),

$$10^\circ = \frac{\pi}{18} = \sqrt{\frac{2\lambda}{dN}}$$

$$\Rightarrow \quad dN = 2\lambda\left(\frac{18}{\pi}\right)^2$$

Since, maximum value of d is $\lambda/2$, minimum number of elements is

$$N = 2\lambda \left(\frac{18}{\pi}\right)^2 \frac{2}{\lambda} = 4\left(\frac{18}{\pi}\right)^2 = 131$$

The minimum HPBW is observed for the broadside array. Therefore, the minimum beam-width seen while scanning is

$$\phi_{HPBW} = \frac{\lambda}{dN} = \frac{\pi^2}{648} = 0.0152 \text{ rad} = 0.873^\circ$$

Therefore, the variation in the HPBW while scanning is from 0.873° to 10°.

9.2.5 Directivity of Uniform Array

As mentioned earlier, the directivity of an antenna defines its power focusing capability. Once the radiation pattern of an antenna is known, its directivity can be calculated using Eqn (8.106).

For an N-element uniform array therefore the directivity is given by

$$D = \frac{4\pi}{\int\int |AF|^2 d\Omega} \qquad (9.66)$$

where AF is the normalized radiation pattern of the array, and is given by Eqn (9.35). Obviously, for a general uniform array the Eqn (9.66) has to be evaluated numerically. For a large array however one can make appropriate approximations to the integral in Eqn (9.66). The integral can be replaced by the solid angle within the half-power beam-width of the array. To get a better insight let us find the directivity of a N element array in the broadside and in the end-fire mode.

Figure 9.18 shows main beam of the array in three dimensional space. It is apparent that although the planar radiation patterns look similar for the broadside and the end-fire arrays, in three dimensions they have totally different appearances. The three dimensional main beam pattern for a broadside array looks almost like a disk, whereas, for the end-fire array, it appears like an elongated balloon. For any arbitrary direction of maximum radiation, the main beam looks like a hollow cone as shown in Fig. 9.18.

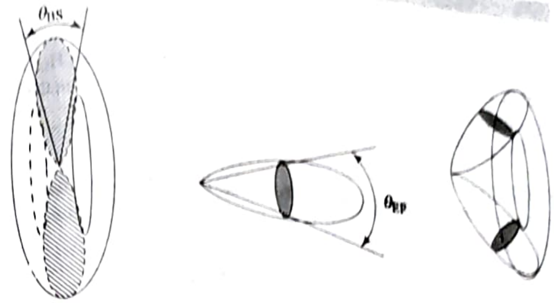
If we denote the HPBW of the broadside and the end-fire array respectively by ϕ_{BS} and ϕ_{EF}, from Eqns (9.64) and (9.65) we get

$$\phi_{BS} = \frac{\lambda}{dN} \qquad (9.67)$$

$$\phi_{EF} = \sqrt{\frac{2\lambda}{dN}} \qquad (9.68)$$

The solid angle for the broadside array is approximately

$$\Omega_{BS} \simeq 2\pi\phi_{BS} \qquad (9.69)$$



(a) Broadside array (b) End-Fire array (c) General array

Fig. 9.18 *Three-dimensional view of the radiation patterns of the broadside, end-fire and a general array.*

and that for the end-fire array is

$$\Omega_{EF} \simeq \pi\left(\frac{\phi_{EF}}{2}\right)^2 \qquad (9.70)$$

The directivities of the two arrays therefore are

Broadside Array:

$$D_{BS} \simeq \frac{4\pi}{2\pi\phi_{BS}} = \frac{2dN}{\lambda} \qquad (9.71)$$

End-fire Array:

$$D_{EF} \simeq \frac{4\pi}{\pi(\frac{\phi_{EF}}{2})^2} = \frac{16}{(\frac{2\lambda}{dN})} = \frac{8dN}{\lambda} \qquad (9.72)$$

This result is quite interesting. The directivity of the end-fire array is about 4 times higher than that of the broadside array of the same length. However, if we compare the HPBW of the two arrays we note that

$$\frac{\phi_{BS}}{\phi_{EF}} = \frac{\lambda}{dN}\sqrt{\frac{dN}{2\lambda}} = \sqrt{\frac{\lambda}{2dN}} \qquad (9.73)$$

Since, $(dN) >> \lambda$ (large array), we get $\phi_{BS} << \phi_{EF}$. In the first look then one may wrongly conclude that the broadside array has high directivity since it has narrower beam compared to the end-fire array. One should, therefore, learn an important lesson from this, and that is, one must always make conclusions on the basis of the three dimensional radiation patterns and not only on the basis of planar radiation patterns. In fact, one must develop a habit of always visualising the antenna radiation patterns as three dimensional solids.

EXAMPLE 9.10 A two element array consists of collinear Hertz dipoles. The element spacing is $\lambda/2$. Find the directivity of the array when the elements are excited in phase.

Solution:

Since the antenna element is the Hertz dipole, the primary element radiation pattern is $\sin\theta$.

The radiation pattern of the array is

$$|E| = 2\cos(\psi/2)\sin\theta$$

$$\text{where } \psi = \frac{2\pi}{\lambda}d\cos\theta + \delta$$

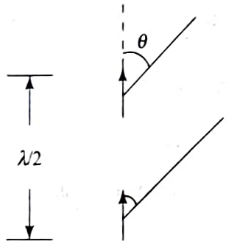


Fig. 9.19 *Array of colinear Hertz dipoles.*

Elements are excited in phase and hence $\delta = 0$. The array phase ψ is

$$\psi = \frac{2\pi}{\lambda}\frac{\lambda}{2}\cos\theta = \pi\cos\theta$$

The normalized field pattern is

$$\Rightarrow \qquad |E| = \sin\theta\cos\left(\frac{\pi\cos\theta}{2}\right)$$

$$\text{Directivity } D = \frac{4\pi}{\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\sin^2\theta\cos^2(\frac{\pi\cos\theta}{2})\cdot\sin\theta\cdot d\theta d\phi}$$

$$D = \frac{4\pi}{2\pi\int_0^{\pi}\sin^3\theta\cos^2(\frac{\pi\cos\theta}{2})d\theta} = \frac{2}{I} \text{ (say)}$$

$$\text{Substituting for } \frac{\pi\cos\theta}{2} = t, \quad \sin\theta d\theta = -\frac{2}{\pi}dt$$

$$I = \int_{-\pi/2}^{\pi/2}\frac{2}{\pi}\left(1 - \frac{4t^2}{\pi^2}\right)\cos^2 t\cdot dt$$

$$= 0.8677$$

The directivity of the array is

$$D = \frac{2}{0.8677} = 2.3$$

EXAMPLE 9.11 Derive expressions for the directivity of two element broadside and end-fire arrays. How much the directivity varies from one case to the other when $d = \lambda/4$.

Solution:

For an array in general

Fig. 9.20 *Two element array for Example 9.11.*

$$\psi = \frac{2\pi}{\lambda}d\cos\theta + \delta$$

and the radiation pattern is

$$|E| = \cos(\psi/2)$$

$$\text{Directivity of the array } D = \frac{4\pi}{\int\int |E|^2\sin\theta d\theta d\phi}$$

Broadside array

$$\Rightarrow \qquad \psi = \frac{2\pi}{\lambda}d\cos\theta$$

$$D = \frac{4\pi}{2\pi\int_0^{\pi}\cos^2(\frac{\pi d}{\lambda}\cos\theta)\sin\theta d\theta}$$

Integral can be solved by putting $\pi \frac{d}{\lambda} \cos\theta = t$

$$\Rightarrow \qquad D = \frac{2}{1 + \frac{\sin(2\pi d/\lambda)}{(2\pi d/\lambda)}} = \frac{2}{1 + (\sin x)/x}$$

where $x = 2\pi d/\lambda$.

End-fire Array

In this case, $\delta = -\beta d = -\frac{2\pi d}{\lambda}$.

$$\Rightarrow \qquad \psi = \frac{2\pi d}{\lambda}(\cos\theta - 1) = x(\cos\theta - 1)$$

$$D = \frac{4\pi}{2\pi \int_0^\pi \cos^2\left\{\frac{\pi d}{\lambda}(\cos\theta - 1)\right\} \sin\theta d\theta}$$

$$= \frac{2}{1 + \sin(2x)/(2x)}$$

When $d = \lambda/4$, $x = \frac{2\pi}{\lambda} \times \frac{\lambda}{4} = \pi/2$.

Directivity of the broadside array

$$D_{BS} = \frac{2}{1 + \sin(\pi/2)/(\pi/2)} = 1.222$$

Directivity of the end-fire array

$$D_{EF} = \frac{2}{1 + \sin(\pi)/\pi} = 2$$

9.2.6 Grating Lobe

For a uniform array the direction of the main beam corresponds to $\psi = 0$. For $\psi = 0$, the fields due to individual elements add to give the maximum radiation from the array. The obvious thing then is, if the radiation is maximum for $\psi = 0$, so it would be for $\psi = 2m\pi$, where m is an integer. However, since $\psi = \beta d \cos\phi + \delta$, the angle ϕ for $\psi = 0$ and $\psi = 2m\pi$ are not the same. In other words, there may be some angles other than ϕ_{max}, where, the field would be maximum (identical to that of the main beam). It means that, there may be beams in a radiation pattern which are identical to the main beam. These beams are called the 'grating lobes' of the array. Hence, a grating lobe is a beam identical to the main beam but in undesired direction. Depending upon the inter element spacing of the array, there may be multiple grating lobes.

Since, in the array design the primary objective is to send the power in the direction of the main beam, presence of a grating lobe reduces the power efficiency in the direction of the main beam. The power which ideally should go in the main beam now gets distributed amongst the various grating lobes. The grating lobe therefore is an undesirable feature of a radiation pattern and should be avoided in the array design. In the following we find the condition under which there would be no grating lobe.

In the expression for the radiation pattern of a uniform array, Eqn (9.35), ψ represents a phase which is a combination of the space phase and the electrical phase. If we look at Eqn (9.35) however as a mathematical expression, we can plot the radiation pattern as a function ψ for all values of ψ between $-\infty$ to ∞. Obviously, only those values of ψ would be physically meaningful which correspond to $0 \leq \phi \leq \pi$. The range of ψ corresponding to $0 \leq \phi \leq \pi$ is called the '*visible range*' of ψ. Substituting for two limits of ϕ in Eqn (9.28) we get the visible range of ψ as

$$\underbrace{-\beta d + \delta}_{\psi_{min}} \leq \psi_{\text{visible}} \leq \underbrace{\beta d + \delta}_{\psi_{max}} \tag{9.74}$$

If the direction of the main beam is ϕ_{max}, $\delta = -\beta d \cos\phi_{max}$ and Eqn (9.74) can be re-written as

$$-\beta d(1 + \cos\phi_{max}) \leq \psi_{\text{visible}} \leq \beta d(1 - \cos\phi_{max}) \tag{9.75}$$

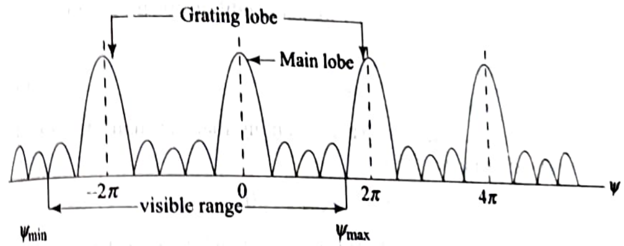


Fig. 9.21 *Radiation pattern of an array as a periodic function of ψ.*

From Eqns (9.74) and (9.75) it is clear that as the direction the of main beam changes, the visible range of ψ slides along the ψ axis but the width of the visible range $(\psi_{max} - \psi_{min})$ is always $2\beta d$. Figure 9.21 shows the array pattern as a function of ψ, and some arbitrary visible range of ψ. (Note that the array pattern is a periodic function of ψ with period 2π). In this case, there are two beams within the visible range of ψ. The beam corresponding to $\psi = 0$ is the main beam, and the beam corresponding $\psi = -2\pi$ is the grating lobe. Clearly, for wider visible range, i.e. for larger inter-element spacing, there would be more grating lobes. It then appears that to avoid grating lobes, one must reduce the inter-element spacing of the array. However, reduction in the inter-element spacing is not favourable as it reduces the length of the array (for given N) and consequently the directivity of the array. The question then reduces to - 'what is the maximum permissible inter-element spacing for which there is no grating lobe?'. We will see as follows that the answer depends upon the choice of the direction of the main beam.

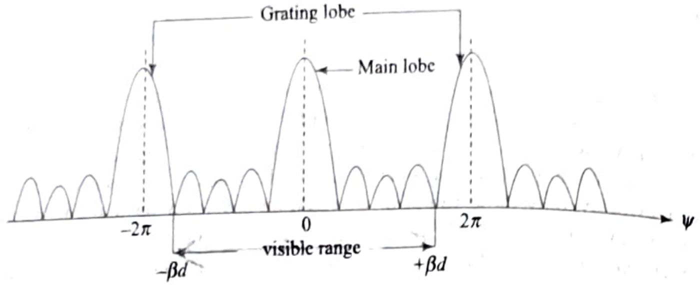


Fig. 9.22 *Radiation pattern and visible range of a broadside array.*

For a broadside array, $\phi_{max} = \frac{\pi}{2}$, and the visible range of ψ is

$$-\beta d \le \psi_{\text{visible}} \le \beta d \tag{9.76}$$

The visible range is symmetrically placed around $\psi = 0$ as shown in Fig. 9.22.

From Fig. 9.22 it is clear that to avoid the peak of the grating lobes, βd should be less than 2π. That is, for a broad-side array, if $\beta d < 2\pi$, there would be no grating lobe. Substituting for $\beta = \frac{2\pi}{\lambda}$, we get condition for no grating lobe of a broadside array as

$$d < \lambda \tag{9.77}$$

For an end fire array, on the other hand, assuming that the main beam is along $\phi_{max} = 0$, the visible range of ψ is

$$-2\beta d \le \psi_{\text{visible}} \le 0 \tag{9.78}$$

The array pattern and the visible range of ψ is shown in Fig. 9.23. To avoid grating lobe in this case we must have

$$2\beta d < 2\pi \tag{9.79}$$

$$\Rightarrow \quad d < \frac{\lambda}{2} \tag{9.80}$$

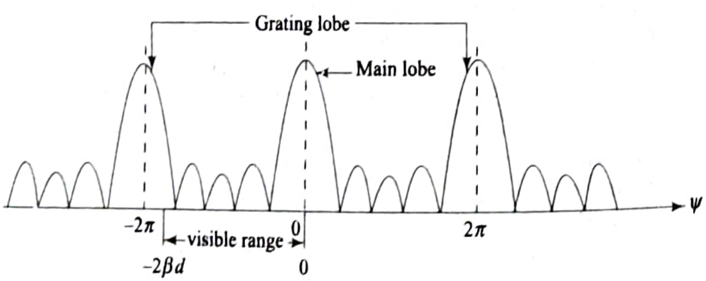


Fig. 9.23 *Radiation pattern and visible range of an end-fire array.*

From the above discussion we conclude that if the inter-element spacing is $< \frac{\lambda}{2}$, there is no grating lobe irrespective of the direction of the main beam. In practice therefore all the scanning beam phased arrays have inter-element spacing of $\frac{\lambda}{2}$. Note that to avoid grating lobes, strictly the inter-element spacing should be $< \frac{\lambda}{2}$, but in practice the scanning range of the main beam is a little short of 0 to π and therefore $d = \frac{\lambda}{2}$ is used without getting any grating lobe.

Before closing the discussion on the grating lobes, it should be mentioned that the grating lobes can also be suppressed by properly choosing the primary radiation pattern of the array elements. With non-isotropic array elements, it is possible to increase the inter-element spacing beyond $\frac{\lambda}{2}$ without encountering the grating lobes.

EXAMPLE 9.12 A two element array of isotropic antennas is located at a height of $\lambda/4$ above the ground as shown in the Fig. 9.24(a). What maximum scanning angle of the main beam is possible so as not to have grating lobe equal to the main beam? The element spacing is λ.

Solution:

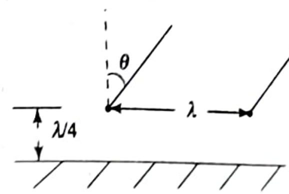


Fig. 9.24(a)

As discussed earlier, the ground can be replaced by the images of the radiating elements. Each element can be combined with its image to form the so called primary element as shown in Fig. 9.24(b).

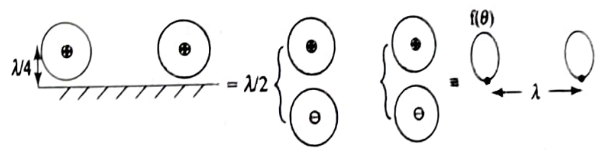


Fig. 9.24(b)

The element and its image form a two-element array. The image is 180° out of phase with respect to the element. For this array then we have the array phase

$$\psi_1 = \frac{2\pi}{\lambda}\frac{\lambda}{2}\cos\theta - \pi$$
$$= \pi(\cos\theta - 1)$$

The primary element pattern $f(\theta) = \cos(\psi_1/2) = \cos\left[\frac{\pi(\cos\theta - 1)}{2}\right]$

The two isotropic elements above the ground plane are equivalent to two elements in the free-space with primary radiation pattern $f(\theta)$.

For the horizontal array $\psi = \frac{2\pi}{\lambda}\lambda \sin\theta + \delta = 2\pi \sin\theta + \delta$

The array factor for the horizontal array is

$$AF = \cos(\psi/2) = \cos\left\{\left(\pi \sin\theta + \frac{\delta}{2}\right)\right\}$$

The total array pattern is

$$E(\theta) = f(\theta)\cdot AF = \cos\left\{\frac{\pi(\cos\theta - 1)}{2}\right\}\cos\left\{\pi\sin\theta + \frac{\delta}{2}\right\}$$

Now when the array is excited for broadside beam,

$\delta = 0$ giving $\psi = 2\pi \sin\theta$.

The direction of the grating lobe corresponds to $\psi = \pm 2m\pi$

$$\Rightarrow \quad 2\pi \sin\theta_g = \pm 2m\pi$$

$$\Rightarrow \quad \theta_g = \pm\pi/2$$

When the array is broadside, there are two grating lobes one at $\theta = -\pi/2$ and other at $\theta = +\pi/2$, i.e. both are along the horizon.

However, since $f(\theta)$ has the nulls along $\theta = \pm\pi/2$, the grating lobe is suppressed. That is to say that due to the presence of the ground the grating lobe is suppressed.

As we now scan the beam towards say +ve θ (clockwise), due to variation in $f(\theta)$ the main beam amplitude reduces. At the same time the grating lobe along $\theta = -\pi/2$ moves in clockwise direction, and its amplitude increases. Since, $f(\theta)$ is symmetric around $\theta = 0$, the grating lobe and the main beam become equal in amplitude when two are symmetrically placed around $\theta = 0$. That is, $\theta_g = -\theta_0$. For main beam we have $\psi = 0$ and for grating lobe we have $\psi = -2\pi$. Hence we get

$$\psi = 2\pi\sin\theta_0 + \delta = 0$$

$$\text{and} \quad 2\pi\sin\theta_g + \delta = -2\pi$$

$$\Rightarrow \quad 2\pi\left\{\sin\theta_0 - \sin\theta_g\right\} = 2\pi$$

Substituting $\theta_g = -\theta_0$, $\sin\theta_g = -\sin\theta_0$, we get

$$\sin\theta_0 = 1/2$$

$$\Rightarrow \quad \theta_0 = \pi/6 = 30^\circ$$

The main beam therefore can be scanned by $\pm 30^\circ$ around broadside direction without having grating lobe equal to or greater than the main beam.

9.3 ARRAY SYNTHESIS

In the previous sections we investigated the array analysis problem. That is, finding the array pattern when the array is specified. As a simpler case, uniform array was investigated to get a specific type of radiation pattern. By controlling three parameters namely, element spacing, number of elements and the progressive phase shift, the directions of the main beam, and nulls were controlled. However, we noticed, that with only three controlling parameters we can atmost control three features of the radiation pattern. Also, we find that for a uniform array, we hardly have any control over the side-lobe level.

In practice we need to design arrays for a variety of radiation patterns with a very tight control over the location of nulls and the side-lobe level. This obviously cannot be achieved within the frame work of the uniform arrays. To increase the degrees of freedom, we essentially have to control individual array elements. In general, of course this can be done by controlling both the complex current and the location of the elements. In the following discussion, however, we keep the elements uniformly spaced and control only the complex currents of the individual elements.

The array synthesis is the inverse of the array analysis. In array synthesis, the radiation pattern is specified and we are asked to find the array excitation. Array analysis is an academic exercise whereas, the array synthesis is an engineering design problem. In practice, a user specifies the radiation pattern and the antenna engineers are asked to design an array to get that radiation pattern.

9.3.1 Radiation Pattern for a General Array

Let us consider an array of $(N+1)$ isotropic elements with uniform spacing d. Let the array elements be excited with currents $I_0, I_1, \ldots, I_N$ as shown in Fig. 9.25. The currents in general are complex, and the phase and amplitude of any current are measured with respect to the N^{th} element. Then, by definition $I_N = 1\angle 0$

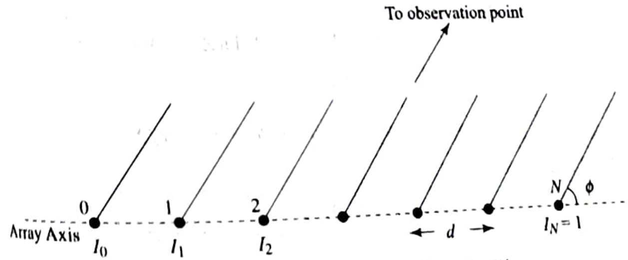


Fig. 9.25 *General N-element linear array.*

If we define now $\psi = \beta d\cos\phi$, the array factor can be written as

$$AF = |I_0 + I_1 e^{j\psi} + I_2 e^{j2\psi} + \ldots + e^{jN\psi}| \qquad (9.81)$$

Note that the progressive phase shift δ which we used in case of the uniform array, has now been absorbed in the phases of the individual currents.

If we now define a complex variable $Z \equiv e^{j\psi}$, the array factor can be written as

$$AF = |I_0 + I_1 Z + I_2 Z^2 + \dots\dots\dots + Z^N| \tag{9.82}$$

From Eqn (9.82) we note that the radiation pattern of an array can be represented by a polynomial of a complex variable Z, with complex coefficients. Since, Z is only a phase function, (i.e. $|Z| = 1$) Z lies on the unit circle in the complex plane.

Now a polynomial of N^{th} degree can be factorized into N monomials to give

$$AF = |(Z - \zeta_1)(Z - \zeta_2)\dots(Z - \zeta_N)| \tag{9.83}$$

where $\zeta_1, \zeta_2, \dots, \zeta_N$ are the zeros of the polynomial.

When $Z = \zeta_1$ or ζ_2 or $\cdots \zeta_N$, the array factor $AF = 0$. That is, $Z = \zeta_1$, $Z = \zeta_2, \cdots$ etc. correspond to the nulls of the radiation pattern. It should be noted however that physically there will be a null in the radiation pattern provided the zeros of the polynomial $\zeta_1, \zeta_2, \cdots$ etc. lie on the unit circle in the Z plane, because for all physical angles ϕ, Z must lie on the unity circle. Said differently, only those nulls of the array are visible for which the zeros of the array polynomial lie on the unity circle.

EXAMPLE 9.13 A uniform broadside array of 4 elements has inter-element spacing of $\lambda/3$. Find the directions of the nulls. Are the nulls visible?

Solution:

For a uniform broadside array, all the currents are equal. The array factor is given as

$$AF = 1 + Z + Z^2 + Z^3$$

which can be factorised to get

$$AF = (Z + 1)(Z^2 + 1) = (Z + 1)(Z + j)(Z - j)$$

The nulls are at $Z = e^{j\psi} = -1, \pm j$

$$\Rightarrow \qquad \psi = \frac{2\pi}{\lambda} d \cos\phi = \pi, \pm \pi/2$$

For $d = \lambda/3$,

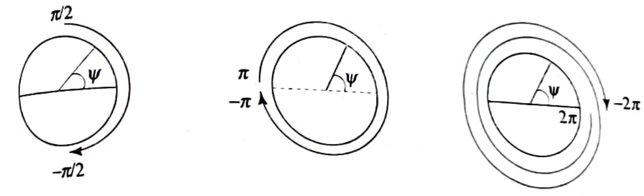
$$\psi = \frac{2\pi}{\lambda}\frac{\lambda}{3}\cos\phi = \pi, \pm\pi/2$$

$$\Rightarrow \qquad \cos\phi = 3/2, \pm 3/4$$

The null correesponding to $\psi = \pi$ is invisible since $\cos\phi$ is greater than unity. The visible nulls are $\psi = \pm\pi/2 \Rightarrow \cos\phi = \pm 3/4$. The nulls therefore are at $\phi = 41.4°, 138.69°$.

9.3.2 Circle Diagram

We have seen above, the physical angle, ϕ, ranges over 0 to π and the visible range of ψ is from βd to $-\beta d$. Therefore, as the angle ϕ varies from 0 to π, the point Z moves on the unity circle in the clockwise direction from βd to $-\beta d$. Total movement on the unit circle is equal to the visible range of ψ, i.e. $2\beta d$. Depending upon the inter-element spacing, the movement may be a fraction of the unity circle, or the full circle or many rotations around the circle. For example if $d = \frac{\lambda}{4}$, the ψ variation is from $\frac{\pi}{2}$ to $-\frac{\pi}{2}$ (see Fig. 9.26(a)). If $d = \frac{\lambda}{2}$, ψ varies from π to $-\pi$ (see Fig. 9.26(b)), and if say $d = \lambda$, ψ varies from $+2\pi$ to -2π i.e. two full clockwise rotations (see Fig. 9.26(c)). The Z for physical nulls now should not only lie on the unit circle but should also lie within the visible range of ψ. Also if the visible of range of $\psi > 2\pi$, then a null marked on the unity circle would correspond to two physical angles.



(a) $d = \lambda/4$ (b) $d = \lambda/2$ (c) $d = \lambda$

Fig. 9.26 *Circular diagrams for different inter-element spacings.*

For example, in Fig. 9.26(c) if there is a null at $\psi = \frac{\pi}{2}$ there will also be a null at $\psi = -\frac{3\pi}{2}$ since both angles correspond to the same point on the unit circle. The direction of null corresponding to $\psi = \frac{\pi}{2}$ is $\cos^{-1}(1/4) = 75.52°$ and that corresponding to $\psi = -\frac{3\pi}{2}$ is $\cos^{-1}(-3/4) = 138.59°$.

The important thing to note is, if the visible range of ψ is greater than 2π, the two nulls are not independent. As we change the direction of one null, the direction of the other null also changes correspondingly. It can therefore be concluded that if we want independent control of all the nulls, or for that matter every part of the radiation pattern, the visible range should be less than 2π. That is, the inter element spacing, d, should be less than $\frac{\lambda}{2}$. The unit circle with the nulls and the visible range marked on it is called the 'circle diagram' of the array.

9.3.3 Arrays Specified by Nulls

The circle diagram or Eqn. (9.83) now can be used to synthesize an array which is completely specified by its nulls. The array synthesis procedure is as follows.

Step 1 : From desired directions of nulls, ϕ_n, find corresponding ψ_n

Step 2 : Find the roots of the array polynomial, $\zeta_n = e^{j\psi_n}$

Step 3 : Substitute ζ_n in Eqn (9.83)

Step 4 : Expand the polynomial in Eqn (9.83) in ascending powers of Z

Step 5 : The coefficients of the polynomial are the complex excitation currents.

Note that, in this synthesis, one does not have any control over the direction of the main beam or the level of side-lobes, etc. This type of synthesis is useful in an environment where there are many intefering sources in different directions and the array has to minimize the effect of interference on the signal. The array is designed such that the nulls of the radiation pattern point in the directions of the interfering sources.

EXAMPLE 9.14 Design a linear array with minimum number of elements to have nulls at $\phi = 0, 45^\circ$ and 120° where ϕ is the angle measured from the axis of the array. Choose inter-element spacing of $\lambda/2$.

Solution:

For $d = \lambda/2$, we have $\psi = \frac{2\pi}{\lambda}\frac{\lambda}{2}\cos\phi = \pi\cos\phi$

Values of ψ for the three nulls are

$$\psi_1 = \pi\cos 0^\circ = \pi$$

$$\psi_2 = \pi\cos 45^\circ = \pi/\sqrt{2}$$

$$\psi_3 = \pi\cos 120^\circ = -\pi/2$$

The polynomial roots are

$$\zeta_1 = e^{j\psi_1} = e^{j\pi} = -1$$

$$\zeta_2 = e^{j\pi/\sqrt{2}} = \cos(\pi/\sqrt{2}) + j\sin(\pi/\sqrt{2})$$

$$\zeta_3 = e^{-j\pi/2} = -j$$

The array factor is

$$\begin{aligned}\text{AF} &= (Z-(-1))(Z-e^{j\pi/\sqrt{2}})(Z-(-j))\\ &= (Z+1)(Z-e^{j\pi/\sqrt{2}})(Z+j)\\ &= -je^{-j\pi/\sqrt{2}} + (j-(1+j)e^{j\pi/\sqrt{2}})Z + (1+j-e^{j\pi/\sqrt{2}})Z^2 + Z^3\\ &= (-0.7457+0.6057j) + (1.4014+0.81j)Z + (1.6+0.2j)Z^2 + Z^3\end{aligned}$$

The array has four elements. The array excitation coefficients are $(-0.7457 + j0.6057)$, $(1.4014 + j0.81)$, $(1.6 + j0.2)$ and 1.

EXAMPLE 9.15 For an array $\psi = \frac{\pi}{2}\cos\phi + \pi/4$. Find the visible range of ψ, and an array which has four equally spaced nulls in the visible range of ψ. Draw the circle diagram and find the directions of nulls.

Solution:

The visible range of ψ is from $\frac{\pi}{2}\cos 0^\circ + \frac{\pi}{4} = 3\pi/4 = 135^\circ$ to $\frac{\pi}{2}\cos 180^\circ + \frac{\pi}{4} = -\pi/4 = -45^\circ$.

The circle diagram and the equally spaced nulls are shown in the Fig. 9.27.

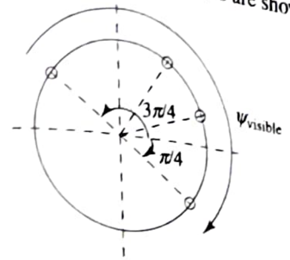


Fig. 9.27 *Circle diagram with nulls and visible range marked on it.*

The equi-spaced nulls are at $\psi = 135^\circ, 75^\circ, 15^\circ, -45^\circ$.

The directions of the nulls are

$$\phi = \cos^{-1}\left\{\frac{2}{\pi}(\psi - \pi/4)\right\}$$

Therefore the nulls are at $\phi = 0^\circ, 70.52^\circ, 109.47^\circ, 180^\circ$.

The array factor is

$$\begin{aligned}AR &= (Z-e^{j135^\circ})(Z-e^{j75^\circ})(Z-e^{j15^\circ})(Z-e^{-j45^\circ})\\ &= -1 - 1.2247(-1+j)Z + (2j)Z^2 - 1.2247(1+j)Z^3 + Z^4\end{aligned}$$

The array coefficients are $-1, -1.2247(-1+j), (2j), -1.2247(1+j)$and1.

9.3.4 General Array Synthesis

In Chapter 8, we established the Fourier transform relationship between the current distribution and the radiation pattern. An array is nothing but a discrete current distribution. An array can also be thought of as a sampled version of a continuous current distribution. Since, the radiation pattern is periodic over 2π, the Fourier transform relationship essentially reduces to the Fourier series. In the following sections we formulate the synthesis problem directly for the discrete current distribution, the antenna array.

Let us consider an array of $(2N+1)$ elements, and let us excite the elements with conjugate symmetry around the center element. The center element is the reference element. The currents on different elements are shown in Fig. 9.28. Since, the central element is taken as the reference element, by definition, its phase is zero i.e. I_0 is a real quantity.

Again defining $\psi = \beta d\cos\phi = \frac{2\pi}{\lambda}d\cos\phi$, where d is the inter-element spacing, the array pattern can be written as

$$|\text{AF}| = |I_N^* Z^{-N} + I_{N-1}^* Z^{-(N-1)} + \ldots + I_0 + \ldots + I_N Z^N| \tag{9.84}$$

$$|AF| = |I_N^* e^{-jN\psi} + I_{N-1}^* e^{-j(N-1)\psi} + \dots + I_0 + \dots + I_N e^{jN\psi}| \quad (9.85)$$

$$= |I_0 + (I_1 e^{j\psi} + I_1^* e^{-j\psi}) + \dots + (I_N e^{jN\psi} + I_N^* e^{-jN\psi})| \quad (9.86)$$

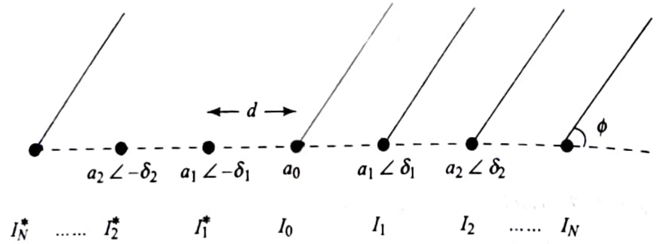


Fig. 9.28 *General array with conjugate current symmetry.*

Since, $I_1^* e^{-j\psi} = (I_1 e^{j\psi})^*$ and so on, we can combine the terms inside the backets to get

$$|AF| = |I_0 + 2Re(I_1 e^{j\psi}) + 2Re(I_2 e^{j2\psi}) + \dots + 2Re(I_N e^{jN\psi})| \quad (9.87)$$

Writing the complex current $I_k = a_k + jb_k$, and $I_0 = a_0 + j0$, we get

$$|AF| = \left| a_0 + 2\left\{ \sum_{k=1}^{N} a_k \cos k\psi - \sum_{k=1}^{N} b_k \sin k\psi \right\} \right| \quad (9.88)$$

In Eqn (9.88) the RHS is the Fourier series exapnsion of the array factor in ψ, with ψ having period of 2π.

The general array synthesis problem therefore reduces to Fourier expansion of the radiation pattern over ψ. The coefficients of the Fourier series give the complex currents on the array elements. It should be emphasized here that for exact representation of a function, in general, we need infinite terms in the Fourier series. The array factor however is a truncated Fourier series. Consequently, the synthesized array factor may be an approximation to the specified array pattern. Larger the number of elements in the arrays, closer is the synthesized pattern to the specified pattern. The array design may therefore be done under two constraints.

(i) *Acceptable root mean square error between the specified and the synthesized patterns*. In this case there is no control over the number of elements.

(ii) *Pre-decided number of elements*. In this case the Fourier synthesis gives the best fit (in the least square sense) between synthesized and the specified patterns.

The Fourier synthesis needs the radiation pattern specified in the ψ-domain, whereas in practice the pattern is specified as a function of ϕ. Conversion from ϕ to ψ needs the knowledge of the inter-element spacing, d. The array designer therefore has to select a priori the inter element spacing, d.

As we have seen above, the visible range of ψ is less than 2π if $d < \lambda/2$ and it is $> 2\pi$ if $d > \lambda/2$. The period for the Fourier series however is always 2π. Therefore, if d is chosen less than $\lambda/2$, there is some region of ψ within

the period 2π, which is outside the visible range. Since, the array pattern chosen outside the visible range of ψ does not affect the actual radiation pattern, one may choose appropriate function for the invisible portion of the pattern such that the number of elements in the array gets reduced. There is no systematic way to do this however. One may arbitrarily try different invisible patterns to give the best possible design.

On the other hand, if $d > \lambda/2$, since the visible range of ψ folds over itself, some portion of the radiation pattern cannot be controlled independently. The optimum choice therefore seems to be $d = \lambda/2$ for which there is uniqueness in the array design and also every part of the array pattern is independently controlled.

Following examples clarify these aspects.

EXAMPLE 9.16 Design a 5 element broadside array with 60° wide sectoral radiation pattern (see Fig. 9.29(a)). Assume inter-element spacing to be $\lambda/2$.

Solution:

Since, the array is the broadside array, the desired radiation pattern is given as

$$AF = 1 \qquad \text{for } \pi/3 < \phi < 2\pi/3$$
$$= 0 \qquad \text{otherwise}$$

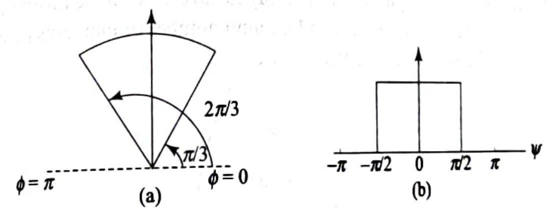


Fig. 9.29 *(a) sectoral radiation pattern (b) radiation pattern as a function of* ψ

Since, $\psi = \frac{2\pi}{\lambda}\lambda/2 \cos\phi = \pi \cos\phi$,

$$AR = f(\psi) = 1 \qquad \text{for } -\pi/2 < \psi < \pi/2$$
$$= 0 \qquad \text{for } \pi/2 < \psi < \pi$$
$$= 0 \qquad \text{for } -\pi < \psi < -\pi/2$$

Since, $f(\psi)$ is an even function, in the Fourier expansion we get $b_k \equiv 0$.

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\psi)d\psi = 1/2$$

$$a_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\psi)\cos k\psi d\psi$$

$$= \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} \cos k\psi d\psi = \frac{1}{k\pi}\sin\frac{k\pi}{2}$$

The array coefficients are

$$a_0 = 1/2, \quad a_1 = 1/\pi, \quad a_2 = 0, \quad a_3 = -1/3\pi, \quad a_4 = 0$$

The array pattern is

$$AR = -\frac{1}{3\pi}Z^{-3} + 0 + \frac{1}{\pi}Z^{-1} + \frac{1}{2} + \frac{1}{\pi}Z + 0 - \frac{1}{3\pi}Z^{-3}$$

$$= \frac{1}{\pi}\left\{-\frac{Z^{-3}}{3} + Z^{-1} + \frac{\pi}{2} + Z - \frac{Z^3}{3}\right\}$$

Note: Since $a_2 = 0$, we have included a_3 in the expansion. If a_2 was not equal to zero, a_0, a_1, a_2 would have made 5 elements in the array. The array elment spacing and the current excitations of the designed array are shown in the following Fig. 9.29(c).

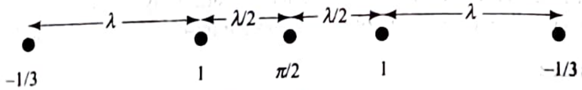


Fig. 9.29(c)

EXAMPLE 9.17 Design an end-fire array to have sectoral beam only in one direction along the axis of the array. Maximum number of elements permitted are seven and the width of the beam is 90°.

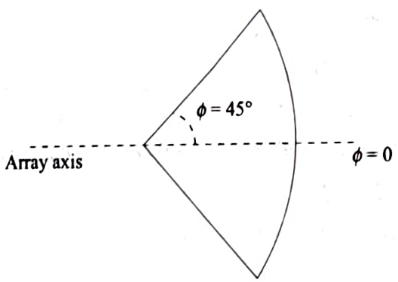


Fig. 9.30 *Desired radiation pattern.*

Solution:

The desired radiation pattern is (as shown in Fig. 9.30)

$$f(\phi) = 1 \qquad 0 < \phi < \pi/4$$
$$= 0 \qquad \pi/4 < \phi < \pi$$

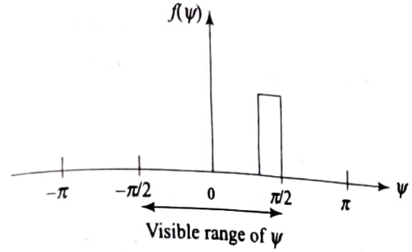


Fig. 9.31 *Radiation pattern as a function of ψ.*

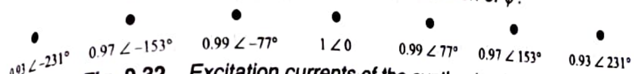


Fig. 9.32 *Excitation currents of the synthesized array.*

Note: The radiation pattern is a figure of revolution around $\phi = 0$ line. For an end-fire array, to have a 90° wide beam, the beam above $\phi = 0$ direction extends only over 45°. Since, we want beam only in one direction, d should be less than $\lambda/2$. Taking $d = \lambda/4$, we get

$$\psi = \frac{2\pi}{\lambda}\frac{\lambda}{4}\cos\phi = \frac{\pi}{2}\cos\phi$$

The visible range of ψ is from $\pi/2$ to $-\pi/2$ and the pattern is

$$f(\psi) = 1 \qquad \frac{\pi}{2\sqrt{2}} < \psi < \frac{\pi}{2}$$
$$= 0 \qquad -\frac{\pi}{2} < \psi < \frac{\pi}{2\sqrt{2}}$$

$f(\psi)$ as a function of ψ is shown in Fig. 9.31. Outside the visible range of ψ, $f(\psi)$ can be arbitrarily chosen. Choosing $f(\psi) = 0$ outside the visible range we can obtain a_0, a_k, b_k as

$$a_0 = \frac{1}{2\pi}\int_{\pi/2\sqrt{2}}^{\pi/2} d\psi = \frac{1}{2\pi}\left\{\frac{\pi}{2} - \frac{\pi}{2\sqrt{2}}\right\} = \frac{1}{4}\left(1 - \frac{1}{\sqrt{2}}\right) = 0.0732$$

$$a_k = \frac{1}{2\pi}\int_{\pi/2\sqrt{2}}^{\pi/2} \cos k\psi\, d\psi = \frac{1}{2\pi}\left[\frac{\sin k\psi}{k}\right]_{\pi/2\sqrt{2}}^{\pi/2}$$

$$b_k = -\frac{1}{2\pi}\int_{\pi/2\sqrt{2}}^{\pi/2} \sin k\psi\, d\psi = \frac{1}{2\pi}\left[-\frac{\cos k\psi}{k}\right]_{\pi/2\sqrt{2}}^{\pi/2}$$

The currents, therefore, are

$$I_0 = a_0 = 0.0732$$
$$I_1 = a_1 + jb_1 = 0.0165 + j0.0707$$

$$I_2 = a_2 + jb_2 = -0.0633 + 0.0314j$$
$$I_3 = a_3 + jb_3 = -0.043 - 0.0521j$$

Normalizing with respect to I_0, we get

$$I_0 = 1.0 = 1.0\angle 0^\circ$$
$$I_1 = 0.225 + j0.966 = 0.99\angle 77^\circ$$
$$I_2 = -0.865 + j0.43 = 0.97\angle 153^\circ$$
$$I_3 = -0.59 - 0.72 = 0.93\angle 231^\circ$$

The array configuration and excitation is as shown in Fig. 9.32. Fig. 9.33 shows the radiation pattern of the synthesized array.

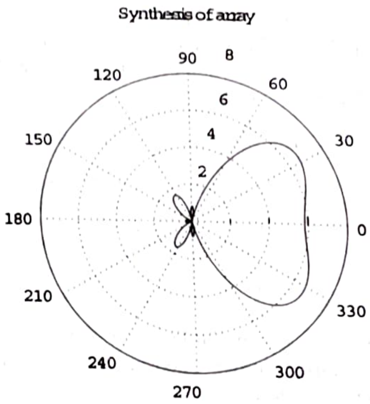


Fig. 9.33 *Synthesized radiation pattern.*

9.4 SPECIAL TYPES OF ARRAYS

In practice, we find some special types of arrays, like the Binomial array, Chebyshev array, Super directive arrays etc. In this section, we briefly discuss the concept of these arrays.

9.4.1 Binomial Array

This array is special in a sense that it has no side lobes. Since, there is always a sidelobe between two nulls (except the nulls around the main beam), no side-lobe means no nulls also, except the two nulls around the main beam. Now consider a two-element broadside array with $\lambda/2$ spacing. The radiation pattern for the array

can be written as a polynomial of 1st degree as

$$AF = 1 + Z \tag{9.89}$$

and the visible range of ψ is $+\pi$ to $-\pi$.

The array has a null at $Z = -1$, i.e. at $e^{j\beta(\lambda/2)\cos\phi} = -1 \Rightarrow \frac{2\pi}{\lambda}\frac{\lambda}{2}\cos\phi = \pm\pi \Rightarrow \phi = 0, \pi$. The maximum radiation is along $Z = 1$, i.e. at $\phi = \pi/2$.

If we now take an array which has a radiation pattern $(1+Z)^N$, neither the direction of maximum radiation nor the directions of the nulls change. However, the beam-width of the array reduces increasing the array directivity.

Expanding $(1+Z)^N$ we get

$$AF = (1+Z)^N = 1 + {}^N C_1 Z + {}^N C_2 Z^2 + \cdots + {}^N C_N Z^N \tag{9.90}$$

The polynomial coefficients ${}^N C_K$ are binomial coefficients. We, therefore, find that if an array of $(N+1)$ elements is excited with binomial coefficients, the array does not develop any side-lobes and additional nulls. It should be noted that since the binomial coefficients give tapered current distribution on the array, from Fourier transform property, its beam-width is larger compared to that of a uniform array of same size. The radiation pattern of a 9-element binomial array is shown in Fig. 9.34. A two-element array pattern is also shown for the sake of comparison.

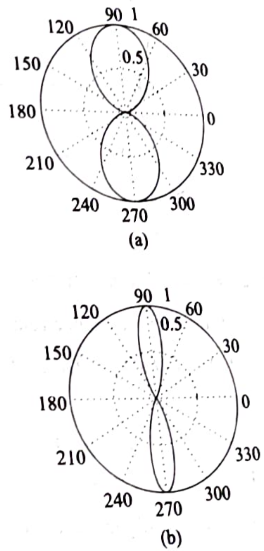


Fig. 9.34 *Radiation patterns of (a) 2-element and (b) 9-element binomial arrays.*

9.4.2 Chebyshev Array

Let us now consider a problem of designing an array with narrowest main beam for a given side-lobe level, or said differently, the lowest sidelobe level for a given main beam-width. In a uniform array, the sidelobe level decreases as we move away from the main beam. For a given sidelobe level, the sidelobe amplitudes need not decrease with the side-lobe number. In fact, the best situation would be equal amplitudes of all the side-lobes. The array synthesis problem then is to find location of the nulls which would make all sidelobes amplitudes equal. Chebyshev polynomial is precisely the polynomial which has this property.

Chebyshev polynomials are defined as

$$T_m(x) = \cos(m\cos^{-1}x) \qquad -1 < x < 1$$
$$T_m(x) = \cosh(m\cosh^{-1}x) \qquad |x| > 1 \tag{9.91}$$

where m is the degree of the polynomial. The general shape of the polynomial is shown in Fig. 9.35.

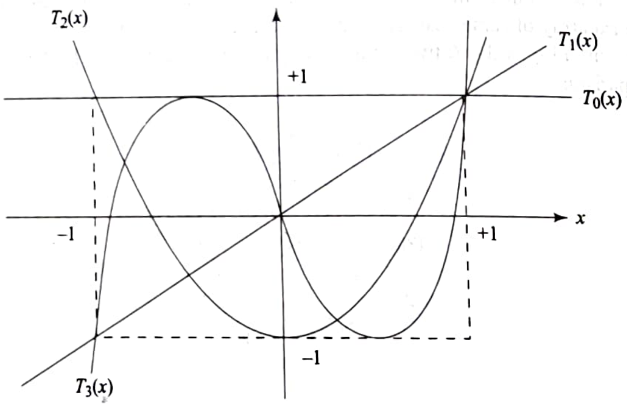


Fig. 9.35 *Plots of Chebyshev polynomials.*

The polynomials have the following recurrence relation.

$$T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x) \tag{9.92}$$

By inspection we get

$$T_0(x) = 1 \tag{9.93}$$

and

$$T_1(x) = x \tag{9.94}$$

The recurrence relation can be used to get higher order polynomials. For example,

$$T_2(x) = 2xT_1(x) - T_0(x) = 2x^2 - 1$$
$$T_3(x) = 2xT_2(x) - T_1(x) = 2(2x^2-1)x - x = 4x^3 - 3x \tag{9.95}$$

and so on.

Chebyshev polynomials have following interesting properties:

(i) The polynomials of all orders pass through point (1, 1).
(ii) For $-1 \le x \le +1$, the polynomials have values between +1 and −1 only.
(iii) All roots of the polynomial lie between −1 and +1, and all maxima and minima have values +1 and −1 respectively.
(iv) For $x > 1$ the polynomial increases monotonically.
(v) The polynomials also satisfy relationship

$$T_m(\cos\gamma) = \cos m\gamma \quad \text{if } \gamma \text{ is real} \tag{9.96}$$

and

$$T_m(\cosh\gamma) = \cosh m\gamma \quad \text{if } \gamma \text{ is complex} \tag{9.97}$$

(vi) From (v) we note that the zeros of the polynomials are given by

$$T_m(x) = T_m(\cos\gamma) = \cos m\gamma = 0 \tag{9.98}$$

$$\Rightarrow \qquad \gamma = \frac{1}{m}(2n+1)\frac{\pi}{2} \qquad n = 0, 1, 2, (m-1) \tag{9.99}$$

The corresponding values of x are

$$x = \cos\gamma = \cos\left\{\frac{1}{m}(2n+1)\frac{\pi}{2}\right\}, \qquad n = 0, 1, \cdots (m-1) \tag{9.100}$$

Let us now make use of the Chebyshev polynomials for the array synthesis. First observe that $T_1(x) = \cos\gamma$, $T_2(x) = \cos 2\gamma$, $T_3(x) = \cos 3\gamma$, and so on.

Now consider

$$x = a + b\cos\psi \tag{9.101}$$

Then, we get

$$T_1(x) = x = a + b\cos\psi \tag{9.102}$$

$$T_2(x) = 2x^2 - 1 = 2(a + b\cos\psi)^2 - 1$$
$$= (2a^2 + b^2 - 1) + 4ab\cos\psi + b^2\cos 2\psi \tag{9.103}$$

$$T_3(x) = 4x^3 - 3x = 4(a + b\cos\psi)^3 - 3(a + b\cos\psi)$$
$$= (4a^3 + 6ab^2 - 3a) + (12a^2b + 3b^3 - 3b)\cos\psi$$
$$+ 6ab^2\cos 2\psi + b^3\cos 3\psi \tag{9.104}$$

and so on.

Now, we make an important observation. $T_1(x)$ is equivalent to a Fourier series upto $\cos\psi$, $T_2(x)$ is equivalent to a Fourier series upto $\cos 2\psi$ and so on. In general, $T_m(x)$ is then equivalent to a Fourier series upto $\cos m\psi$. Since, a radiation pattern can be expressed as a Fourier series in ψ, the Chebyshev polynomials can be employed for synthesizing an array pattern.

Let us consider a broadside symmetrical array of $(2N+1)$ elements. The array factor is given by (For braodside array all currents are in phase, i.e. all currents are real)

$$AF = a_0 + 2\sum_{k=1}^{N} a_k \cos k\psi \tag{9.105}$$

Equation (9.105) can be equated to $T_N(a + b\cos\psi)$. The constants a and b are chosen in such a way that the visible range of ψ corresponds to the values of x in $T_N(x)$ that range from $x = -1$ to some value $x = x_1$ where $x_1 > 1$. Since, x_1 is greater than 1, $T_N(x_1)$ is always greater than 1. Now the value of ψ corresponding to $x = x_1$ gives the main beam of the array pattern whereas the side lobes correspond to those values x which lie between -1 and $+1$ and where $T_N(x)$ is ± 1 (see Fig. 9.36)

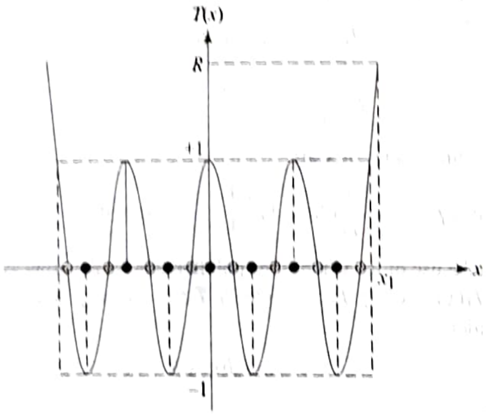


Fig. 9.36

The values of x corresponding to dots in Fig. 9.36 indicate locations of the sidelobes, and that corresponding to circles indicate the locations of the nulls of the array pattern. It is interesting to note that for this array, the sidelobe amplitudes are always 1 irrespective of the array size. The amplitude of the main beam, however, can be increased arbitrarily by increasing the value of x_1. In other words, for a given number of elements in the array, the sidelobe level relative to the main beam can be arbitrarily reduced by choosing higher value of x_1.

Assuming the interelement spacing $d \le \lambda/2$, the visible range of ψ is from βd to $-\beta d$. For broadside array, as we have seen above, the direction of main beam corresponds to $\psi = 0$. Point x_1 should then correspond to $\psi = 0$ giving

$$a + b = x_1 \tag{9.106}$$

Now, since, the extreme point on the radiation pattern corresponds to $x = -1$, $\psi = \beta d$ and $-\beta d$ should correspond to $x = -1$. That is,

$$a + b\cos(\beta d) = a + b\cos(-\beta d) = -1 \tag{9.107}$$

Solving Eqns (9.106) and (9.107) we get

$$a = \frac{1 + x_1\cos\beta d}{1 - \cos\beta d} \tag{9.108}$$

$$b = \frac{1 + x_1}{1 - \cos\beta d} \tag{9.109}$$

Let the ratio of the main beam to sidelobe be denoted by R. Then, we drive $T_N(x_1) = R$, since sidelobe amplitude is always 1. Hence,

$$T_N(x_1) = T_N(\cosh\gamma) = \cosh N\gamma = R \tag{9.110}$$

$$\Rightarrow \quad \gamma = \frac{1}{N}\cosh^{-1} R \tag{9.111}$$

$$\Rightarrow \quad x_1 = \cosh\gamma = \cosh\left\{\frac{1}{N}\cosh^{-1} R\right\} \tag{9.112}$$

Substituting x_1 in Eqns (9.108) and (9.109) we get a and b which after substituting in $T_N(a + b\cos\psi)$ and expanding the polynomial in Fourier series gives the array coefficients. The procedure will be clear after a few examples are worked out.

We can also find the array coefficients by making use of nulls of the array as has been done earlier.

EXAMPLE 9.18 Design a broadside array of 5 elements which has main beam to side lobe ratio of 10 and smallest possible beam-width. The inter-element spacing is $\lambda/2$.

Solution:

This is a problem of designing a Chebyshev array. The number of elements $2N+1 = 5 \Rightarrow N = 2$. The radiation pattern therefore is given by $T_2(x)$. Main beam to side-lobe ratio, $R = 10$.

$$\Rightarrow \quad x_1 = \cosh\left(\frac{1}{2}\cosh^{-1} 10\right) = 2.3452$$

Since, $d = \lambda/2$, $\quad \beta d = \frac{2\pi}{\lambda}\cdot\frac{\lambda}{2} = \pi \quad \Rightarrow \quad \cos\beta d = -1$

Using Eqns (9.108) and (9.109) we get

$$\Rightarrow a = \frac{1 - x_1}{2} = -0.6726$$

$$b = \frac{1 + x_1}{2} = 1.6726$$

The array coefficeints can be obtained in two ways:

FIRST APPROACH

From Eqn (9.103) we have

$$T_2(x) = a_0 + 2a_1 \cos\psi + 2a_2 \cos 2\psi$$
$$= (2a^2 + b^2 - 1) + 4ab\cos\psi + b^2\cos 2\psi$$

The array coefficient are

$$a_0 = 2a^2 + b^2 - 1 = 2.7024$$
$$a_1 = 2ab = -2.25$$
$$a_2 = b^2/2 = 1.3988$$

The 5 elements of the array are excited with currents 1.39, −2.25, 2.70, −2.25, 1.39 i.e. the currents are in the ratio 1 : −1.61 : 1.94 : −1.61 : 1

SECOND APPROACH

The half range of $\psi = \beta d - \pi$ corresponds to $x = -1$ to $x = x_1 = 2.3452$.

$$\psi = \frac{\beta d}{1 + x_1}(x_1 - x)$$

The location of the nulls of $T_2(x)$ are given by

$$x = \cos\left\{\frac{1}{2}(2n+1)\frac{\pi}{2}\right\} \qquad n = 0, 1$$

The nulls are at

$$x = x_{n1} = \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$$

and $$x = x_{n2} = \cos\frac{3\pi}{4} = -\frac{1}{\sqrt{2}}$$

Now from Eqn (9.101) we have $\cos\psi = (x - a)/b$

Substituting for a and b we get $\cos\psi = (x - (-0.6726))/1.6726$.

For the two nulls x_{n1} and x_{n2} we get

$\cos\psi_1 = 0.8249$ and $\cos\psi_2 = -0.021$.

Now, for a broadside array there is symmetry of nulls around $\psi = 0$. The other two nulls, therefore, are

$$\psi_1' = -\psi_1$$
$$\psi_2' = -\psi_2$$

The radiation pattern is

$$AF = (Z - e^{j\psi_1})(Z - e^{j\psi_1'})(Z - e^{j\psi_2})(Z - e^{j\psi_2'})$$
$$= (Z^2 - 2\cos\psi_1 Z + 1)(Z^2 - 2\cos\psi_2 Z + 1)$$

$$= 1 - 2(\cos\psi_1 + \cos\psi_2)Z + (2 + 4\cos\psi_1\cos\psi_2)Z^2 - 2(\cos\psi_1 + \cos\psi_2)Z^3 + Z^4$$
$$= 1 - 1.61Z + 1.94Z^2 - 1.61Z^3 + Z^4$$

The array coefficients are 1.0, −1.61, 1.94, −1.61, 1.0 which are same as that obtained from the first approach.

9.4.3 Superdirective Arrays

In case of the Chebyshev array we see that by choosing proper value of x_1 the ratio of the main beam to side-lobe can be arbitrarily increased. Also, since, the Chebyshev polynomials $T_m(x)$ increase steeply for $x > 1$, with increasing m, we can get higher main beam to side lobe ratio with fixed x_1 but higher order of the polynomial, i.e. higher number of elements in the array. It should be noted that increasing number of elements does not mean larger array. It means that more densely packed elements in the given length of the array.

One can also note that, with increasing the order of the polynomial, point x_1 comes closer to $x = 1$ for a given main beam to sidelobe ratio, and the location of first null also moves closer to $x = 1$. That is to say that by increasing the order of the polynomial the beam-width can be arbitrarily reduced and consequently the directivity of the array can be arbitrarily increased. By this technique one can then obtain much higher directivity compared to that of a uniform array of the same length. This array then is called the 'Superdirective Array'.

Let us investigate a super directive array of fixed length say, $L = \lambda/4$ having 7 elements, and main beam to sidelobe ratio of 20 dB, ($R = 100$). The interelement spacing in this case is $d = L/6 = \lambda/24$. The Chebyshev array for these parameters have $a = -73.01544219$, $b = 74.55587192$ and $x_1 = 1.54042973$. The relative values of the currents in the elements are

$$a_0 = -3.99200886 \times 10^6$$
$$a_1 = 3.006383214 \times 10^6$$
$$a_2 = -1.2175861 \times 10^6$$
$$a_3 = 2.072123161 \times 10^6$$

For the broadside array, the maximum radiation is proportional to the algebraic sum of all the currents since they all add in phase in the broadside direction. The maximum radiation, therefore, is proportional to

$$a_0 + 2a_1 + 2a_2 + 2a_3 = 10.0002 \qquad (9.113)$$

That is to say that radiation equivalent to 10 A of current is received from the array when the actual currents in the elements are of the order of million Amperes. The large antenna currents give high ohmic losses in the antenna elements. The superdirective array consequently has a very poor efficiency. The important thing here to note is, the directivity of a superdirective array is very large but its gain is very low.

The superdirective array although is a very interesting concept, its practical utility is very low due to its unacceptably low efficiency. In real life, therefore, the superdirective arrays hardly find any application.

9.5 SUMMARY

The antenna array analysis and synthesis is a very vast subject and it is employed in a variety of fields like, radio astronomy, radars, mobile communication, seismology, sonography, etc. The antenna array decouples the radiation properties from the terminal characteristics of the antenna. Complex radiation patterns can be realized with help of arrays. Apart from the arrays mentioned above, there are other types of arrays like correlation arrays, adaptive arrays, multiple beam arrays which find active applications today in science and technology. In this chapter only the basic principles involved in array design have been highlighted. For more in-depth studies of arrays, readers may refer to other advanced material on the subject.

Review Questions

9.1 What is an antenna array?

9.2 Why are antenna arrays preferred over other types of antennas?

9.3 For a N-element array, how many degrees of freedom does one have?

9.4 Why is the spacing between array elements not made arbitrarily small?

9.5 What are linear, planar and volume arrays?

9.6 In a two element array which parameter affects the direction of main beam?

9.7 What is the effect of inter-element spacing on the array pattern?

9.8 Why cannot the direction of maximum radiation and a null both be controlled independently in a two element array?

9.9 What is a uniform array?

9.10 What is progressive phase shift in a uniform array?

9.11 Why are current amplitudes made equal in a uniform array?

9.12 What is a primary radiation pattern?

9.13 What is the relation between progressive phase shift and direction of main beam?

9.14 Why does the direction of main beam not depend upon the number of elements in a uniform array?

9.15 What is the maximum number of nulls in an N-element array?

9.16 What is HPBW?

9.17 What is BWFN?

9.18 Why is HPBW a more meaningful parameter than BWFN?

9.19 What is visible range of ψ?

9.20 What is the amplitude of the first sidelobe in a uniform array?

9.21 Why does the HPBW increase as the main beam direction tilts from broadside to end-fire direction?

9.22 For a given large array, when is the HPBW maximum, when the array is broadside or when the array is end-fire?

9.23 For a given large array when is the directivity maximum? In the broadside configuration or in the end fire configuration? Why?

9.24 Name applications of a highly directive array.

9.25 What is a phased array antenna?

9.26 What is a grating beam? Why should it be avoided?

9.27 What is the condition under which there are no grating beams for any scanning of the main beam?

9.28 What is array synthesis?

9.29 What is a Binomial array? What is its speciality?

9.30 What is speciality of the Chebyshev array?

9.31 What is Superdirective array?

9.32 Why are Superdirective arrays not used in practice?

Problems

9.1 A two-element array is excited with currents $1\angle 0$ and $1.5\angle 45^\circ$. The spacing between the elements is 1.5λ. Find the directions of maximum and minimum radiation.

9.2 Two parallel Hertz dipoles separated by a distance of 2λ, are excited with equal currents. Find the directions of maximum radiation and the nulls.

9.3 Two isotropic radiators are located at coordinates $(-\lambda, 0, -\lambda/2)$ and $(\lambda, 0, \lambda/2)$. Find the radiation patterns in xy, yz and zx planes.

9.4 A Hertz dipole of 10 cm length carrying 1 A current at 100 MHz is horizontally located at a height of 1 m above a perfect ground plane. Find the electric and magnetic field at a distance of 10 km in the direction of 45° from the horizon in the vertical plane containing the axis of the dipole. Also find the directions in which there is no radiation field.

9.5 A vertical 1.5λ long dipole is located at a height of 2λ above a perfect ground plane. Find the directions of maximum radiation and the nulls.

9.6 Two isotropic radiators are seperated by a distance of $\frac{\lambda}{2}$ and are excited with equal currents. Find the half power beam width and directivity of the array when (i) the array is broadside (ii) the array is end-fire.

9.7 For a two element array of isotropic radiators, the maximum radiation is at an angle of 45° from the axis of the array and the minimum radiation is along the broadside direction. The ratio of maximum and minimum electric field is 2 : 1. Find the complex current ratio for the two elements and seperation between the elements.

9.8 Find the radiation pattern of an array of two parallel halfwave dipoles carrying currents I_1 and I_2 respectively in the plane containing the two dipoles when (i) $I_1 = I_2$ (ii) $I_1 = -I_2$ (iii) $I_1 = I_2\angle 60°0$

9.9 A 5-element uniform array with inter-element spacing of 0.6λ is excited to have main beam along 30° from the broadside direction. Find the progressive phase shift and the direction of the nulls for the array.

9.10 A uniform array has main beam at an angle of 45° from the array axis. If the interelement spacing is $\frac{\lambda}{2}$ find the number of elements in the array to get BWFN of 30°. Find the directions of all the nulls.

9.11 In a 3-element uniform array the inter element spacing is 0.75λ. Find the phase gradients on the array so as to get nulls in the broadside and end-fire directions. Find the direction of the main beam in each case.

9.12 For the two-element array of the Hertz dipoles shown in Fig. 9.37, find the directions of the nulls.

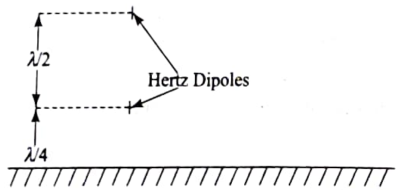


Fig. 9.37

9.13 A uniform 4-element aray has the progressive phase shift of 45° and interelemnent spacing of 0.4λ. Find the directions of the nulls and directions of the sidelobes. How many sidelobes are visbile in the radiaiton pattern.

9.14 A 10 element uniform array has interelement spacing of 0.7λ. Find the approximate HPBW and the directivity of the array when it is excited as (i) broadside array (ii) end-fire array

9.15 The element of a 6-element aray of $\frac{\lambda}{2}$ long parallel dipoles are separated by a $\frac{\lambda}{3}$ distance. Draw approximately the radiation patterns in the plane (i) containing the dipoles (ii) passing through the array axis and perpendicular to the dipole (iii) perpendicular to the array axis and passing through the center of the array.

9.16 A uniform scanning array of colinear Hertz dipoles is to be designed to scan the main beam over ±60° around the braodside direction. What is the maximum permissible inter-element spacing so that the grating lobe is less than the main beam. For that spacing find the directivity of the array when it is broadside if the number of elements in the array is 15.

9.17 Two uniform end-fire arrays have same BWFN of 60° but different number of elements. Will the two arrays have same nulls and directivities? If the number of elements in the two arrays are 5 and 8 respectively, find the directivities and the nulls for the two arrays.

9.18 A uniform array has inter element spacing of $\frac{5\lambda}{8}$. If the main beam of the array is at an angle of 30° form the array axis, find the visible range of ψ.

9.19 A horizontal 4-element uniform array of isotropic antennas with $d = 0.4\lambda$ and $\delta =$ 45° is located at a height of 0.3λ above a perfect ground plane. Find the directions of the nulls, main beam and the sidelobes.

9.20 A 3-element array of length 0.7λ is to be excited such that radiation pattern has nulls along $\phi = 50°$ and 120°, where ϕ is the angle measured from the array axis. Find the complex currents of the array.

9.21 A 6-element array has to have equi-spaced nullls over $\phi = 0$ to 180°. If the inter element spacing is $\frac{\lambda}{3}$ find the excitation of the array elements.

9.22 The visible range ψ of a 5-element uniform array is from −45° to +120°. Find the directions of the main beam and the nulls.

9.23 In Problem 22 if the nulls are equispaced in the visible range of ψ, find the directions of the nulls and the currents on the array elements.

9.24 A broadside 7-element binomial array has nulls at ±30° from the broadside. Find the inter-element spacing and the currents on the array. What is the HPBW and the directivity of the array.

9.25 An end-fire 7-element binomial array has nulls at ±30° from the end-fire direction. Find the inter-element spacing and the currents on the array. What is the HPBW and the directivity of the array.

9.26 Design a 5-element array so as to get uniform radiation intensity over angular sectors $30° \leq \theta \leq 90°$ and $120° \leq \theta \leq 180°$ where θ is the angle measured from the array axis.

9.27 A linear array of 7-element is to be designed for satellite application. The array is parallel to the earth's surface. Find the current distribution on the array so that the radiation field strength is uniform on the ground along a line parallel to the array axis. Assume the height of the satellite is 800 km and the width of the beam on the ground is 1600 km.

9.28 Design a linear array to have a radiation pattern which is maximum in the broadside direction and linearly tapers off to zero at 30° on either side of the broadside direction. The maximum rms deviation between the realised and the expected radiation pattern is 5 %. Use computer if needed.

9.29 A 5-element uniform sidelobe array with broadside beam is to be designed to get the sidelobe amplitude 10% of the main beam. Find the array currents if the inter element spacing is $\lambda/2$. Also find the HPBW of the array.

CHAPTER 10

Propagation of Radio Waves

In the previous chapters, we studied the propagation of electromagnetic waves in an unbound as well as bound media at the conceptual level. Coming to the practical aspects of this we will now investigate the propagation of radiowaves from a transmitter to a receiver in a communication link. The word 'radio' is used here in a broad sense to indicate any frequency at which normally the long distance communication takes place. Any frequency beyond few hundred kHz, therefore, can be called a radio frequency. Depending upon the span of the communication link and the requirement of the bandwidth, different frequency bands are used for different applications. For example, local AM radio broadcasting stations use frequencies of few hundred kHz whereas the local FM radio stations use frequencies of few hundred MHz. The satellite communication uses frequencies in the range of few GHz to few tens of GHz. A typical radio frequency bands used in short and long distance wireless communication are shown in Fig 10.1.

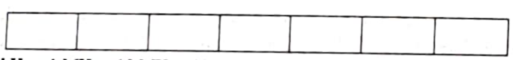


Fig. 10.1 *Radio frequency spectrum.*

Generally, lower frequency bands are used for audio broadcasting services whereas higher frequency bands are used for video transmission and more professional services.

The propagation of different frequency bands get affected in different ways in the presence of earth and its gaseous atmosphere. This is due to the fact that the medium properties like dielectric constant, conductivity, etc. of earth and its atmosphere vary as a function of frequency.

In this chapter, we investigate various ways in which an electromagnetic energy can be sent from one point to another on the surface of the earth.

Figure 10.2 shows the earth's surface and the atmospheric layers above the earth surface. The atmospheric layer which is very close to the earth's surface is called the troposphere. This layer extends for few tens of km above the earth's surface and primarily consists of neutral gas molecules. The layer called the 'ionosphere', extends from 100 km to 1000 km above the earth's surface and consists of free electrons and ionized molecules. Both the layers affect the propagation of electromagnetic waves. In addition, the earth's surface acts like a large reflecting boundary, hence, affects the propagation of EM waves.

Figure 10.2 shows various paths by which an EM wave can reach from one point to another. When an EM wave propagates along the earth's surface the wave is called the 'Ground Wave'. Generally, the ground wave signals are divided into two, the 'space wave' and the 'surface wave'. The space wave propagates in the troposphere and is made of two waves namely the direct wave and the ground-reflected wave. As the name suggests, the direct wave travels by the direct path from the transmitter to the receiver whereas, the ground reflected wave arrives at the receiver after reflecting from the earth's surface.

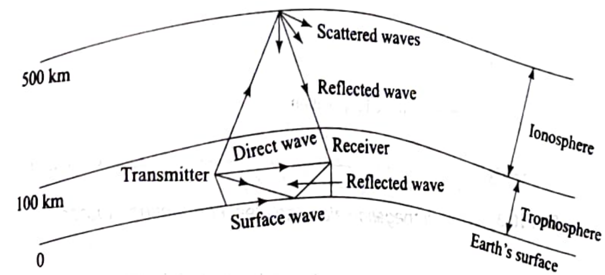


Fig. 10.2 *Modes of propagation of electromagnetic waves.*

The surface wave is guided by the earth's surface similar to the way an EM wave is guided by a conducting surface. The surface wave therefore is possible for those frequencies for which the earth's surface acts like a good conductor. As seen in Chapter 4, a medium behaves like conductor at low frequencies. The surface wave propagation therefore is more effective at low frequencies of few hundred kHz. It should, however, be pointed out here that the conductivity of the earth is rather low and consequently the earth's surface behaves like a lossy conductor giving high loss to the surface wave. An EM wave can be transmitted only over few tens of km using this mode.

The wave so called the 'Sky Wave' is the wave which is returned to the earth by the ionosphere. The sky wave is a combination of systematically reflected wave from the layered structure of the ionosphere and the scattered waves due to random fluctuations in the ionosphere. The ionospheric propagation is effective for short wave communication. An EM wave can be transmitted over few hundred km using this propagation. It will be shown later however, that this mode is not very suitable for short distance communication. We, therefore, find that the ground wave and sky wave provide complementary coverage on the ground. For short distances, the ground wave is more effective whereas, for long distances the sky wave is more effective.

10.1 SURFACE WAVE PROPAGATION

As mentioned above, surface wave is a wave which is guided along the surface of the earth as if the earth is a good conductor. Let us also assume that the conducting surface is flat as the effect of the curvature of the earth is negligibly small over short distances. As we have seen in Chapter 4, there is no penitration of electromagnetic wave inside an ideal conductor and the wave exists only in the dielectric medium above the conducting surface. Since, there is only one conducting boundary, the wave propagation can again be transverse electromagnetic with the electric field normal and the magnetic field tangential to the conducting boundary as shown in Fig. 10.3.

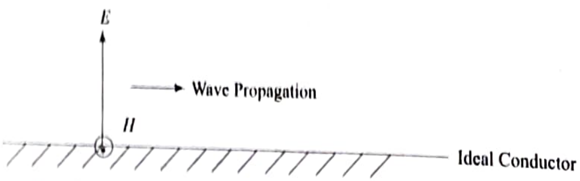


Fig. 10.3 *Wave propagation along an ideal conducting surface.*

Note that the boundary conditions demand tangential component of the electric field and the normal component of the magnetic field to be zero at the conducting boundary. The normal component of E and tangential component of H give Poynting vector along the conducting surface and therefore the transverse electromagnetic wave propagates along the conducting surface. It should be emphasized at this point that, since, a horizontal electric field is not supported by the ideal conducting surface, a surface wave is essentially a vertically polarized wave. For medium wave radio broadcasting, which is through the surface waves, we must have an antenna which can generate a vertically polarized wave. The wave travels without any loss since we have assumed the conductor to be ideal.

Let us now consider the real earth which is not an ideal conductor. However, one thing is clear that, we have to make sure that the earth must behave like a good conductor for surface wave propagation to take place. The frequency of wave ω should, therefore, be much smaller than σ/ϵ, σ is the conductivity of the earth and ϵ is its permittivity. The imperfect conducting nature of the earth affects the wave propagation in two ways:

(i) It develops a tangential component of electric field at the surface.

(ii) There is ohmic loss due to finite conductivity which attenuates the wave.

10.1.1 Wave Tilt

Let a TEM wave with vertical polarization be launched along the earth surface. Let the electric and magnetic field for the wave be E_v and H as shown in Fig 10.4.

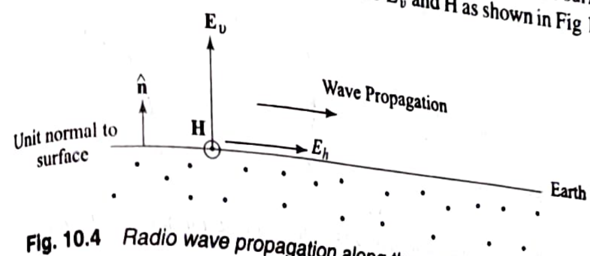


Fig. 10.4 *Radio wave propagation along the surface of the earth.*

If the dielectric constant of the earth is ϵ_r, the surface impedance of the earth is

$$Z_s = \sqrt{\frac{j\omega\mu_0}{\sigma + j\omega\epsilon_0\epsilon_r}} \tag{10.1}$$

Now, since the surface impedance is non zero, there has to be an electric field tangential to the earth's surface and in the direction of the surface current. The surface current $\hat{\mathbf{n}} \times \mathbf{H}$ flows in the direction of wave propagation and consequently the tangential electric field is also in the direction of the wave propagation. Let us denote this tangential field by E_h. We now have following relations.

$$\frac{E_v}{H} = \eta_0 \tag{10.2}$$

(assuming medium above the earth is free-space)

and

$$\frac{E_h}{H} = Z_s \tag{10.3}$$

The ratio of E_v and E_h therefore is

$$\frac{E_v}{E_h} = \frac{\eta_0}{Z_s} = \frac{\sqrt{\frac{\mu_0}{\epsilon_0}}}{\sqrt{j\omega\mu_0/(\sigma + j\omega\epsilon_0\epsilon_r)}} \tag{10.4}$$

$$= \sqrt{\frac{\sigma + j\omega\epsilon_0\epsilon_r}{j\omega\epsilon_0}} \tag{10.5}$$

Two things should be noted from Eqn (10.5). First, since E_h is non-zero, the wave is no more a transverse electromagnetic wave. Second, since E_v and E_h are not in phase and in general since the $|E_v| \neq |E_h|$, the resultant electric field is elliptically polarized. Note however that, the ellipse of polarization is not in the transverse plane but is in the longitudinal plane. Figure 10.5 shows the ellipse of polarisation for different values of $\sigma/\omega\epsilon$ for the earth.

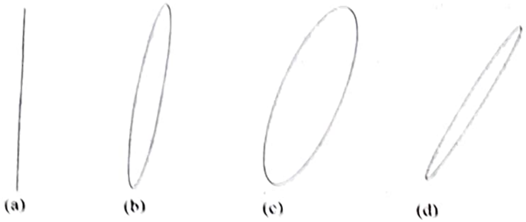


Fig. 10.5 *Wave tilt for various values of loss-tangent for the earth.*

For ideal conductor since $\sigma = \infty$ the wave is vertically linearly polarized. As σ decreases, the axial ratio of the ellipse decreases (ellipse becomes fatter) and the major axis of the ellipse tilts. This phenomenon is called the 'wave tilt'. As will be seen later, wave tilt is a direct measure of the loss.

EXAMPLE 10.1 A vertically polarized wave having power density of 100 W/m^2 is launched as a surface wave along a smooth ground having dielectric constant 20 and conductivity 10^{-2} ℧/m. The frequency of the wave is 2 MHz. Find the wave tilt and the power loss per unit area of the ground.

Solution:

Power density of the wave

$$P = \frac{|E_v|^2}{\eta_0}$$

$$\Rightarrow \quad |E_v| = \sqrt{P\eta_0} = \sqrt{100 \times 120\pi}$$

$$= 194.1626 \text{ V/m}$$

The surface impedance

$$Z_s = \sqrt{\frac{j\omega\mu_0}{\sigma + j\omega\epsilon_0\epsilon_r}} = \sqrt{\frac{j2\pi \times 2 \times 10^6 \times 4\pi \times 10^{-7}}{10^{-2} + j2 \times 2\pi \times 10^6 \times 20 \times \epsilon_0}}$$

$$= 30.62 + 24.57j$$

The horizontal electric field induced is

$$|E_h| = \frac{|Z_s|\,|E_v|}{\eta_0}$$

The wave tilt is approximately

$$\tan^{-1}\left(\frac{|E_h|}{|E_v|}\right) = \tan^{-1}\left(\frac{39.2624}{120\pi}\right) = 0.1038 \text{ rad}$$

The power loss per unit area is

$$P_L = R_s\,|H|^2 = R_s\left|\frac{E_h}{Z_s}\right|^2 = R_s\left|\frac{E_v}{\eta_0}\right|^2$$

(Factor 1/2 is not there because E is already the rms value of the electric field.)

$$P_L = Re\{Z_s\}\left|\frac{E_v}{\eta_0}\right|^2 = 30.62 \times \frac{(10)^2}{120\pi} = 8.122 \text{ W}$$

10.1.2 Attenuation of the Surface Wave

Following the analysis in Chapter 6, the attenuation constant of the wave is given as

$$\alpha = \frac{\text{Power loss per unit distance}}{2 \times \text{Power carried by the system}} \tag{10.6}$$

If we now assume that the wave guided by the earth's surface is a uniform plane wave, the power carried by the wave will be infinite (as the wave front is of infinite height) and we will not be able to calculate the attenuation constant. Let us, therefore, consider a more practical case, a wave launched by a vertical small monopole antenna as shown in Fig. 10.6.

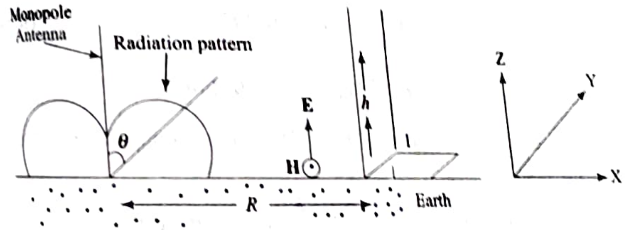


Fig. 10.6 *Loss due to imperfect conducting surface.*

The radiation pattern of a small monopole antenna mounted above the perfect conducting ground is same as that of the Hertz dipole and therefore the electric field at any height is given as (Refer Fig. 10.7)

$$E = \frac{K}{r}\sin\theta \tag{10.7}$$

where K is a constant and r is the radial distance from the antenna.

Let us now consider a strip of unit width in the yz-plane at a distance R along the x-axis. The electric find in this strip is given as

$$E = \frac{K}{r}\sin\theta = \frac{K}{R}\sin^2\theta \tag{10.8}$$

$$H = \frac{E}{\eta_0} = \frac{K}{R\eta_0}\sin^2\theta \tag{10.9}$$

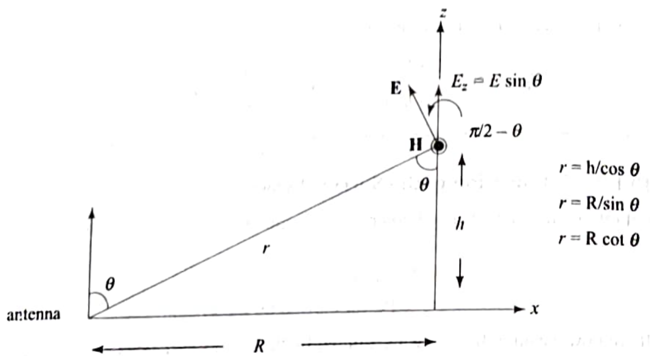


Fig. 10.7

Now the power carried by the wave perpendicular to the strip (along x-direction) is

$$W = \int_0^\infty \frac{1}{2}Re\{E_z \times H^*\}dh \tag{10.10}$$

$$= \int_0^\infty \frac{1}{2}Re\{E\sin\theta \times H^*\}dh \tag{10.11}$$

Substituting for E and H from Eqns (10.8) and (10.9) and noting that $h = R\cot\theta$, we get (Refer Fig. 10.7)

$$W = \int_0^{\pi/2} \frac{1}{2}\cdot\frac{K}{R}\sin^3\theta\cdot\frac{K}{d\eta_0}\sin^2\theta\cdot R^2\theta d\theta \tag{10.12}$$

$$= \frac{K^2}{2d\eta_0}\int_0^{\pi/2}\sin^3\theta d\theta \tag{10.13}$$

$$W = \frac{K^2}{2R\eta_0}\cdot\frac{2}{3} = \frac{K^2}{3R\eta_0} \tag{10.14}$$

Now, we note that the power loss per unit distance for a unit strip width is nothing but the power loss per unit area on the earth's surface, which is given as

$$P_L = \frac{1}{2}R_S|J_s|^2 = \frac{1}{2}R_s|H|^2_{\text{surface}} \tag{10.15}$$

where R_s is the surface resistance of the earth ($R_s = Re(Z_s)$). At the earth's surface, $\theta = \pi/2$ and we have from Eqns (10.3) and (10.4),

$$|H|_{\text{surface}} = \frac{|E_h|_{\text{surface}}}{Z_s} = \frac{|E|_{\text{surface}}}{\eta_0} = \frac{K}{R\eta_0} \tag{10.16}$$

Substituting Eqn (10.16) in (10.15) we get,

$$P_L = \frac{1}{2}R_s\frac{K^2}{R^2\eta_0^2} \tag{10.17}$$

The attenuation constant, therefore, is given as

$$\alpha = \frac{1}{2}\frac{P_L}{W} = \frac{3}{4}\frac{R_s}{R\eta_0} \tag{10.18}$$

10.2 SPACE WAVE

The space wave propagation normally takes place at high frequencies between two points which are elevated above the earth surface. This mode of transmission therefore is employed for TV transmission or communication between two microwave towers. The electromagnetic wave travels almost like a uniform plane wave (in a straight line) in this case. Both types of polarizations, horizontal and vertical can be used in the space wave propagation.

Let us first investigate the space wave propagation assuming the flat earth. Let the transmitter and receiver be mounted on two towers of heights h_1 and h_2 respectively, separated by a distance 'd' (Fig. 10.8).

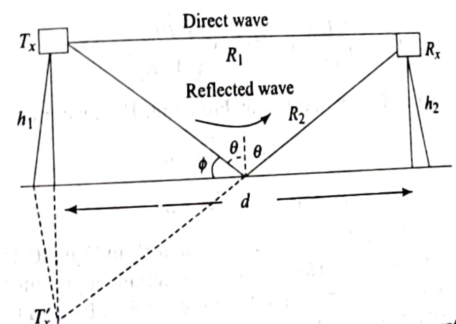


Fig. 10.8 *Space wave propagation between two antennas mounted on towers.*

Let us assume that the transmitting and receiving antennas are isotropic. Now the wave from transmitter reaches the receiver by two paths, the direct path and the ground reflected path. In reality, the distance between the towers is much greater than the heights of the towers (i.e. $d >> h_1, h_2$). Consequently, the angle of incidence θ is very close to $\pi/2$, and angle ϕ is very close to 0. Since the wavelength is much shorter now, the reflection seems to be a mirror like reflection.

When seen from the receiver, the reflected wave seems to be originating from the image of the transmitter T_x'. In fact, the reflection point O is obtained by joining the image of the transmitter and receiver by a straight line. The reflection coefficient at the earth's surface depends upon the polarization of the wave.

Assuming that the earth has conductivity σ and dielectric constant ϵ_r. The intrinsic impedance of earth is

$$\eta_e = \sqrt{\frac{j\omega\mu_0}{\sigma + j\omega\epsilon_0\epsilon_r}} \tag{10.19}$$

The reflection coefficient for the two polarizations can be written (from Chapter 5) as

$$\Gamma_H = \frac{\sqrt{\sigma + j\omega\epsilon_0\epsilon_r}\cos\theta_i - \sqrt{j\omega\epsilon_0}\cos\theta_t}{\sqrt{\sigma + j\omega\epsilon_0\epsilon_r}\cos\theta_i + \sqrt{j\omega\epsilon_0}\cos\theta_t} \tag{10.20}$$

$$\Gamma_V = \frac{\sqrt{j\omega\epsilon_0}\cos\theta_i - \sqrt{\sigma + j\omega\epsilon_0\epsilon_r}\cos\theta_t}{\sqrt{j\omega\epsilon_0}\cos\theta_i + \sqrt{\sigma + j\omega\epsilon_0\epsilon_r}\cos\theta_t} \tag{10.21}$$

Let us now assume that the transmitting antenna radiates P watts of power isotropically. The radiation electric field E at a distance R then is

$$E = \sqrt{\frac{P\eta_0}{4\pi R^2}}\, e^{-j\beta R} \tag{10.22}$$

where $\beta = 2\pi/\lambda$.

The electric field at the receiver, which is sum of the electric fields of the direct and the reflected waves can be written as

$$E_r = \sqrt{\frac{P\eta_0}{4\pi R_1^2}}\, e^{-j\beta R_1} + \Gamma\sqrt{\frac{P\eta_0}{4\pi R_2^2}}\, e^{-j\beta R_2} \tag{10.23}$$

Γ is the reflection coefficient of the earth. From Fig. 10.8, we find

$$R_1 = \sqrt{(h_1 - h_2)^2 + d^2} \tag{10.24}$$

$$R_2 = \sqrt{(h_1 + h_2)^2 + d^2} \tag{10.25}$$

In general, we can obtain E_r by substituting R_1 and R_2 in Eqn (10.23). However, in practice, where $d >> h_1$ and h_2, we can make certain approximations. Firstly, since in this case $\theta \approx \pi/2$, the reflection coeficient Γ_H and Γ_V are almost equal to -1. That is to say that the reflection coefficient $\Gamma \sim -1$ irrespective of the polarization of the wave. Secondly, we can approximate R_1 and R_2 as

$$R_1 = d\left\{1 + \frac{(h_1 - h_2)^2}{d^2}\right\}^{1/2} \approx d\left\{1 + \frac{(h_1 - h_2)^2}{2d^2}\right\}$$

$$\text{and} \quad R_2 = d\left\{1 + \frac{(h_1 - h_2)^2}{d^2}\right\}^{1/2} \approx d\left\{1 + \frac{(h_1 - h_2)^2}{2d^2}\right\} \tag{10.26}$$

Rewriting expression for the received electric field we get

$$E_r = \sqrt{\frac{P\eta_0}{4\pi}}\left\{\frac{e^{-j\beta R_1}}{R_1} + \Gamma\frac{e^{-j\beta R_2}}{R_2}\right\}$$

$$= \sqrt{\frac{P\eta_0}{4\pi}}\frac{e^{-j\beta R_1}}{R_1}\left\{1 + \frac{\Gamma R_1}{R_2}e^{-j\beta(R_2 - R_1)}\right\} \tag{10.27}$$

Now for $d >> h_1$ and h_2, we have $\Gamma \approx -1$ and $R_2 - R_1 \approx 2\frac{h_1h_2}{d}$. Using arguments similar to that used in Chapter 8, we make $R_1 \approx R_2 \approx d$ in the amplitude term whereas in the phase term we substitute actual value of $R_2 - R_1$. We, therefore, have

$$E_r = \sqrt{\frac{P\eta_0}{4\pi}}\frac{e^{j\beta d}}{d}\left(1 - e^{-j2\beta h_1h_2/d}\right)$$

$$\Rightarrow \quad |E_r| = \sqrt{\frac{P\eta_0}{4\pi}}\frac{2}{d}\left|\sin\left(\frac{2\pi h_1h_2}{\lambda d}\right)\right| \tag{10.28}$$

The electric field strength $|E_r|$ of the space wave has decaying sinusoidal variation as shown in Fig. 10.9.

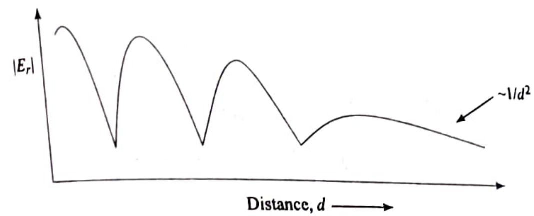


Fig. 10.9 *Variation of the space wave field strength as a function of distance.*

For short distance the field strength varies from maximum to null. However, as the distance becomes large the field strength monotonically decreases as a function of distance. For large d, we can approximate

$$\sin\left(\frac{2h_1h_2}{\lambda d}\right) \approx \frac{2\pi h_1h_2}{\lambda d} \tag{10.29}$$

$$\text{giving} \qquad |E_r| = \frac{4\pi h_1h_2}{\lambda d^2}\sqrt{\frac{P\eta_0}{4\pi}} \tag{10.30}$$

At large distance, therefore, the field strength varies as $1/d^2$ and consequently decreases much rapidly compared to the $1/d$ variation of a spherical wave. It is clear that the sharp decrease in the field strength is due to the reflected wave which cancels the electric field of the direct wave. In presence of the earth, therefore, the space wave also provides a mode of transmission only over short distances. Of course this conclusion has been reached for an isotropic transmitting antenna. By using directional antennas the effect of the reflected wave can be reduced giving higher transmission range.

EXAMPLE 10.2 Two isotropic microwave antennas are located on two towers of height 100 m and 60 m respectively. The distance between the towers is 2 km. Assuming flat earth, find the distance of the reflection point for the ground reflected wave. If the frequency of the signal is 4 GHz, find the phase difference between the direct and the reflected waves at the receiver due to propagational delay.

Solution:

Referring to Fig. 10.10, we get

$$\frac{x}{h_1} = \frac{d-x}{h_2}$$

$$x = \frac{h_1 d}{h_1 + h_2} = \frac{0.1 \times 2}{0.1 + 0.06}$$

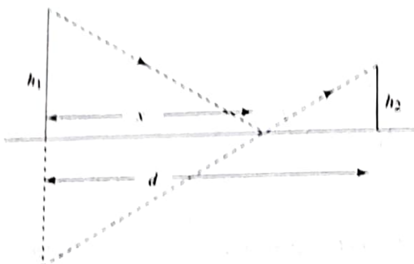


Fig. 10.10 *Reflection of space wave from the earth's surface.*

Distance of reflection point $x = 1.25$ km

The wavelength of the wave =

$$\frac{3 \times 10^8}{4 \times 10^9} = 0.075 \text{ m}$$

Phase constant

$$\beta = \frac{2\pi}{\lambda} = \frac{2\pi}{0.075} = \frac{80\pi}{3} \text{ rad/m}$$

The phase difference between the direct and the reflected wave is

$$\phi = \beta(R_2 - R_1) = \frac{80\pi}{3}\left\{\sqrt{(h_1+h_2)^2 + d^2} - \sqrt{(h_1-h_2)^2 + d^2}\right\}$$

Substituting $h_1 = 100$ m $h_2 = 60$ m and $d = 2000$ m we get, $\phi = 501.3$ rad

EXAMPLE 10.3 In Example 10.2, if the transmitting isotropic antenna radiates 10 W power in a horizontally polarized wave, and if the ground has $\epsilon_r = 10$ and $\sigma = 10^{-3}$ ℧/m, find the net electric field at the receiving antenna. If the receiving antenna has an effective aperture of 0.2 m^2, find the power received by the receiving antenna.

Solution:

The reflection coefficient from Eqn (10.20) for the horizontally polarized wave is $\Gamma_H = 0.5195$

The received electric field

$$E_r = \sqrt{\frac{10 \times 120\pi}{4\pi}}\left\{\frac{e^{-j\beta R_1}}{R_1} + \Gamma_H \frac{e^{-j\beta R_2}}{R_2}\right\}$$

Taking $R_1 \approx R_2 \approx 2000$ m in the denominator

$$E_r = \frac{\sqrt{300}}{2000}\{e^{-j\beta R_1} + 0.5195 e^{-j\beta R_2}\}$$

$$|E_r| = 0.0121 \text{ V/m}$$

The effective power density at the receiving antenna is

$$P_r = \frac{|E_r|^2}{\eta_0} = 3.88 \times 10^{-7} \text{ W/m}^2$$

Power received by the antenna

$$W = P_r \cdot A_{eff} = 3.88 \times 10^{-7} \times 0.2 = 7.77 \times 10^{-8} \text{ W}$$

10.2.1 Space Wave with Directional Antennas

Let us assume that the radiation patterns of the transmitting and receiving antennas be given as $f_T(\theta)$ and $f_R(\theta)$ as shown in Fig. 10.11, where θ is the elevation angle. Let the gains of the two antennas be denoted by G_T and G_R respectively. For $d >> h_1$ and h_2, $\theta_1 \approx 0$, and maximum radiation is along $\theta = 0$.

The received electric field can now be written as

$$E_r = \sqrt{\frac{P\eta_0 G_T}{4\pi R_1^2}} e^{-j\beta R_1} + \Gamma\sqrt{\frac{P\eta_0 G_T}{4\pi R_2^2}} f_T(\theta_T) f_R(\theta_R) e^{-j\beta R_2} \qquad (10.31)$$

Again, applying same approximations as above, we get

$$|E_r| = \sqrt{\frac{P\eta_0 G_T}{4\pi}}\left|\frac{e^{-j\beta d}}{d}\right|\left\{1 - f_T(\theta_T) f_R(\theta_R) e^{-j2\beta\frac{h_1 h_2}{d}}\right\}\Bigg| \qquad (10.32)$$

By making antennas highly directional, we get $f_T(\theta_T) f_R(\theta_R) << 1$ and consequently the second term in the bracket is $<< 1$. The field strength then

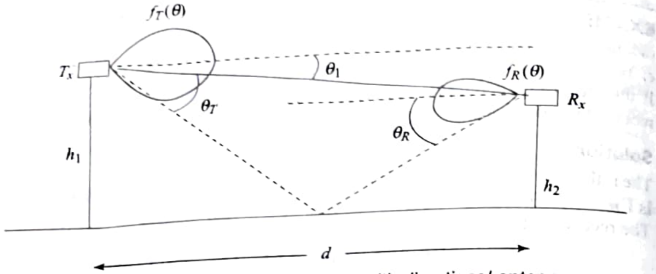


Fig. 10.11 *Space wave transmission with directional antennas.*

is

$$|E_r| = \sqrt{\frac{P\eta_0 G_T}{4\pi}}\frac{1}{d} \tag{10.33}$$

The strength of the received electric field then varies as, $1/d$ and not as $1/d^2$ as in the case of the isotropic antennas. The range of communication is much longer in this case. Highly directional antennas are used in microwave links not only for better line of sight focusing of the electromagnetic wave but also for suppressing the ground reflected waves.

EXAMPLE 10.4 Two microwave towers are seperated by a distance of 20 km. The transmitter and receiver towers have heights 150 m and 10 m respectively. If the transmitted power is 100 W find the received power when (i) Both antennas are isotropic (ii) Both antennas look into each other and have directivity of 30 dB. Wavelength of the signal is 3 cm.

Solution:

When both antennas are isotropic, both direct and ground reflected waves interfere, whereas for highly directional antenna there is only direct wave.

(i) Since $d >> h_1$ and h_2, for isotropic antennas (from Eqn (10.30)) we have

$$|E_r| = \frac{4h_1h_2}{\lambda d^2}\sqrt{\frac{P\eta_0}{4\pi}}$$

$$= \frac{4 \times 150 \times 100}{0.03 \times (20 \times \omega^3)^2}\sqrt{\frac{100 \times 120\pi}{4\pi}}$$

$$= 0.2739$$

Effective received power density

$$P_r = \frac{|E_r|^2}{\eta_0} = \frac{(4h_1h_2)^2 p}{4\pi\lambda^2 d^4}$$

$$= 0.1989 \text{ mW/m}^2$$

For isoptropic receiving antenna, the directivity is 1, giving effective aperture

$$A_e = \frac{\lambda^2}{4\pi} = \frac{(0.03)^2}{4\pi}$$

Power received

$$W = P_r \cdot A_e = 1.4248 \times 10^{-8} \text{ W}$$

(ii) For directional antennas the directivity D is 30 dB = 1000. The effective aperture therefore is

$$A_e = \frac{1000\lambda^2}{4\pi} = \frac{1000 \times (0.03)^2}{4\pi}$$

$$= 7.16 \times 10^{-2} \text{ m}^2$$

Since, there is no reflected wave,

$$|E_r| = \sqrt{\frac{P\eta_0}{4\pi}} \cdot \frac{D}{d}$$

Power density at the receiving antenna is

$$P_r = \frac{|E_r|^2}{\eta_0} = \frac{P}{4\pi}\left(\frac{D}{d}\right)^2$$

$$P_r = \frac{100}{4\pi}\left(\frac{1000}{20 \times 10^3}\right)^2 \text{ W/m}^2$$

Power received

$$W = P_r \cdot A_e = 1.424 \text{ mW}$$

10.2.2 Space Wave Over Spherical Earth

The above analysis of the space wave assumes a flat earth. However, in reality the earth is spherical and the curvature of the earth cannot be neglected beyond a distance of few tens of km. The curvature of the earth affects the space wave in two ways:

(i) For a given height of transmitter/receiver tower, there is an upper limit on the distance over which there is line of sight between the transmitter and the receiver.

(ii) The angle between the direct wave and the reflected waves reduces making it difficult to supress the reflected wave using directional antennas.

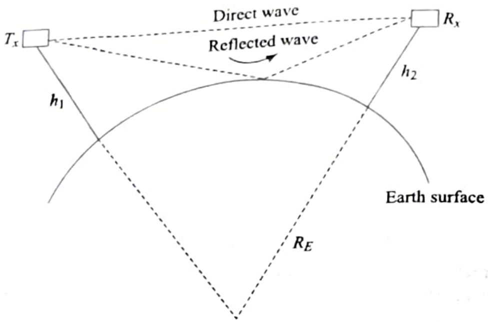


Fig. 10.12 *Space wave propagation over the curved earth surface.*

Figure 10.12 shows the space wave propagation above the curved earth surface. Maximum distance between the transmitter and receiver is obtained when both the direct and the reflected waves become tangential to the earth's surface. Of course in this situation the two waves completely cancel each other making net field at the receiver zero. However, just from line of sight consideration we may say that this is the maximum distance allowed between the transmitter and the receiver. Figure 10.13 shows the geometry for the maximum transmission range; R_E is the radius of the earth.

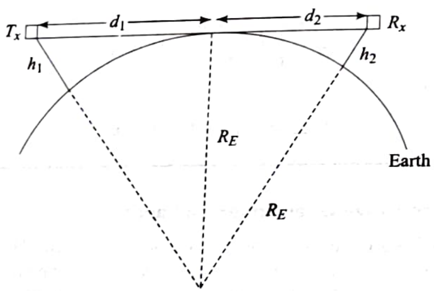


Fig. 10.13 *Maximum repeater spacing for the space wave.*

From Fig. 10.13 we have

$$d_1 = \sqrt{(R_E + h_1)^2 - R_E^2} = \sqrt{2R_E h_1 + h_1^2} \tag{10.34}$$

$$d_2 = \sqrt{(R_E + h_2)^2 - R_E^2} = \sqrt{2R_E h_2 + h_2^2} \tag{10.35}$$

In practice h_1 and $h_2 << R_E$ giving

$$d_1 \simeq \sqrt{2R_E h_1} \tag{10.36}$$

$$d_2 \simeq \sqrt{2R_E h_2} \tag{10.37}$$

The maximum transmission range is

$$d_{max} = d_1 + d_2 = \sqrt{2R_E h_1} + \sqrt{2R_E h_2} \tag{10.38}$$

It is, therefore, clear that by increasing the height of transmitter and/or receiver the transmission range can be increased. If the distance over which the signals are to be sent is larger than d_{max}, multiple towers are to be used.

10.3 IONOSPHERIC PROPAGATION

As we rise above the earth's surface, we encounter a partially ionized gaseous layer called the 'Ionosphere.' The inonosphere is a medium which consists of free electrons and ions, and neutral gas molecules. The electron density gradually increases above the earth's surface, reaches a maximum at a height of about 500 km, and then gradually decreases above that. Of course this is a very simple description of the inonosphere, what is called the Chapman's model of the inonosphere. In reality the inosphere is in the form of layers within which there are electron clouds with random density variation. In the sky wave communication, the radio signals sent towards the ionosphere get systematically deflected by the layers of the ionosphere and get randomly scattered by the electron clouds. The ionospheric propagation therefore has to be investigated for both, the layered structure and the random structure. In the following sections we investigate propagation of electromagnetic waves in a layered ionized medium in detail. Towards the end of the chapter, however, we shall give a brief explanation on the propagation of electromagnetic waves in a random medium.

10.3.1 Structure of Ionosphere

The ionosphere is essentially created due to ionization of the earth's atmospheric gases by solar radiation. The atmospheric density is highest near the earth's surface and it monotonically decreases with the altitude. The solar radiation on the contrary is strong at higher altitudes and its intensity decreases due to atmospheric absorption as it moves towards the earth's surface. At higher altitude therefore the radiation intensity is high but there are not sufficient molecules to get ionized. Consequently the electron density is low at higher altitudes of 1000 kms or so. As the altitudes decreases, the molecular density increases and consequently the electron density increases. For further decrease in the altitude, the molecular density continues to increase but the intensity of radiation decreases due to absorption in the upper layers. Consequently the electron density again decreases and practically becomes zero below an altitude of 80–90 kms. The electron density is maximum at about 500–600 kms above the earth's surface.

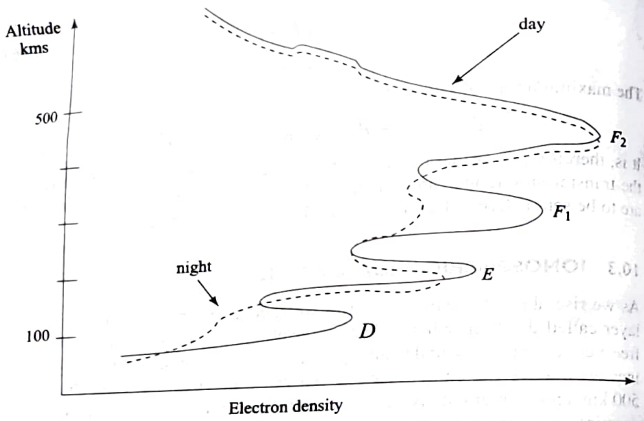


Fig. 10.14 *Electron density variation as a function of height in the ionosphere.*

The single peak profile of the electron density variation is expected if the earth's atmosphere consists of homogeneous mixture of different gases. In practice, however, different gases have different density profiles and they have different ionization energies. As a result, instead of single peak electron density profile, we have a density profile with multiple peaks as shown in Fig 10.14. The regions around the peaks are identified by layers called D, E, F layers. Typical heights of the D, E and F layers are 80 km, 200 km and 500 km respectively. Generally, E and F layers affect the radio wave propagation. The average electron density in these layers has large diurnal variation. This is expected since the angle of solar radiation which is the casue of ionization, changes with the time of the day. At noon, when the solar radiation directly falls on the earth's atmosphere all the layers are fully ionized. The F-layer consists of two layers F_1 and F_2. At night when the lower atmosphere is under the shadow of the earth, the D, E and F_1 layer tend to disappear. F_2 layer however remains more or less unaffected. In the night although, there is no systematic E layer, there are electron density clouds which are formed due to random recombination of electrons and ions after the sun set. The E-layer, therefore, is called the 'sporadic E layer'.

It is, therefore, clear that during day time the sky wave communicaton can be established using either of the D, E, F_1, F_2 layers, whereas, in the night only F_2 layer is available for the sky wave communication.

10.3.2 Dielectric Constant of the Ionosphere

In the previous sections we have seen dielectrics which have only bound changes. The ionosphere on the contrary has free mobile charges in the form of ions and electrons. Ionosphere in fact is a collection of electrons, ions and neutral molecules. Such media are generally referred to as plasmas. The ionospheric propagation essentially is propagation of electromagnetic waves in a plasma.

A plasma is a rather complex medium especially when it is magnetized. Here, we first investigate the dielectric properties of a plasma in the absence of magnetic field. The magnetized plasmas are dealt with subsequently.

When an electric field is impressed upon the plasma, the mobile charges start moving constituting the conduction current. If we assume that the electric field is oscillating at a high frequency, the conduction current is primarily due to the motion of electrons since the ions are much heavier and therefore are less mobile. Let us, therefore, assume that as far as the high frequency fields are concerned, the ions are stationary. While moving, the electrons colloid with the neutral molecules and the ions, and there is momentum transfer from electrons to the molecules and ions. This causes decceleration of the electrons which are accelerated by the applied electric field.

Let there be sinusoidal electric field of angular frequency ω and amplitude E_0 applied to an electron. We, therefore, have

$$\mathbf{E} = \mathbf{E}_0 e^{j\omega t} \tag{10.39}$$

Due to this field, the velocity of the electron varies sinusoidally giving

$$\mathbf{v} = \mathbf{v}_0 e^{j\omega t} \tag{10.40}$$

where $\mathbf{v}_0$ is the peak velocity of the electron. If the collision frequency of electrons with ions and molecules is ν, and the mass of electron is m, the rate of decrease of momentum due to collision would be $m\nu\mathbf{v}$. This quantity is also known as frictional drag.

The net force on the electron is the force due to the electric field minus the frictional drag. The equation of motion of an electron can therefore, be written as

$$m\frac{d\mathbf{v}}{dt} = -e\mathbf{E} - m\nu\mathbf{v} \tag{10.41}$$

$$\Rightarrow \quad jm\omega\mathbf{v} = -e\mathbf{E} - m\nu\mathbf{v} \tag{10.42}$$

$$\Rightarrow \quad \mathbf{v} = \frac{-e\mathbf{E}}{jm\omega + m\nu} \tag{10.43}$$

Here, e is the charge of an electron.

The conduction current density due to motion of electrons is

$$\mathbf{J} = -N e \mathbf{v} \tag{10.44}$$

where N is the electron density in the plasma. Substituting for $\mathbf{J}$ in the Maxwell's equation we get

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E} + \mathbf{J} = j\omega\epsilon_0\mathbf{E} - Ne\mathbf{v} \tag{10.45}$$

$$= j\omega\epsilon_0\mathbf{E} + \frac{Ne^2\mathbf{E}}{j\omega + m\nu} \tag{10.46}$$

Here, it is assumed that $\epsilon \approx \epsilon_0$ since the gas density is low.
Rationalizing the second term on RHS we get

$$\begin{aligned}
\nabla \times \mathbf{H} &= j\omega\epsilon_0\mathbf{E} + \frac{Ne^2}{m}\frac{(\nu - j\omega)}{\nu^2 + \omega^2}\mathbf{E} \\
&= j\omega\epsilon_0\left\{1 - \frac{Ne^2/m\epsilon_0}{\omega^2 + \nu^2}\right\}\mathbf{E} + \frac{Ne^2\nu}{m(\omega^2 + \nu^2)}\mathbf{E} \\
&= j\omega\epsilon_0\left[\left\{1 - \frac{Ne^2/m\epsilon_0}{\omega^2 + \nu^2}\right\} - j\frac{Ne^2\nu/m\epsilon_0}{\omega(\omega^2 + \nu^2)}\right]\mathbf{E} \\
&= j\omega\epsilon_0\left[\epsilon_r' - j\epsilon_r''\right]\mathbf{E} \\
&\equiv j\omega\epsilon_0\epsilon_r'\mathbf{E} + \sigma\mathbf{E}
\end{aligned} \tag{10.47}$$

where ϵ_r' and σ are effective dielectric constant and effective conductivity respectively of the plasma.
An ionized medium therefore has complex dielectric constant in general.

As discussed in Chapter 4, the imaginary part of the dielectric constant represents loss in the medium. In this case the loss is due to collision of electrons with the ions and molecules. The plasma becomes low loss when ω is much higher compared to the collision frequency ν.

The quantity $\sqrt{Ne^2/m\epsilon_0}$ is called the plasma angular frequency and is denoted by ω_p ($= 2\pi f_p$). f_p is the natural frequency of oscillation of a plasma slab. Substituting for electronic charge $e = 1.59 \times 10^{-19}$ Coulombs, electronic mass $m = 9 \times 10^{-31}$ kg and $\epsilon_0 = 8.86 \times 10^{-12}$F/m, we get

$$f_p = \sqrt{81N} \tag{10.48}$$

where N is the electron density per cubic meter and f_p is in Hertz. The dielectric constant of a plasma in the frequency range $\omega >> \nu$ therefore is

$$\epsilon_r' = 1 - \frac{\omega_p^2}{\omega^2} = 1 - \frac{f_p^2}{f^2} = 1 - \frac{81N}{f^2} \tag{10.49}$$

One can make a very important observation here, that the dielectric constant of a plasma is always less than unity. The dielectric constant may even become negative. This medium is then very different than other dielectrics where the dielectric constant is always greater than unity. Since, the refractive index of a medium is the square root of the dielectric constant ($n = \sqrt{\epsilon_r'}$), a plasma has a refractive index less than unity, or even imaginary if $f < f_p$. A typical variation of the refractive index in the ionosphere then is as follows:

At low heights when N is small, $f >> f_p$, and the refractive index is almost unity. As the height increases, N increases and consequently the refractive index decreases gradually (and it can become imaginary if $f < f_p$ at certain height) upto F_2 layer, where the electron density is maximum. Beyond F_2 layer, the refractive index again increases gradually to become almost unity at large heights.

As a function of height, the collision frequency ν also varies and consequently the conductivity of the ionosphere varies. Since, the collision frequency is more governed by the neutral molecular density, the collision frequency decreases as a function of height.

We have seen in Chapter 4, that a dielectric medium becomes lossy if its conductivity is non-zero. Higher the conductivity, more is the loss in the medium. An electromagnetic wave at certain frequency ω when propagates through the ionosphere, suffers maximum loss at that height where the conductivity is maximum. Making the derivative of the conductivity with respect to ν, zero, we get

$$\frac{\partial}{\partial\nu}\left\{\frac{Ne^2\nu/m}{(\omega^2 + \nu^2)}\right\} = 0$$

$$\Rightarrow \qquad \omega = \nu \tag{10.50}$$

(Note that although N and ν both are functions of atmospheric density, they are taken independent since N depends on the ionization index whereas as ν depends upon the neutral molecular density.)

An electromagnetic wave undergoes maximum loss in a layer for which $\omega = \nu$. Generally, for radio wave frequencies the wave has maximum absorption in the E-layer.

10.3.3 Group Velocity of a Wave in Plasma

To make the analysis simple, let us assume that the collisional frequency is negligibly small compared to f and therefore the plasma has negligible loss. The dielectric constant of the plasma is then

$$\epsilon_r = 1 - \frac{\omega_p^2}{\omega^2} = n^2 \tag{10.51}$$

A uniform plane wave has a phase constant in the plasma as

$$\beta = \beta_0\sqrt{\epsilon_r} = \frac{\omega}{c}\sqrt{\epsilon_r} = \frac{\omega}{c}\left(1 - \frac{\omega_p^2}{\omega^2}\right)^{1/2} \tag{10.52}$$

The group velocity of the wave then is

$$v_g = \frac{\partial\omega}{\partial\beta} = \frac{c^2\beta}{\omega} = c\sqrt{1 - \frac{\omega_p^2}{\omega^2}} = c\sqrt{1 - \frac{81N}{f^2}} = cn \tag{10.53}$$

In general, plasma is a dispersive medium since the group velocity varies with frequency. Equation (10.52) is similar to that for a waveguide, with ω_c replaced by ω_p. The plasma frequency ω_p of an ionized medium is identical to the cut-off frequency of a waveguide. In a waveguide, when $\omega >> \omega_c$, the group velocity is almost same as the velocity of light and when $\omega \to \omega_c$, the group velocity approaches zero. Similarly, inside a plasma, the group velocity $v_g \approx c$ when $\omega >> \omega_p$ and $v_g \to 0$ when $\omega \to \omega_p$. Below the plasma frequency (i.e. $\omega < \omega_p$), there is no propagation of wave similar to that in a waveguide where the wave propagation ceases below the cut-off.

EXAMPLE 10.5 In the E-layer of the ionosphere, the electron density is 10^{10} electrons/m^3 and the collision frequency is 10^6 per sec. Find the complex dielectric constant of the layer at 2 MHz.

Solution:

Plasma frequency of the layer is

$$f_p = \sqrt{81N} = 9\sqrt{10^{10}} = 9 \times 10^5 \text{ Hz}$$
$$= 0.9 \text{ MHz}.$$
$$\omega_p = 2\pi \times f_p = 1.8\pi \text{ M rad/sec}$$

The complex dielectric constant from Eqn. (10.47) is

$$\epsilon_r = \left\{1 - \frac{\omega_p^2}{\omega^2 + \nu^2}\right\} - j\frac{\omega_p^2 \nu}{\omega(\omega^2 + \nu^2)}$$

$$\omega = 2\pi \times 2 \times 10^6 = 4\pi \text{ M rad/sec}$$

$$\epsilon = \left\{(1 - \frac{(1.8\pi \times 10^6)^2}{(4\pi \times 10^6)^2 + (10^6)^2}\right\} - j\frac{(1.8\pi \times 10^6)^2 \times 10^6}{4\pi \times 10^6((4\pi \times 10^6)^2) + (10^6)^2}$$

$$= \left\{1 - \frac{(1.8\pi)^2}{(4\pi)^2 + 1}\right\} - j\frac{(1.8\pi)^2}{4\pi(16\pi^2 + 1)}$$

$$= 0.7988 - 0.016j$$

10.4 WAVE PROPAGATION THROUGH THE IONOSPHERE

As seen above, the refractive index of the ionosphere is a strong function of frequency. Consequently, the radio waves of different frequencies get affected differently by the ionosphere. At very low frequencies, where the wave length is comparable to the thickness of the layers, the ionosphere behaves like a dielectric slab of average layer properties and the wave is reflected back as it would be reflected from a dielectric slab. This case is not of great practical relevance as these low frequency waves cannot be generated very efficiently in practice. Also, these

waves can be transmitted more efficiently through the ground wave propagation. On the other extreme, when $f >> f_p$, the refractive index of the ionosphere is almost unity and the wave propagation is practically unaffected.

The most interesting and useful case is where the frequency of the wave f, becomes comparable to the plasma frequency f_p at some layer of the ionosphere. In this case, the refractive index significantly deviates from unity and consequently affects the radio wave propagation maximally. This case can be analyzed by tracing the path of the wave in the medium having gradually varying refractive index. The ionosphere is visualized as a layered medium with different refractive indices in different layers as shown in Fig 10.15.

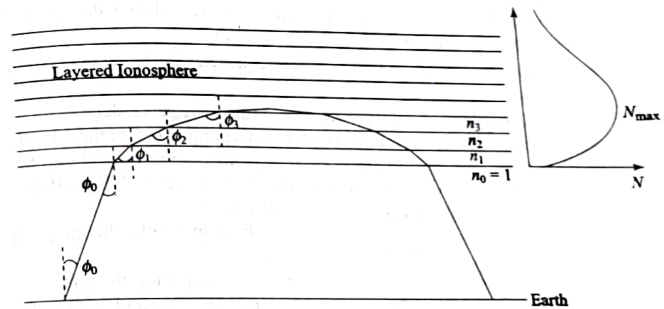


Fig. 10.15 *Propagation of a uniform plane wave in the layered ionosphere.*

Let an electromagnetic wave be launched towards the ionosphere at an angle ϕ_0 from the vertical direction. Applying Snell's law at each interface of the layers we get,

$$n_1 \sin\phi_1 = \sin\phi_0 \quad (10.54)$$

$$n_2 \sin\phi_2 = n_1 \sin\phi_1 = \sin\phi_0 \quad (10.55)$$

and so on.

In general, at any layer, then we have

$$\sin\phi_0 = n \sin\phi \quad (10.56)$$

where n is the refractive index of the layer and ϕ is the direction of the wave in that layer. It is clear that as we go deeper in the ionosphere, n decreases and consequently ϕ increases since $\sin\phi_0$ is constant for any given launching angle. At some height ϕ becomes $\pi/2$ and the wave travels along the layer. Any downward perturbation to the wave direction at this height pushes the wave downwards and the wave is reflected to the earth. At the reflecting layer we have

$$\sin\phi_0 == n\sin(\pi/2) = n = \sqrt{1 - f_p^2/f^2} = \sqrt{1 - \frac{81N}{f^2}} \qquad (10.57)$$

where N is the electron density in the reflecting layer. Simplifying Eqn (10.57) we get

$$f = f_p \sec\phi_0 = \sqrt{81N}\sec\phi_0 \qquad (10.58)$$

Following important observations can be made from Eqn (10.58).

(a) For a given launching angle ϕ_0, higher frequencies get reflected from higher layers (Here, we are assuming that the reflecting layer is below the F_2 layer where the density is maximum in the ionosphere)

(b) By increasing the launching angle, higher frequencies can be reflected from a given layer.

(c) For vertical launching ($\phi_0 = 0$ i.e. $\sec\phi_0 = 1$) the wave gets reflected from a layer for which $f = f_p$. One may wonder at this point, whether the relation (10.58) is applicable in this case as ϕ_0 and ϕ will be zero through out the propagation! In other words one can ask a question, 'is the result Eqn (10.57) valid for $\phi_0 = 0$?' This can be answered in two ways:
   (i) since, the Eqn (10.57) is valid for any arbitarily small value of ϕ_0, it will also be valid in the limit for $\phi_0 = 0$.
   (ii) when frequency becomes equal to the plasma frequency, the refractive index goes to zero and the wave propagation ceases beyond that height. The entire wave energy then gets reflected from that layer.

(d) For vertical incidence the highest frequency which can be reflected by a layer is called the '*critical frequency*' and is given by

$$f_{cr} = \sqrt{81N_{max}} \qquad (10.59)$$

(e) Signals having frequencies higher than f_{cr} can be reflected by a layer provided they are launched at an angle such that (from Eqn (10.58))

$$f = f_{cr}\sec\phi_0 \qquad (10.60)$$

This is called the '*Secant law*' for ionospheric reflection.

10.4.1 Skip Distance

If the launching angle ϕ_0 is less than $\cos^{-1}(f_{cr}/f)$ the wave penetrates the ionosphere and never returns to the earth. A typical picture of wave paths for some $f > f_{cr}$ are shown in Fig 10.16.

From Fig. 10.16 we note that the waves launched at larger ϕ_0 reach far away distances on the ground. As we decrease the launching angle ϕ_0, the point of return comes closer to the transmitter. However, below certain value of ϕ_0, the wave is refracted and there is no reflection of the wave. For a given frequency

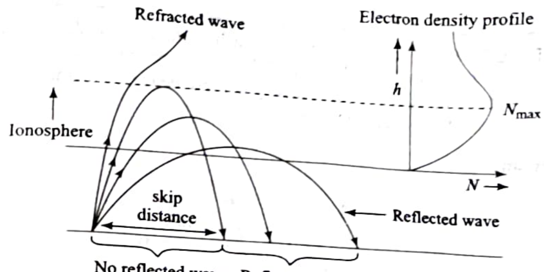


Fig. 10.16 *Reflection of radio waves as a function of launching angle. Concept of skip distance.*

$f > f_{cr}$ we, therefore, do not have reflected wave within certain distance. The minimum distance from the transmitter over which the reflected wave can reach is called the 'skip distance'.

The ionospheric propagation is essentially suited for long distance communication. The ionospheric mode of transmission, therefore, is complementary to the ground and space mode of transmission. The spatial locations reached by ionospheric waves cannot be reached by the ground/space waves and vice versa. Of course, one may argue at this point that by choosing $f < f_{cr}$, we can eliminate the skip distance. This will, however, reduce the frequency range for communication significantly. Also, as will become clear later, the low frequencies undergo higher collisional loss in the ionosphere.

EXAMPLE 10.6 The F_1 layer of the ionosphere has a height of 400 km and electron density of $10^{11}/\text{m}^3$. What is the maximum possible elevation angle for a 10 MHz radiation to get reflected from the layer? What is the skip distance?

Solution:

Plasma Frequency $f_p = 9\sqrt{10^{11}} = 2.846$ MHz

$$\text{Now } f_p \sec\phi = f$$

$$\Rightarrow \quad \phi = \sec^{-1}\left(\frac{f}{f_p}\right) = \sec^{-1}\left(\frac{10}{2.846}\right)$$

$$= 73.47^\circ$$

Maximum elevation angle is $\theta = 90^\circ - \phi = 16.535^\circ$

The skip distance

$d_{skip} = 2h\tan\phi = 2 \times 400 \times \tan(73.47^\circ) = 2695$ km

EXAMPLE 10.7 A circular antenna dish pointed vertically upwards has the effective beam width of 90° at 12 MHz. An ionospheric layer has electron density of $10^{12}/m^3$ and is at 500 km height. Find the region on the ground over which there will be reflected wave.

Solution:

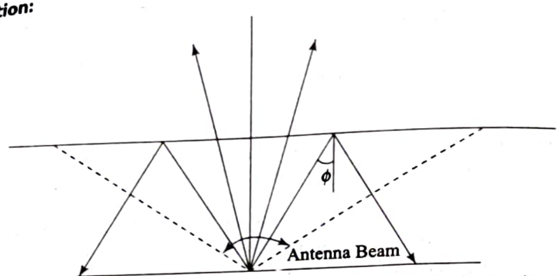


Fig. 10.17 *Reflected and refracted waves from the ionosphere.*

Plasma frequency of the layer

$$f_p = 9\sqrt{10^{12}} = 9 \text{ MHz}$$

Minimum angle of incidence ϕ is given by

$$f_p \sec \phi_{min} = f$$

$$\phi_{min} = \sec^{-1}\left(\frac{12}{9}\right) = 41.4°$$

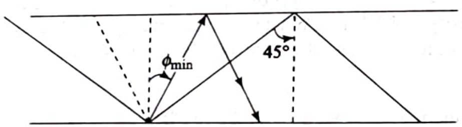


Fig. 10.18 *Maximum and minimum angle of the reflected waves.*

Maximum angle of incidence is decided by the beam of the antenna i.e. $\phi_{max} =$ 45°

Radiation going within $\phi_{min} = \pm 41.4°$ is not reflected. We then have a circular region around the antenna where there is no reflection. The radiation is reflected within a circular annular region formed by radii

$$r_{min} = 2h \tan \phi_{min} = 2 \times 500 \tan(41.4°) = 881.9 \text{ km}$$

$$r_{max} = 2h \tan 45° = 1000 \text{ km}$$

10.4.2 Virtual Height

Let us now consider the trajectory of an electromagnetic pulse inside the ionosphere (Fig. 10.19). The pulse is reflected from the ionosphere. Let the pulse be launched at an angle ϕ_0 and let at some height the pulse be travelling at an angle ϕ with respect to the vertical direction with group velocity v_g. The horizontal and vertical components of the group velocity of the pulse are

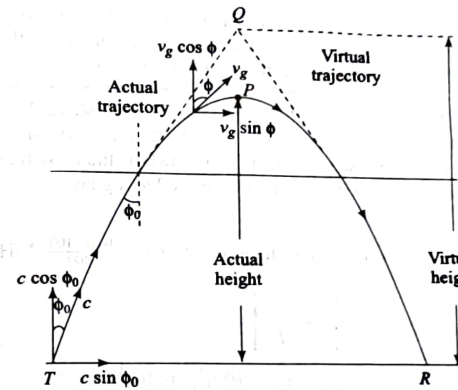


Fig. 10.19 *Actual and virtual paths of a wave in the ionosphere.*

$$v_{gH} = v_g \sin \phi = cn \sin \phi \quad (10.61)$$

$$v_{gV} = v_g \cos \phi = cn \cos \phi \quad (10.62)$$

Substituting for $n \sin \phi = \sin \phi_0$ from Eqn (10.56) we have

$$v_{gH} = c \sin \phi_0 = \text{constant for given } \phi_0 \quad (10.63)$$

$$v_{gV} = c\sqrt{n^2 - \sin^2 \phi_0} \quad (10.64)$$

The pulse, therefore, has constant horizontal velocity throughout the propagation. The vertical velocity decreases till the pulse reaches the reflecting layer where the vertical velocity becomes zero. The vertical velocity then increases downward and the wave returns to the ground. The pulse has uniform velocity in the horizontal direction and decelaration in the upward direction. The pulse, therefore, has a projectile motion inside the ionosphere.

Let us now imagine that there is no ionosphere and the wave is launched at an angle ϕ_0. This pulse also has $v_{gH} = c \sin \phi_0$. In this case however the pulse will reach point Q travelling in a straight line in a time same as that taken by the pulse to travel in the ionosphere along the actual trajectory upto point P.

Let us say that the travel time of the pulse from transmitter T or receiver R is t_d. Then the height of point Q above the earth's surface can be estimated to be

$$H_{vir} = \frac{1}{2} c \cos\phi_0 \cdot t_d \tag{10.65}$$

We call this height 'virtual height' because the pulse virtually appears to be reflected from this height.

The pulse delay measurement, therefore, provides height of a hypothetical layer from which there is mirror like reflection of the pulse. In practice this technique is used to measure the heights of the ionopsheric layers. Indeed, the heights estimated by this technique are the virtual heights which are always greater than the true heights of the reflecting layers. In fact, the layered nature of the ionosphere was established using this technique at multiple frequencies.

EXAMPLE 10.8 In a simplified model of ionosphere let the plasma frequency vary linearly from 0 at a height of 100 km to 10 MHz at a height of 500 km. A 5 MHz pulse is vertically sent towards the ionosphere. Find the round trip delay and the virtual height of the reflecting layer.

Solution:

Plasma frequency as a function of height is $f_p(h) = \frac{10(h-100)}{400}$ MHz

The refractive index

$$n = \left\{1 - \frac{f_p^2}{f^2}\right\}^{1/2}$$

$$= \left\{1 - \frac{(h-100)^2/1600}{(5)^2}\right\}^{1/2}$$

The velocity of the pulse in the ionosphere is $v_g = cn$

$$v_g = c\left\{1 - \frac{(h-100)^2}{4 \times 10^4}\right\}^{1/2}$$

v_g will become zero at h = 300 km.

Round trip delay

$$t_d = 2\left\{\frac{100\ km}{c} + \int_o^{300\ km} \frac{dh}{v_g}\right\}$$

$$= 2\left\{\frac{100\ km}{c} + \int_o^{300\ km} \frac{dh}{c\left\{1 - \frac{(h-100)^2}{4\times10^4}\right\}^{1/2}}\right\}$$

$$= 2\left\{\frac{100}{c} + \frac{400\pi}{3c}\right\}$$

$= 3.4592\ \mu s$

The virtual height $= \frac{ct_d}{2} = \frac{3\times10^8 \times 3.4952\times10^{-6}}{2} = 524.25$ km

10.4.3 Attenuation of Waves in the Ionosphere

As seen above, in presence of the frictional drag, the ionosphere has complex permittivity, i.e. it has non-zero conductivity. The attenuation constant for a medium having dielectric constant ϵ_r and conductivity σ is given by (see Eqn (4.110))

$$\alpha = \omega\sqrt{\frac{\mu_0\epsilon_0\epsilon_r}{2}\left(\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon_0^2\epsilon_r^2}} - 1\right)} \tag{10.66}$$

Substituting for $\epsilon_r = 1 - \frac{Ne^2/m\epsilon_0}{\omega^2+\nu^2}$ and $\sigma = \frac{Ne^2\nu}{m(\omega^2+\nu^2)}$ from Eqn (10.47), we, can get an expression for α which for frequency $\omega >> \omega_p$ reduces to

$$\alpha = \frac{60\pi Ne^2\nu}{\sqrt{\epsilon_r}m(\nu^2+\omega^2)} \tag{10.67}$$

$$\approx \frac{60\pi Ne^2\nu}{\sqrt{\epsilon_r}m\omega^2} \qquad \omega^2 >> \nu^2 \tag{10.68}$$

The attenuation constant for a wave varies inversely with square of the frequency. Higher frequencies, therefore, have less attenuation compared to the lower frequencies. Hence to reduce the propagation loss higher frequencies are preferred. However, higher frequencies require higher launching angle (by Secant law) giving higher skip distance.

Most of the attenuation of the radio waves takes place in D and E layers of the ionosphere where $\omega \sim \nu$. In higher layers $\omega >> \nu$ and the propagation is practically lossless.

10.4.4 Maximum Usable Frequency (MUF)

In the earlier sections we observed that frequencies higher than the critical frequency can be reflected by a layer by launching them at an angle. Of course for the flat ionosphere (and flat earth), there is no limit on the launching angle and it appears that an arbitarily high frequency can be reflected by the ionosphere. In reality, however, this is not true as the ionosphere is spherical in nature as shown in Fig 10.20. Even if we launch a wave tangential to the earth surface, the launchng angle at the ionsophere is not $\pi/2$ but is only ϕ_{max} given by

$$\sin\phi_{max} = \frac{R}{R+h} \tag{10.69}$$

where R is the radius of the earth and h is the height of the ionospheric layer (F-layer) above the earth's surface. Substituting $R \approx 6400$ km and $h \simeq 500$ km we get $\phi_{max} \approx 68°$. The maximum frequency which can be reflected by the F- layer then would be

$$f_{max} = f_{cr}\sec\phi_{max} \approx 2.7 f_{cr} \tag{10.70}$$

This is then the maximum usable frequency (MUF) for the F_2 layer of the ionosphere. It should be noted that, since, higher the frequency higher is the skip distance, the MUF has the largest skip distance.

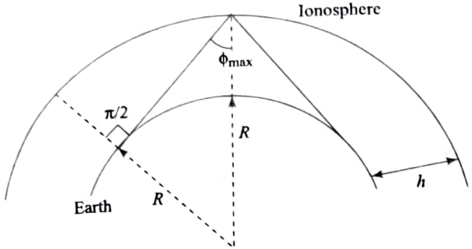


Fig. 10.20 *Maximum launching angle and maximum useable frequency.*

Due to diurnal variation in the electron density, the MUF also varies during the day. The frequencies which have reflection during day time might escape the ionosphere in the night, disrupting the communication. Typically MUF can vary by a factor of 2–3 during the day.

10.5 DIELECTRIC CONSTANT OF MAGNETIZED PLASMA

The motion of electrons and ions is influenced not only by the electric field of the propagating wave but is also affected by the earth's magnetic field. Even at a height of few hundred kms above the earth's surface, the earth's magnetic field is large enough to deflect the path of the electrons accelerated by the electric field of the propagating wave. The ionosphere is, therefore, a magnetized plasma . In general, we first investigate the EM wave propagation inside a magnetized plasma, and then specifically study the behavior of the ionosphere. It will be shown that the permittivity of magnetized plasma is no more a scalar quantity. The magnetized plasma is an anisotropic medium and consequently the propagation characteristics are different in different directions. The permittivity of the magnetized plasma can be derived by incorporating the Lorentz force in the equation of motion of the electrons.

Let us assume that a plasma is magnetized with a DC magnetic flux density $\mathbf{B_0}$. Also, without losing generality, let us assume that the DC magnetic field is oriented in z-direction ($\mathbf{B_0} = B_0\hat{\mathbf{z}}$). Let us now consider propagation of an eletromagnetic wave with electric field $\mathbf{E}$ and magnetic field $\mathbf{H}$ (magnetic flux density $\mathbf{B}$). The total magnetic flux density acting on a change, therefore, is

$$\mathbf{B}_{Total} = \mathbf{B}_0 + \mathbf{B}e^{j\omega t} = \mathbf{B}_0 + \mu\mathbf{H}e^{j\omega t} \tag{10.71}$$

Assuming again that, the ions are stationary, the equation of motion of an electron is

$$m\frac{d\mathbf{v}}{dt} = -e(\mathbf{E} + \mathbf{v}\times\mathbf{B}_{Total}) - m\nu\mathbf{v} \tag{10.72}$$

$$\Rightarrow \quad j\omega m\mathbf{v} = -e(\mathbf{E} + \mathbf{v}\times(\mathbf{B_0} + \mu\mathbf{H})) - m\nu\mathbf{v} \tag{10.73}$$

Generally, we have $\mu\mathbf{H} << \mathbf{B_0}$, giving

$$j\omega m\mathbf{v} = -e(\mathbf{E} + \mathbf{v}\times\mathbf{B_0}) - m\nu\mathbf{v} \tag{10.74}$$

Now, the vector conduction current density

$$\mathbf{J} = -eN\mathbf{v} \tag{10.75}$$

$$\Rightarrow \quad \mathbf{v} = -\frac{\mathbf{J}}{eN} \tag{10.76}$$

where N is the electron density.
Substituting for $\mathbf{v}$ in Eqn (10.74) and rearranging we get

$$-m(j\omega + \nu)\frac{\mathbf{J}}{eN} - \frac{\mathbf{J}}{N}\times\mathbf{B_0} = -e\mathbf{E} \tag{10.77}$$

$$\Rightarrow \quad (j\omega + \nu)\mathbf{J} + \frac{eB_0}{m}\mathbf{J}\times\hat{\mathbf{z}} = \frac{e^2N}{m}\mathbf{E} \tag{10.78}$$

The quantity (eB_0/m) is called the cyclotron frequency and is denoted by ω_c. Noting that $e^2N/m = \epsilon_0\omega_p^2$, we can rewrite Eqn (10.78) as

$$(j\omega + \nu)\mathbf{J} + \omega_c\mathbf{J}\times\hat{\mathbf{z}} = \epsilon_0\omega_p^2\mathbf{E} \tag{10.79}$$

Equation (10.79) can be written in a matrix form as

$$\begin{bmatrix} j\omega+\nu & \omega_c & 0 \\ -\omega_c & j\omega+\nu & 0 \\ 0 & 0 & j\omega+\nu \end{bmatrix}\begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} = \epsilon_0\omega_p^2\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{10.80}$$

Inverting the equation we get

$$\begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} = \frac{\epsilon_0\omega_p^2}{(\omega_c^2 - \omega^2 + \nu^2 + 2j\omega\nu)(j\omega+\nu)}$$

$$\begin{bmatrix} (j\omega+\nu)^2 & -\omega_c(j\omega+\nu) & 0 \\ \omega_c(j\omega+\nu) & (j\omega+\nu)^2 & 0 \\ 0 & 0 & (j\omega+\nu)^2+\omega_c^2 \end{bmatrix}\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{10.81}$$

The above Eqn (10.81) is of the form

$$\mathbf{J} = \overline{\overline{\sigma}} : \mathbf{E} \tag{10.82}$$

Where $\bar{\bar{\sigma}}$ is a tensor (3×3 matrix). We therefore find that for the magnetized plasma the conductivity is a tensor. The operation in Eqn (10.82) represents product of a matrix and a vector.

Now from the Maxwell's curl equation we have

$$\nabla \times \mathbf{H} = j\omega\epsilon_0 \mathbf{E} + \bar{\bar{\sigma}} : \mathbf{E} \tag{10.83}$$

Noting that $\mathbf{E} = \bar{\bar{\mathbf{I}}} : \mathbf{E}$, where $\bar{\bar{\mathbf{I}}}$ is a (3×3) identity matrix, we get

$$\nabla \times \mathbf{H} = j\omega\varepsilon_0 \left(\bar{\bar{I}} + \frac{\bar{\bar{\sigma}}}{j\omega\varepsilon_0} \right) : \mathbf{E} = j\omega\epsilon_0 \bar{\bar{\varepsilon}}_r : \mathbf{E} \tag{10.84}$$

The complex dielectric constant of the magnetized plasma, therefore, is

$$\bar{\bar{\varepsilon}}_r = \bar{\bar{I}} + \frac{\bar{\bar{\sigma}}}{j\omega\epsilon_0} \tag{10.85}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{10.86}$$

$$- \frac{j\omega_p^2}{\omega(\omega_c^2 - \omega^2 + \nu^2 + 2j\nu\omega)(j\omega + \nu)}$$

$$\times \begin{bmatrix} (j\omega+\nu)^2 & -\omega_c(j\omega+\nu) & 0 \\ \omega_c(j\omega+\nu) & (j\omega+\nu)^2 & 0 \\ 0 & 0 & (j\omega+\nu)^2 + \omega_c^2 \end{bmatrix}$$

$$\equiv \begin{bmatrix} K_1 & -jK_2 & 0 \\ jK_2 & K_1 & 0 \\ 0 & 0 & K_3 \end{bmatrix} \tag{10.87}$$

$$\text{where } K_1 = 1 - \frac{\omega_p^2(1 - j\nu/\omega)}{\omega^2 - \omega_c^2 - \nu^2 - 2j\omega\nu} \tag{10.88}$$

$$K_2 = \frac{\omega_p^2\omega_c/\omega}{\omega^2 - \omega_c^2 - \nu^2 - 2j\omega\nu} \tag{10.89}$$

$$K_3 = 1 - \frac{\omega_p^2}{\omega^2(1 - j\nu/\omega)} \tag{10.90}$$

$\bar{\bar{\varepsilon}}_r$ is a tensor, and hence the magnitized plasma is an anisotropic medium, i.e. the directions of the electric field vector and the electric displacement vector are not the same inside a magnetized plasma.

One can verify from the above expression that for a non magnetized plasma when $\omega_c = 0$, $K_2 \equiv 0$ and hence $\bar{\bar{\epsilon}}_r$ is a diagonal matrix with equal diagonal elements making the medium an isotropic medium.

10.6 PLANE WAVE PROPAGATION IN THE IONOSPHERE

In general, propagation of a wave in the ionosphere (magnetized plasma) is a rather complex phenonomenon. Waves travelling in different directions see different dielectric constants and therefore behave differently. To develop basic understanding of the wave propagation in the ionosphere, let us investitgate two specific cases (i) Longitudinal Propagation (ii) Transverse Propagation. In longitudinal propagation the wave travels along $\mathbf{B_0}$, whereas in the transverse propagation, the wave travels perpendicular to $\mathbf{B_0}$. For transverse propagation we may further consider two cases, one with polarization perpendicular to $\mathbf{B_0}$, and other with polarization parallel to $\mathbf{B_0}$. Overall then we have to investigate three cases, one case of longitudinal propagation and two cases of transverse propagation.

For a uniform plane wave, the electric and magnetic fields can be written as

$$\mathbf{E} = \mathbf{E}_0 e^{-j\mathbf{k}\cdot\mathbf{r}} \tag{10.91}$$

$$\mathbf{H} = \mathbf{H}_0 e^{-j\mathbf{k}\cdot\mathbf{r}} \tag{10.92}$$

where $\mathbf{k}$ is the vectors propagation constant of the wave. It has been shown in Chapter 5 (Eq. 5.15) that

$$\nabla \times \equiv -j\mathbf{k}\times \tag{10.93}$$

$$\text{and} \qquad \mathbf{H} = \frac{1}{\omega\mu}\mathbf{k}\times\mathbf{E} \tag{10.94}$$

substituting for $\mathbf{H}$ in (10.84) we get

$$\frac{1}{\omega\mu}(-j)\mathbf{k}\times(\mathbf{k}\times\mathbf{E}) = j\omega\epsilon_0\bar{\bar{\epsilon}}_r : \mathbf{E} \tag{10.95}$$

$$\Rightarrow \qquad \mathbf{k}\times\mathbf{k}\times\mathbf{E} + \omega^2\mu\epsilon_0\bar{\bar{\epsilon}}_r : \mathbf{E} = 0 \tag{10.96}$$

$$\Rightarrow \qquad \mathbf{k}\times\mathbf{k}\times\mathbf{E} + k_0^2\bar{\bar{\epsilon}}_r : \mathbf{E} = 0 \tag{10.97}$$

where, $k_0^2 = \omega^2\mu\epsilon_0$.

Equation (10.97) is the wave equation in the ionosphere. The equation is to be solved for the three special cases mentioned above.

10.6.1 Transverse Propagation with Parallel Polarization

The wave propagation geometry for this case is shown in Fig 10.21. In the presence of the electric field of the wave, the electrons acquire velocity $\mathbf{v}$, in the direction parallel to the magnetic field $\mathbf{B_0}$. Consequently, $\mathbf{v}\times\mathbf{B_0}$ force is identically zero on the electrons and the magnetic field does not have any influence on the motion of the electrons. The wave would then propagate as it would propagate in the non-magnetized plasma. In this configuration, nothing special happens to the wave and hence the wave is called the 'ordinary wave'. Substituting for $\mathbf{k} = k\hat{\mathbf{x}}$ and $\mathbf{E} = E\hat{\mathbf{z}}$ in the wave Eqn (10.97), we get

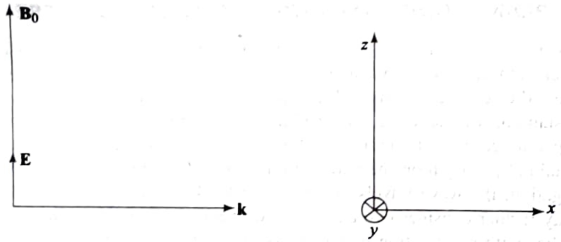


Fig. 10.21 *Propagation of a wave with the E-vector along the direction of the magnetic field.*

$$k^2\hat{\mathbf{x}} \times \hat{\mathbf{x}} \times \hat{\mathbf{z}}E + k_0^2\overline{\overline{\epsilon}}_r : (E\hat{\mathbf{z}}) = 0$$

$$\Rightarrow \begin{bmatrix} k_0^2K_1 & -jk_0^2K_2 & 0 \\ jk_0^2 & k_0^2K_1 & 0 \\ 0 & 0 & k_0^2K_3 - k^2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ E \end{bmatrix} = 0 \qquad (10.98)$$

Since, E cannot be zero, we should have

$$k_0^2K_3 - k^2 = 0$$

$$\Rightarrow \quad k = k_0\sqrt{K_3} = k_0\sqrt{1 - \frac{\omega_p^2}{\omega^2(1 - j\nu/\omega)}} \qquad (10.99)$$

This propagation constant is same as that would be obtained in a non magnetized plasma.

10.6.2 Transverse Propagation with Perpendicualr Polarization

The propagation geometry in this case is as shown in Fig. 10.22. The wave vector is in x-direction and the electric field is in y-direction which is perpendicular to $\mathbf{B}_0$. Due to electric field, the electrons start moving in y-direction and experience a $\mathbf{v} \times \mathbf{B}_0$ force in the x-direction. The electrons then start gyrating around the magnetic field $\mathbf{B}_0$ and consequently give conduction current which has both, x and y components.

Since, the electric field is oriented in y-direction, we have $\mathbf{E} = E\hat{\mathbf{y}}$. The wave

equation now is

$$k^2\hat{\mathbf{x}} \times \hat{\mathbf{x}} \times \hat{\mathbf{y}}E + k_0^2\overline{\overline{\epsilon}}_r : (E\hat{\mathbf{y}}) = 0$$

$$\Rightarrow \begin{bmatrix} k_0^2K_1 & -jk_0^2K_2 & 0 \\ jk_0^2K_2 & k_0^2K_1 - k^2 & 0 \\ 0 & 0 & k_0^2K_3 \end{bmatrix} \begin{bmatrix} 0 \\ E \\ 0 \end{bmatrix} = 0 \qquad (10.100)$$

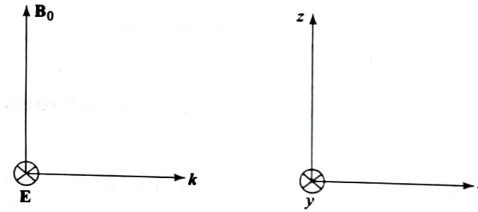


Fig. 10.22 *Propagation of a TEM wave with the E-vector perpendicular to the direction of the magnetic field.*

Equation (10.100) cannot be satisfied without making K_2 identically zero. However, K_2 is directly proportional to ω_c. Making K_2 zero makes ω_c also zero which is the case of non magnetized plasma. We can, therefore, conclude that only E_y component cannot exist for transverse propagation. Due to motion of electrons in x-direction, an electric field in the direction of wave vector will be induced. The wave, therefore, does not remain purely transverse. The electric field in this case then have to be written as

$$\mathbf{E} = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} \qquad (10.101)$$

Equation (10.100) then becomes

$$\begin{bmatrix} k_0^2K_1 & -jk_0^2K_2 & 0 \\ jk_0^2K_2 & k_0^2K_1 - k^2 & 0 \\ 0 & 0 & k_0^2K_3 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ 0 \end{bmatrix} = 0 \qquad (10.102)$$

Making determinant of the matrix zero we get

$$k = k_0\sqrt{\frac{K_1^2 - K_2^2}{K_1}} \qquad (10.103)$$

Also from Eqn (10.102) we get

$$\frac{E_x}{E_y} = j\frac{K_2}{K_1} \tag{10.104}$$

From Eqn (10.104) we note that $|E_x| \neq |E_y|$ since $K_1 \neq K_2$. Also the two components E_x and E_y are not in phase (the phase difference is $\pi/2$). The wave therefore is elliptically polarized. The ellipse of polarization is in the plane which is perpendicular to $\mathbf{B_0}$ but parallel to the wave vector (see Fig. 10.23).

Fig. 10.23 *Rotation of the electric field vector due to the presence of a longitudinal field.*

This wave has a very special behavior and hence called the 'extra-ordinary wave'.

For a general case of transverse propagation, when the electric field is neither along $\mathbf{B}_0$ nor perpendicular to it, the electric field can be resolved into two components, one along $\mathbf{B}_0$ and other perpendicualr to $\mathbf{B}_0$. The wave, therefore, gets splits into two waves: the ordinary wave corresponding to the parallel polariztion and the extra-ordinary wave corresponding to the perpendicular polarization. The two waves have different propagation constants, i.e. they travel with different phase velocities and undergo different attenuation.

10.6.3 Longitudinal Propagation

In this case the wave propagates along $\mathbf{B}_0$ i.e. $\mathbf{k} = k\hat{\mathbf{z}}$. The propagation geometry is shown in Fig 10.24. Here, since $\mathbf{E}$ can be oriented in any direction in the xy-plane we can assume

$$\mathbf{E} = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} \tag{10.105}$$

Substituting $\mathbf{E}$ in the wave Eq. (10.97) we get

$$k^2\hat{\mathbf{z}} \times \hat{\mathbf{z}} \times (E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}}) + k_0^2\bar{\bar{\epsilon}}_r : (E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}}) = 0 \tag{10.106}$$

$$\Rightarrow \quad k^2\left\{-\hat{\mathbf{x}}E_x - \hat{\mathbf{y}}E_y\right\} + k_0^2\bar{\bar{\epsilon}}_r : (E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}}) = 0 \tag{10.107}$$

$$\Rightarrow \quad \begin{bmatrix} k_0^2K_1 - k^2 & -jk_0^2K_2 & 0 \\ jk_0^2K_2 & k_0^2K_1 - k^2 & 0 \\ 0 & 0 & k_0^2K_3 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ 0 \end{bmatrix} = 0 \tag{10.108}$$

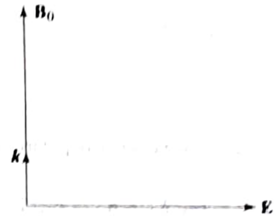


Fig. 10.24 *Propagation of a TEM wave along the direction of the magnetic field; longitudinal propagation.*

Again making determinant of the matrix zero, we get

$$(k_0^2K_1 - k^2)^2 - k_0^4K_2^2 = 0 \tag{10.109}$$

$$\Rightarrow \quad k^2 = k_0^2(K_1 \pm K_2) \tag{10.110}$$

From Eqn (10.110), we get two solutions for k as

$$k = k_L = k_0\sqrt{K_1 + K_2} \tag{10.111}$$

and

$$k = k_R = k_0\sqrt{K_1 - K_2} \tag{10.112}$$

The ratio of E_x and E_y from Eqn (10.108) is

$$\frac{E_y}{E_x} = \frac{k_0^2K_1 - k^2}{-jk_0^2K_2} \tag{10.113}$$

$$\text{For} \quad k = k_L = k_0\sqrt{K_1 + K_2}, \quad E_y/E_x = +j \tag{10.114}$$

$$\text{and for} \quad k = k_R = k_0\sqrt{K_1 - K_2}, \quad E_y/E_x = -j \tag{10.115}$$

The first thing to note from above analysis is that, the wave splits into two waves both travelling in same direction but with different propagation constants, k_L and k_R. The wave which travels with propagation constant k_L has $E_y/E_x = j$. That is, this wave is left circularly polarized. On the other hand, the wave which has propagation constant k_R, has $E_y/E_x = -j$ and consequently is right circularly polarized (see Chapter 4 for polarization of a wave). This means that the LHC and RHC are the fundamental modes of the longitudninal propagation. Any arbitarily polarized wave, when launched along the direction of the magnetic field in the magnetized plasma, is split into LHC and RHC waves which propagate with different phase velocities.

Now, assuming $\omega >> \nu$, i.e. almost a collisionless plasma, we get

$$K_1 = 1 - \frac{\omega_p^2}{\omega^2 - \omega_c^2} \tag{10.116}$$

$$K_2 = \frac{\omega_p^2 \omega_c/\omega}{\omega^2 - \omega_c^2} \tag{10.117}$$

The propagation constants of the two circularly polarized waves then become

$$LHC: \quad k_L = k_0\sqrt{K_1 + K_2} = k_0\sqrt{1 - \frac{\omega_p^2}{\omega(\omega + \omega_c)}} \tag{10.118}$$

$$RHC: \quad k_R = k_0\sqrt{K_1 - K_2} = k_0\sqrt{1 - \frac{\omega_p^2}{\omega(\omega - \omega_c)}} \tag{10.119}$$

It can be noted that when $\omega \to \omega_c$, k_R becomes imaginary and the RHC wave shows attenuation. This is due to the resonanace between the wave and the gyrating electrons. The electrons rotate around the magnetic field with frequency ω_c. When $\omega \to \omega_c$, the electric field of the RHC wave and the electrons rotate almost together giving energy transfer from the wave to the electrons. Consequently, the RHC wave shows high absorption for frequencies close to ω_c .

For LHC on the other hand, the elctric field and the electrons rotate in opposite directions and hence there is no resonance for the LHC wave.

The above discussion tells us that for longitudinal propagation, the fundamental modes of propagation are RHC and LHC. Moreover, the two modes travel with different velocities and undergo different absorption. These differential characteristics of RHC and LHC give rise to a phenomenon what is called the 'Faraday rotation'.

EXAMPLE 10.9 In a magnetised plasma, the magnetic flux density is 0.5 Gauss (0.5×10^{-4} Wb/m^2), and the magnetic field is oriented along z-direction. The electron density in the plasma is 10^{10}/m^3. If a 3 MHz y-oriented electric of 10 V/m is impressed on the plasma, find the conduction and displacement current densities and their directions.

Solution:

The cyclotron frequency

$$\omega = \frac{-eB}{m} = \frac{1.59 \times 10^{-19} \times 0.5 \times 10^{-4}}{9 \times 10^{-31}} = 8.83 \times 10^6 \text{ rad/s}$$

The plasma frequency

$$\omega_p = 2\pi\sqrt{81N} = 18\pi\sqrt{10^{10}}$$
$$= 18\pi \times 10^5 \text{ rad/s}$$

Assuming $\nu = 0$, the conductivity of the plasma is

$$\overline{\overline{\sigma}} = \frac{\epsilon_0 \omega_p^2}{(\omega_c^2 - \omega^2)} \begin{bmatrix} j\omega & -\omega_c & 0 \\ \omega_c & j\omega & 0 \\ 0 & 0 & \frac{\omega_c^2 - \omega^2}{j\omega} \end{bmatrix}$$

For 3 MHz frequency $\omega = 6\pi \times 10^6$ rad/s

$$\overline{\overline{\epsilon}}_r = \overline{\overline{I}} + \frac{\overline{\overline{\sigma}}}{j\omega\epsilon_0}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{\omega_p^2}{j\omega(\omega_c^2 - \omega^2)} \begin{bmatrix} j\omega & -\omega_c & 0 \\ \omega_c & j\omega & 0 \\ 0 & 0 & \frac{\omega_c^2 - \omega^2}{j\omega} \end{bmatrix}$$

$$= \begin{bmatrix} 1 + \frac{\omega_p^2}{\omega_c^2 - \omega^2} & \frac{-\omega_p^2\omega_c}{j\omega(\omega_c^2 - \omega^2)} & 0 \\ \frac{-\omega_p^2\omega_c}{j\omega(\omega_c^2 - \omega^2)} & 1 + \frac{\omega_p^2}{\omega_c^2 - \omega^2} & 0 \\ 0 & 0 & 1 - \frac{-\omega_p^2}{\omega^2} \end{bmatrix} = \begin{bmatrix} 0.8882 & -0.0494j & 0 \\ 0.0494j & 0.8882 & 0 \\ 0 & 0 & 0.09 \end{bmatrix}$$

$$= \begin{bmatrix} 0.8882 & 0 & 0 \\ 0 & 0.8882 & 0 \\ 0 & 0 & 0.09 \end{bmatrix} + j \begin{bmatrix} 0 & -0.0494 & 0 \\ 0.0494 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Electric field vector can be written as $E = 10\hat{\mathbf{y}}$

$$J_d = j8.882 \quad \omega\epsilon_0\hat{\mathbf{y}} = 1.48 \times 10^{-3}\hat{\mathbf{y}}$$
$$J_c = 0.494 \quad \omega\epsilon_0\hat{\mathbf{x}} = 0.082 \times 10^{-3}\hat{\mathbf{x}}$$

10.6.4 Faraday Rotation

Let us consider a linearly polarized wave longitudinally propagating in a magnetized plasma, the ionosphere. Without losing generality let the wave initially be x-polarized. Assuming that the wave propagates in z-direction, a linearly polarized wave can be visualized as a superposition of two circularly polarized wave as

$$\mathbf{E} = E_0 e^{-jk_0 z}\hat{\mathbf{x}} \tag{10.120}$$
$$= \frac{E_0}{2}\hat{\mathbf{L}}e^{-jk_0 z} + \frac{E_0}{2}\hat{\mathbf{R}}e^{-jk_0 z} \tag{10.121}$$

where $\hat{\mathbf{L}} = (\hat{\mathbf{x}} + j\hat{\mathbf{y}})$ and $\hat{\mathbf{R}} = (\hat{\mathbf{x}} - j\hat{\mathbf{y}})$ are LHC and RHC unit vectors respectively. The wave enters the plasma at $z = 0$ (say). For $z > 0$ then, LHC and RHC waves travel with k_L and k_R propagation constants respectively. If the length of the ionosphere is L, the electric field of the exiting wave can be written as

$$\mathbf{E} = \frac{E_0}{2}(\hat{\mathbf{x}} + j\hat{\mathbf{y}})e^{-jk_L L} + \frac{E_0}{2}(\hat{\mathbf{x}} - j\hat{\mathbf{y}})e^{-jk_R L} \tag{10.122}$$

The phase difference between LHC and RHC is

$$\phi = (k_L - k_R)L \tag{10.123}$$

and the amplitudes of LHC and RHC are same, $E_0/2$. The resultant field therefore is still linearly polarized but the plane of polarization is at an angle $\phi/2$ with respect to the x-axis (See Fig. 10.25).

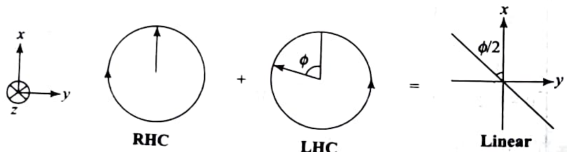


Fig. 10.25 *Linear polarization as a superposition of two circular polarizations.*

We, therefore, find that a linearly polarized wave remains linearly polarized after longitudinal propagation in the ionosphere but the plane of polarization is rotated by $\phi/2$. This rotation is called the *Faraday rotation*. Substituting for k_L and k_R in Eqn (10.123), the Faraday rotation is

$$\phi_{FR} = \frac{(k_L - k_R)L}{2} = \frac{L}{2}k_0\{\sqrt{K_1 + K_2} - \sqrt{K_1 - K_2}\} \tag{10.124}$$

If $\omega >> \omega_c$ and ω_p, we can binomially expand the square root and retain first order terms only to get

$$\phi_{FR} \approx \frac{\omega_p^2 \omega_c k_0 L}{2} = \frac{e^3 B_0}{2cm^2\epsilon_0\omega^2}NL \tag{10.125}$$

The quantity NL gives the total electronic content along the direction of the wave per unit area of the ionosphere. The Faraday rotation is inversely proportional to the square of the frequency. At higher microwave frequencies, the Faraday rotation is negligibly small but at VHF and UHF frequencies the faraday rotation

is hundreds of radians. The electric field vector at these frequencies undergoes many rotations while passing through the ionosphere (see Fig. 10.26).

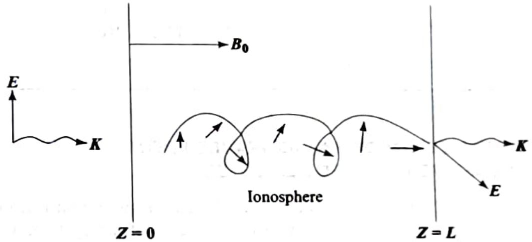


Fig. 10.26 *Farady rotation in the ionosphere.*

As mentioned earlier, the ionosphere is a statistically varying medium. It's electron content keeps fluctuating as a function of time. Consequently, the Faraday rotation also randomly varies as a function of time. Now, imagine a satellite transmitting a linearly polarized wave towards the earth. Due to Faraday rotation, the orientation of the received electric field vector is not the same as that of the tramsmitted wave. Moreover the orientation of the electric field vector keeps varying as a function of time. If we now receive the linearly polarized wave by a linearly polarized antenna, the output of the antenna fluctuates as a function of time due to time dependent polarization mismatch. The Faraday rotation, therefore, gives fading in the received signal which is undesirable for reliable communication.

Instead of a linear polarization, if the satellite transmits a circularly polarized wave, there is no Faraday rotation and consequently there is no fading. The satellite communication, therefore, prefers circular poalrization over linear polarization. Even if the transmitted wave is linearly polarized, fading can be avoided by using circularly polarized receiving antenna. In this case, however, the reception efficiency is half only.

EXAMPLE 10.10 A 15 MHz wave travels a distance of 400 km in the ionosphere. If the average electron density is $10^{11}/\text{m}^3$ and the magnetic flux density is 10μTesla, find the faraday rotation of the wave.

Solution:

Plasma frequency $f_p = 9\sqrt{10^{11}} = 2.846$ MHz

$$\text{Cyclotron frequency } \omega_c = \frac{1.59 \times 10^{-19} \times 10 \times 10^{-6}}{9 \times 10^{-31}}$$
$$= 1.767 \times 10^6 \text{ Hz}$$

Since, $\omega >> \omega_c$ and $f >> f_p$. We can use approximate relation to get Faraday rotation

$$\phi_{FR} = \frac{e^3 \times 10 \times 10^{-6} \times 10^{11} \times 400 \times 10^3}{2 \times 3 \times 10^8 m^2 \epsilon_0 (2\pi \times 15 \times \times 10^6)^2}$$
$$= 42.12 \text{ rad}$$

10.7 DIFFRACTION OF RADIO WAVES FROM IONOSPHERIC IRREGULARITIES

In the previous sections, we treated the ionosphere as a smooth uniform layered medium. In this medium, the radio waves reflect and refract systematically. In reality, however, the ionosphere has random fluctuation in the electron density over and above the mean electron density of a layer. In fact the ionosphere consists of electron clouds with varying density and size. The cloud size may vary from few hundred meters to few thousand meters. The clouds randomly wander in the layer with a speed of few tens of m/s. When radio waves pass through these clouds, instead of systematic refraction, the waves get scattered. The scattering angle depends upon the deviation of the electron density and the size of the cloud. The scattered waves from different clouds interfere and give random fluctuations in the strength of the incoming radio signals. This phenomenon is called *fading*. In the following section we investigate the basic mechanism of signal fading using semi quantitative approach. In general, the problem of scattering of radio waves from random medium is an extremely complex problem. Here, however, we adopt an approach which is not vary accurate but is adequate to provide physical understanding of scattering of EM waves from a random medium.

Let us assume that the ionosphere has a mean electron density N_0 and peak deviation in the density ΔN. The dielectric constant of the ionosphere is

$$\epsilon_r = 1 - \frac{Ne^2/m\epsilon_0}{\omega^2} = 1 - \frac{(N_0 + \Delta N)e^2/m\epsilon_0}{\omega^2} \tag{10.126}$$

If the thickness of the ionosphere is l, the phase change which the wave undergoes after passing through the ionosphere is

$$\phi = \beta_0 \sqrt{\epsilon_r} l = \frac{2\pi l}{\lambda}\left\{1 - \frac{(N_0 + \Delta N)e^2/m\epsilon_0}{\omega^2}\right\}^{1/2} \tag{10.127}$$

Assuming that the frequency ω is much greater than the plasma frequency, we can approximate Eqn (10.127) to get

$$\phi \simeq \frac{2\pi l}{\lambda}\left\{1 - \frac{(N_0 + \Delta N)e^2/m\epsilon_0}{2\omega^2}\right\} \tag{10.128}$$

$$= \frac{2\pi l}{\lambda}\left\{1 - \frac{N_0 e^2/m\epsilon_0}{2\omega^2}\right\} - \frac{2\pi l e^2}{2\lambda m\epsilon_0\omega^2}\Delta N \tag{10.129}$$

$$= \phi_0 - \Delta\phi \quad \text{(say)} \tag{10.130}$$

ϕ_0 is the uniform phase shift of the wave whereas, $\Delta\phi$ gives the random spatial variation of the phase. A constant phase front when passes through the ionosphere develops a random spatial phase variation. The ionospheric layer, therefore, acts like a thin phase scene as shown in Fig 10.27.

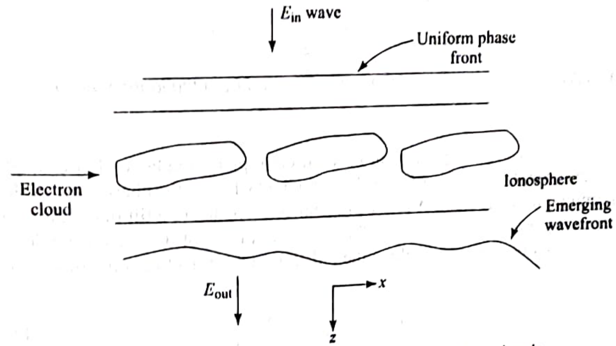


Fig. 10.27 *Ionosphere as a collection of electron clouds.*

If the electric field of the inicident wave is

$$E_{in} = E_o e^{-jk_0 z} \tag{10.131}$$

the electric field of the emerging wave will be

$$E_{out} = E_0 e^{-jk_0 z} \cdot e^{j\phi_0} \cdot e^{-j\Delta\phi(x)} \tag{10.132}$$

$$= E_0' e^{-j\Delta\phi(x)} \quad \text{for a given } z \tag{10.133}$$

To make analysis simpler, let us assume that $\Delta\phi(x)$ has sinusoidal variation, i.e.

$$\Delta\phi(x) = \phi_1 \sin\frac{2\pi x}{\Lambda} \tag{10.134}$$

where Λ is the spatial period of the phase and ϕ_1 is the peak phase deviation above the average. Now, consider one cycle of the phase as shown in Fig. 10.28(a).

Assuming $\Lambda >> \lambda$ (wavelength of the wave), the phase front can be approximated by a piecewise linear function as shown Fig.10.28(b). That is to say that the emerging wavefront can be visualized as a collection of small phase fronts AB,

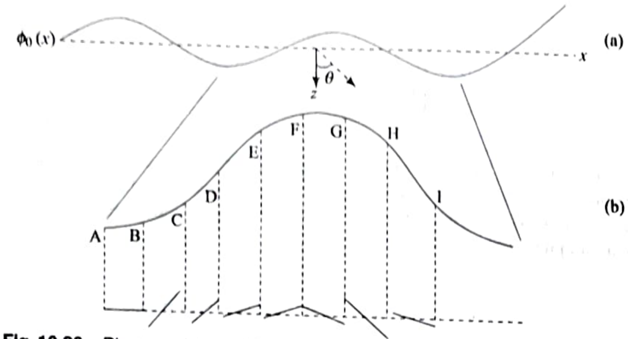


Fig. 10.28 *Phase variation of the wave diffracted from the ionospheric irregularities.*

BC, CD,···, which are planes but have different phase gradients. For example, the phase front FG has almost zero phase gradient whereas the phase front CD has maximum phase gradient. A phase gradient corresponds to a tilt in the direction of the wave propagation. It means that, different pieces of the wave front do not travel vertically downwards along the z-direction but travel in some direction θ such that

$$\sin\theta = \frac{1}{k_0}\frac{d\phi}{dx} = \frac{\lambda}{2\pi}\frac{d\phi}{dx} \tag{10.135}$$

The maximum $\frac{d\phi}{dx}$ is

$$Max\left|\frac{d\Delta\phi(x)}{dx}\right| = Max\left|\frac{2\pi}{\Lambda}\phi_1\cos\frac{2\pi x}{\Lambda}\right| = \frac{2\pi\phi_1}{\Lambda} \tag{10.136}$$

The maximum deviation in the direction of wave is given by

$$\sin\theta_{max} = \frac{\lambda}{2\pi}\frac{2\pi}{\Lambda}\phi_1 = \frac{\lambda}{\Lambda}\phi_1 \tag{10.137}$$

Since, $\Lambda >> \lambda$ and $\phi_1 \leq 2\pi$, we get

$$\theta_{max} \approx \frac{\lambda\phi_1}{\Lambda} \tag{10.138}$$

As the wave may deflect by $\pm\theta_{max}$, the scattering angle, i.e. the cone in which the waves go is

$$\theta_{scat} \simeq 2\theta_{max} \approx \frac{2\lambda\phi_1}{\Lambda} \tag{10.139}$$

Very crudely we can then say that if there are electrons clouds of size Λ the waves get scattered by the cloud over an angle θ_{scat}.

It should be noted here that at the emergence point, the wave has only phase variation but the amplitude of the wave is same for every value of x.

The scattered waves from different electron clouds interfere to give spatial intensity variation along x-direction (See Fig. 10.29). For getting full interference, the wave must travel a minimum distance from the ionosphere such that every cone overlaps with the adjacent cone. We therefore get

$$H\theta_{scat} \geq \Lambda \tag{10.140}$$

$$H \geq \frac{\Lambda}{\theta_{scat}} = \frac{\Lambda^2}{2\lambda\phi_1} \tag{10.141}$$

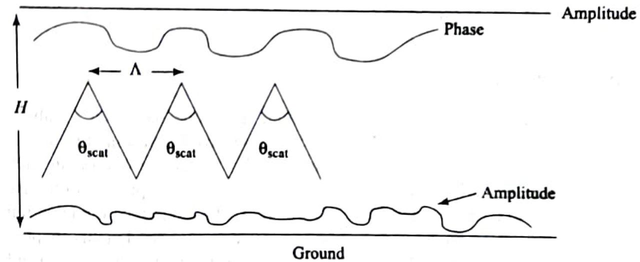


Fig. 10.29 *Interference of the diffracted radio waves from ionospheric irregularities.*

For the ionosphere, taking $\Lambda \approx 300m$, $\phi_1 \approx 1$rad, and wavelength of the wave 10 cm, H should be greater than about 400 kms which is of the order of the height of the ionosphere above the earth. Of course this height is much smaller at longer wavelengths. In general, therefore, we find that for VHF, UHF and microwave frequencies, the scattered waves exhibit good interference and consequently show large spatial amplitude variation. The amplitude of the radio waves on the ground is, therefore, not same at every location. The radio signal is strong at some point and weak at other.

Now, if the electron clouds move randomly in the horizontal direction, the interference pattern also moves on the ground. The spatial amplitude variations then gets converted into temporal fluctuations. The signal strength of a radio wave received from the ionosphere then fluctuates randomly over a time scale of fraction of second to tens of seconds. The signals received from the ionosphere therefore show amplitude variation from very large to almost zero over time scales of few seconds to few tens of minutes. The signal fading is an integral part of the ionospheric propagation.

The wave sent from the ground towards the ionosphere also shows temporal fluctuations after reflecting from the ionosphere, due to the same phenomenon. The fading is more dominant at low frequencies and as the frequency increases the fading reduces.

The same fading phenomenon at optical frequencies is responsible for the twinkling of the stars. Due to air density variation, the optical waves are scattered giving interference pattern on the eye and due to random movement of the interference pattern, the light reaching retina fluctuates making stars twinkle.

The same phenomenon in a more discrete fashion, takes place in cellular communication, where the radio waves get scattered from man made structures. Due to movement of the mobile user, the signal intensity fluctuates over time scales of few msec to seconds.

It is, therefore, interesting that a variety of fields have the same underlying phenomenon of scattering of electromagnetic waves. The propagation of electromagnetic waves in the ionosphere therefore provides understanding of wave propagation in many other situations.

10.8 SUMMARY

Propagation of electromagnetic waves around the globe has been a subject of interest for many decades. The subject has received new impetus due to revolution in the mobile communication. In today's wireless communication systems, practically all the concepts of electromagnetic wave propagation have been exploited to its highest level of sophistication. The modern communication essentially uses all the modes of radio wave propagation discussed in this chapter. The usual radio broadcasting takes place through the surface (medium waves) and space (FM radio) waves. The short wave broadcasting is through the ionosphere. A substantial international communication is routed through the satellite where the propagation is affected by the ionosphere. Propagation of radio waves through random media and the fading of the radio signals is an integral part of the modern mobile communication. A good understanding of the radio wave propagation is vital in designing reliable wireless and mobile communication links.

Review Questions

10.1 What are the various modes by which electromagnetic energy can be sent form one point to another?

10.2 What are strengths and limitations of various modes of propagation?

10.3 What is ground wave propagation?

10.4 What is sky wave propagation?

10.5 What is surface wave propagation?

10.6 What is space wave propagation?

10.7 What polarization is transmitted in ground wave propagation and why?

10.8 What is wave tilt? What does it depend upon?

10.9 What is typical skin depth of the earth at 100 kHz?

10.10 Typically over what distance is the ground wave propagation effective?

10.11 What is the reflection coefficient for a ground reflected wave at grazing angle?

10.12 In presence of ground reflected wave what is variation of the electric field for a wave at longer distance?

10.13 Is ground reflected wave desirable? Why?

10.14 How is the effect of ground reflected wave eliminated or reduced?

10.15 What is polarization used for space wave? Why?

10.16 Why are the microwave link towers erected on top of hills?

10.17 How does the curvature of the earth affects the space wave propagation?

10.18 What is ionosphere?

10.19 Why does the ionosphere have layered structure?

10.20 Which layers are present during day and night time?

10.21 What is plasma frequency or critical frequency?

10.22 If the frequency of a wave is less than the plasma frequency, what will be the dielectric constant of the ionosphere?

10.23 Where does the maximum attenuations of an electromagnetic wave take place inside the ionosphere?

10.24 What is MUF?

10.25 What is skip distance?

10.26 What is virtual height of a layer? Why it is called the virtual height?

10.27 Is the virtual height of a layer greater or lesser than the actual height of the layer?

10.28 What is the special feature of the pulse propagation inside the ionosphere?

10.29 How does the earth's magnetic filed affect the behavior of the ionosphere?

10.30 Physically visualize the tensor nature of the dielectric constant and conductivity of the ionosphere?

10.31 What is cyclotron frequency?

10.32 What are ordinary and extra-ordinary waves and under what circumstances is the extra ordinary wave developed?

10.33 What are the natural states of polarization of an electromagnetic wave travelling along the magnetic field inside the ionosphere?

10.34 What is Faraday Rotation?

10.35 On what factors does the Faraday rotation depend?

10.36 If the frequency of an electromagnetic wave is increased, what happens to its Faraday rotation?

10.37 Why is circular polarization preferred in satellite communication?

10.38 Do we find fading in satellite signals? Why?

10.39 Why does the ionosphere have electron clouds rather than a uniform layer?

10.40 Why do the stars twinkle and the planets do not twinkle?

10.41 Why do signals fade in mobile communication?

10.42 How can the effect of fading be reduced in mobile communication?

Problems

10.1 An AM broadcasting radio station at 550 kHz launches a vertically polarized wave. At certain distance from the station the electric field is 1 V/m. If the dielectric constant of the ground is 4 and the conductivity of the ground is 10^{-2} $\mho$/m, find the tilt angle of the wave.

10.2 In Problem 10.1 find the loss per unit area of the ground surface.

10.3 For a terrain, $\epsilon_r = 15$ and $\sigma = 10^{-3}$ $\mho$/m. Find the surface impedance of the terrain at 1 kHz. For a vertically polarized wave, the loss in the terrain is 10 mW/m^2. Find the power density of the uniform plane wave.

10.4 A vertical monopole of length $\lambda/2$ is mounted on a ground having $\epsilon_r = 20$ and $\sigma = 10^{-3}$$\mho$/m. If 1 kW of power at 2 MHz is supplied to the antenna, find the electric field at a distance of 10 km at the ground.

10.5 Two microwave towers of 100 m height are separated by a distance of 40 km. Two isotropic antennas are mounted on the towers for communication at 2.4 GHz. One of the antenna transmits 100 W power. If the terrain between the towers has $\epsilon_r = 5$ and $\sigma = 10^{-2}$$\mho$/m, find the electric field at the receiving antenna. Assume the transmitted wave to be horizontally polarized.

10.6 A TV station at 160 MHz transmits 10 kW power. The TV station antenna has $\sin\theta$ radiation pattern, where θ is measured from the horizon. The height of the TV station antenna is 150 m. A receiving domestic TV antenna is mounted at a distance of 5 km on a terrace of 20 m height. What is the electric field at the receiving antenna?

10.7 A 900 MHz mobile base station located at a height of 100 m radiates 50 W power using isotropic antenna. A wireless receiver is mounted at a distance of 2 km at a height of 6 m, has an effective aperture of 0.01 m^2. Under fully matched conditions, find the power received by the receiver.

10.8 Two parabolic dish antennas are mounted to look into each other at a distance of 30 km. The directivities of the antennas are 20 dB and 30 dB respectively. The frequency of operation is 6 GHz. Find the power loss in dB between the two antennas.

10.9 Find the number of microwave towers of 100 m height to be installed to cover the perimeter of the earth. Take the radius of the earth to be 6000 km.

10.10 Find the maximum range one can get between two microwave towers of 70 m height. What is the range increase if one of the tower is mounted on a hill of 1000 m height? Take radius of the earth to be 6000 km.

10.11 A medium has free electron density 10^4 per cc. Find the plasma frequency of the medium.

10.12 A medium has electron density of 10^{11}/m^3 and the collision frequency of 3×10^6 per sec. Find the dielectric constant and the conductivity of the medium at 10 MHz.

10.13 A medium has the plasma frequency 2 MHz. A 3 MHz wave is launched in the medium. Find the phase and the group velocities of the wave.

10.14 A 10 MHz radio pulse is transmitted through a plasma slab having electron density 10^{12}/m^3, and thickness 2 m. Find the propagation time of the pulse in the medium.

10.15 The F-layer of the ionosphere has electron density 3×10^{11}/m^3. What is the maximum frequency which is reflected from the layer if (i) launched vertically (ii) launched at an angle of 60° from the horizon?

10.16 The F-layer of the ionosphere starts from a height of 300 km. The electron density at 300 km is 10^{10}/m^3, and it increases linearly with height to reach to a maximum value of 2×10^{12}/m^3 at a height of 500 km. A 12 MHz wave is launched at an angle of 30° from the vertical. Upto what height the wave will penetrate the ionosphere?

10.17 In Problem 10.16, how much time will a 12 MHz pulse take to return to the earth?

10.18 In Problem 10.16, if a 8 MHz pulse is vertically launched, what is the return time of the pulse?

10.19 For E-layer the electron density is 10^{10}/m^3. Find the skip distance for a 2 MHz wave. The height of the layer is 200 km.

10.20 A parabolic dish having 10° wide beam launches a signal at an angle of 45° from the horizon. The signal is reflected by the F-layer having height of 400 km. Find the region on the ground which will be illuminated by the reflected signal.

10.21 Find the MUF of the E-layer. The height of the E-layer is 250 km and its density is 10^9/m^3.

10.22 Inside a magnetized plasma, the cyclotron frequency is 1 MHz, and the plasma frequency is 8 MHz. Assuming the plasma to be collisionless, find the tensor dielectric constant of the medium at 10 MHz.

10.23 In Problem 10.22, find the propagation constant of LH and RH circularly polarized waves travelling along the magnetic field in the plasma.

10.24 While analysing the magnetized plasma, why is the effect of the magnetic field of the wave ignored compared to the external magnetic field? Under what situation will this not be justified?

10.25 In the magnetized ionosphere, $B_0 = 0.1$G and the electron density is 10^{12}/m^3. The thickness of the ionosphere is 200 km. Find the Faraday rotation at 1 GHz.

10.26 The densities of the E and F layers of the ionosphere are 3.5×10^{10} and 10^{12} electrons/m^3 respectively. Two waves at 10 MHz and 2 MHz have the same skip distance of 500 km. Calculate the heights of the two layers. If both signals are transmitted simultaneously in the form of narrow pulses, find the time difference between the arrival of the two signals.

10.27 The ionosphere is assumed to have a triangular electron density distribution above a height of 100 km. The electron density is zero below 100 km and it reaches to its maximum value of 10^{12} electrons/m^3 at an altitude of 500 km. Calculate the skip distance at 12 MHz. For a launching angle of 45°, what is the MUF?

10.28 A 10 MHz plane wave propagates through a distance of 200 km in the ionosphere. The average electron density of the layer is 5×10^{10}/m^3, and the cyclotron frequency $\omega_c = 8 \times 10^6$ rad/sec. Assuming the collisions are negligible, find the amount of Faraday rotation produced. In the presence of collisions if the difference in the attenuation of the circularly polarized waves is 0.03 dB/km, find the state of polarization of the emerging wave. Assume the input wave to be linearly polarized.

CHAPTER **11**

Applications of Electromagnetic Waves

In the previous chapters, we studied the general characteristics of electromagnetic waves. Starting from the wave propagation in an unbound medium, we studied the propagational characteristic of electromagnetic waves in bound structures like waveguides and resonators. We also studied the various modes of transporting radio waves for communication purpose. The concepts of electromagnetic waves are applicable whenever we deal with energy transport at high frequencies. Since, most of the modern communication takes place with high frequency carriers the electromagnetic waves find wide applications. In this chapter, we discuss a few areas where the knowledge of the electromagnetic waves is crucial.

11.1 FIBER OPTIC COMMUNICATION

Most of the modern wide band communication systems use optical fiber as a transport medium. An optical fiber is an ultra wideband medium with extremely small loss. Consequently, very high data rate signals can be transported very efficiently over thousands of kms using the optical fibers.

In fiber optic communications the principles of electromagnetic waves find application in investigating characteristics of optical fibers and the lasers.

11.1.1 Optical Fibers

An optical fiber is a cylindrical dielectric rod called core, surrounded by a dielectric shell called cladding (see Fig 11.1).

Light propagates inside the core. In principle, an optical fiber is a cylindrical dielectric waveguide discussed in Chapter 7. For making the analysis simple, the diameter of the cladding is assumed to be infinite. This reduces the geometry to a single boundary, i.e. the interface between the core and the cladding. As we have

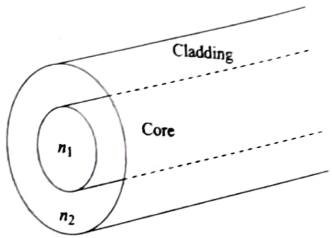


Fig. 11.1 *Basic structure of an optical fiber.*

seen in Chapter 7, for propagation of light inside the core, the refractive index of the core, n_1 has to be greater than the refractive index of the cladding, n_2. Moreover, the light, which is an electromagnetic wave, propagates along the fiber in the form of modes. As seen in Chapter 7, three types of modes can exist inside an optical fiber, namely, the transverse electric (TE) modes, transverse magnetic (TM) modes, and the hybrid (HE) modes. Depending upon the refractive indices of the core and the cladding, and the size of the core, some or all types of modes can propagate inside a fiber at a given frequency (wavelength) of light. All the modes, except the HE_{11} mode, have finite cut off frequencies. The HE_{11} mode does not have cut off frequency and the light at any wavelength can propagate inside an optical fiber in this mode.

If only HE_{11} mode propagates inside the fiber, the fiber is called the single mode fiber. If other modes also propagate, the fiber is called the multimode fiber. A characteristic parameter called the V-number is defined for an optical fiber as

$$V = \frac{\omega a}{c}(n_1^2 - n_2^2)^{1/2} \tag{11.1}$$

Where, a is the radius of the fiber core, n_1 and n_2 are refractive indices of the core and the cladding respectively, and ω is the angular frequency of the light. For a given fiber, the V-number is proportional to ω.

The most important aspect of fiber optic communication is the 'disperison' of the signal inside the optical fiber. The dispersion is defined as the pulse broadening per unit distance on the fiber. The dispersion is caused due to change in velocity of light at different frequencies and in different modes. To investigate dispersion the characteristic equation of the cylindrical dielectric waveguide is solved to get relation between the modal phase constant β and the frequency. From this relation, one obtains the phase and group velocities and subsequently the dispersion. Since, the V-number is proportional to frequency, ω, we obtain a graph between β and V. This graph is called b-V diagram of an optical fiber. A typical b-V diagram is shown in Fig 11.2. $\beta_1(\equiv \omega n_1/c)$ and $\beta_2(\equiv \omega n_2/c)$ are the wave numbers in the core and the cladding respectively.

The modal propagation constant β always lies between β_2 and β_1. As $\beta \to \beta_2$ the mode approaches the cut-off and the fields spread in the cladding. As the

V-number increases the modal propagation constant β also increases monotonically approaching asympotically to β_1. In other words, as the V-number increases, the fields get more confined inside the core.

The slope of the b-V curve $\partial\beta/\partial V$, for a mode is proportional to $\partial\beta/\partial\omega$ (since, $V \propto \omega$). Since $\partial\beta/\partial\omega = (\partial\omega/\partial\beta)^{-1} = 1/v_g$, the slope of the b-V curve is a measure of the group delay (time taken by a pulse to travel unit distance). From Fig. 11.2 we can note that b-V curve is non-linear and hence different frequencies undergo different group delays. Also, for a given frequency, i.e. for a given V-number, different b-V curves have different slopes, therefore, different group delays.

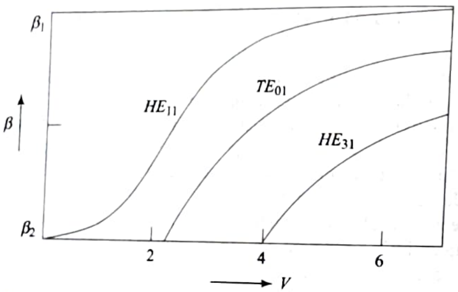


Fig. 11.2 *b-V diagram for an optical fiber. The characteristics of only few modes are shown.*

An optical signal consists of a bunch of wavelengths (typical spectral width is 2–50 nm). Also depending upon the fiber, the signal may get divided between different modes. The optical signal therefore sees a substantial variation in the group delay across its spectrum and across the modes. An optical pulse then gets distorted due to the variation in the group delay. This phenomenon is called dispersion on the optical fiber. In a multi-mode fiber, the signal distortion is dominated by the differential group delay between different modes, and the dispersion is called the 'inter-modal dispersion'. On the other hand, in a single mode fiber, the dispersion is due to differential group delay over the signal spectrum within a mode. This dispersion is then called the 'intra-modal' or 'waveguide dispersion'. The inter-modal dispersion is many orders of magnitude higher than the intra-modal dispersion. The intermodal dispersion is defined as the pulse broadening per unit length of the fiber and is given as

$$D_{\mathrm{inter}} = \frac{n_1(n_1 - n_2)}{n_2 c} \qquad (11.2)$$

D_{inter} has unit ps/km.

The intra-modal dispersion is defined as the pulse broadening per unit length of the fiber per unit spectral width of the signal, and is given as

$$D_{\mathrm{intra}} = \frac{\lambda V}{2\pi c}\frac{\partial^2 \beta}{\partial V^2} \qquad (11.3)$$

D_{intra} has unit ps/km/nm. The pulse broadening $\tau = D_{\mathrm{intra}}\sigma_\lambda L$, where σ_λ is the spectral width of the signal and L is the length of the fiber.

The maximum data rate which the fiber can support is $\approx 1/\tau$. It is then clear that for higher data rates, τ and therfore, dispersion should be as small as possible. Since, the intermodal dispersion is much larger than the intramodal dispersion, ($D_{\mathrm{inter}} >> D_{\mathrm{intra}}$), the single mode fiber has much larger bandwidth.

The single mode optical fiber due to its lowest dispersion is used for long distance high speed data communication. However, to make a fiber single mode, the V-number of the fiber should be ≤ 2.4. (The magic number 2.4 has its origin in the root of J_0 Bessel function. We may recall from Chapter 7 that the variation of the fields inside a cylindrical waveguide is given by Bessel functions.) Since, the V-number is proportional to $(n_1^2 - n_2^2)^{1/2}$, and the core radius a, the V-number can be kept below 2.4 by either reducing the difference between the refractive indices of the core and the cladding $(n_1 - n_2)$, or by reducing the core size. Due to manufacturing reasons, $(n_1 - n_2)$ cannot be made smaller than $\sim 10^{-3}$. The core radius, therefore, has to be reduced to few microns to achieve single mode propagation. In practice, single mode fiber typically has the core diameter 6–10 μm.

Within single mode operation, the V-number should be chosen so as to get the minimum possible dispersion. Since, the dispersion (intramodal for the single mode fiber) is proportional to $\partial^2\beta/\partial V^2$, the fiber should be designed to operate at V-number whose $\partial^2\beta/\partial V^2$, is minimum. From Fig. 11.2 we note that $\partial^2\beta/\partial V^2$, for the HE_{11} curve is lowest around V = 2.4. Most of the single mode fibers, therfore, operate at V-numbers which are close to but less than 2.4.

A variety of optical fibers with complex refractive index profiles have been designed by solving the wave equation numerically. A thorough knowledge of wave propagation inside a dielectric waveguide is essential in designing proper optical fibers.

11.1.2 Lasers

Laser is a device which can coherently amplify light. Depending upon the material used, the light is amplified in specific frequency bands. So, in principle, a laser is an optical amplifier of finite bandwidth. The main constituents of a laser as an amplifier are, a medium with desired energy band structure, an excitation source called the 'pump', and input-output coupling devices. The pump supplies the energy to the material which is then transferred to the signal and the signal is amplified.

The laser is invariably used as a coherent source of light. However, it should be noted that a principle, laser is an amplifier and not a light generating device. Since, any amplifier can be converted into an oscillator with proper frequency selective feedback, a laser, which intrinsically is an amplifier, can also be converted into a coherent source of light. The crucial component of a laser (as an optical oscillator) is a frequency feedback device. Generally, this feedback is provided by creating a resonant cavity around the amplifying medium. A schematic of laser is shown in Fig.11.3.

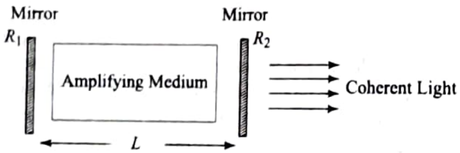


Fig. 11.3 *Schematic of a laser.*

The light waves get reflected by the mirrors and get amplified. Of course only those frequencies which see a regenrative feed back from the mirrors get sustained amplification. For sustained oscillation of a frequency the round trip gain between the mirrors should be unity and the round trip phase must be multiple of 2π. If the gain of the amplifying medium is G, and the reflection coefficients of the mirrors are R_1 and R_2 respectively, the unity round trip gain needs

$$R_1 R_2 e^{2GL} = 1 \tag{11.4}$$

If the refractive index of the amplifying medium is n, the phase constant of a light wave of frequency ω, is $\beta = \omega n/c$. The wave undergoes a round trip phase change of $2\,\beta L = 2\omega n L/c$. The phase condition then needs

$$\frac{2\omega n L}{c} = 2m\pi \quad m \text{ is an integer} \tag{11.5}$$

$$\Rightarrow \qquad \omega = \frac{m\pi}{nL} \tag{11.6}$$

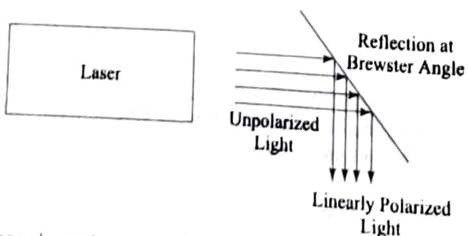


Fig. 11.4 *Changing polarization of light from random to linear using the Brewster angle.*

Only those frequencies which satisfy Eqns (11.5) and (11.6), have sustained oscillations. An extremly narrow band optical signal can be obtained by proper choice of the length of the cavity.

The light generated by the laser shown in Fig. 11.3, has in general random polarization. The Brewster angle concept is used for obtaining a linearly polarized light. The light coming form the laser is reflected by a dielectric slab at the Brewster angle. the reflected light (as mentioned in Chapter 5) has only perpendicular polarization (Fig.11.4).

11.2 SATELLITE COMMUNICATION

Satellite communication has been one of the important modes of long distance communication. The satellite communication offers many advantages like, large bandwidth, mobility, dynamic assignment of resources, etc. A large number of low orbit satellites (few hundred kms height above the earth surface) and geostationary satellites have been launched in last few decades.

In satellite communication the radio signals are transmitted from a transmitter on the earth called the 'earth station', towards the satellite. The satellite receives the signal, changes its frequency and retransmits it towards the earth. The satellite generally offers point-to-multipoint, i.e. broadcast transmission. Generally, microwave frequencies are used for satellite communication.

The design of a satellite communication link exploits almost all aspects of electromagnetic waves discussed in this book. The satellite link design is an important task as it gives the estimate of the power that the satellite would be able to receive from the earth station and the power which a receiving station would receive on the ground. The link design takes into consideration several aspects like, absorption by space between the satellite and the earth, changes in the wave characteristics during propagation, design of the antenna system, etc. Since, the power at the satellite is limited, a highly directional and efficient antenna system is a crucial component of a satellite link.

Figure 11.5 shows schematic of a satellite link. Let the satellite transmit P_t power with an antenna of directive gain G_t. Let the distance between the earth and the satellite be 'd'. The product $P_t G_t$ is called the 'effective isotropic radiated power' (EIRP) of the satellite. This quantity essentially gives the hypothetical power which the satellite would have to transmit with an isotropic radiator to get the same power density on the earth.

The power density on the earth due to satellite is

$$W = \frac{P_t G_t}{4\pi d^2} \tag{11.7}$$

Let the receiving antenna be having the directive gain of G_R and the effective aperture A_R. ($G_R = 4\pi A_R/\lambda^2$).

The power received by the receiving station is

$$P_R = W \cdot A_R \cdot \eta_p \tag{11.8}$$

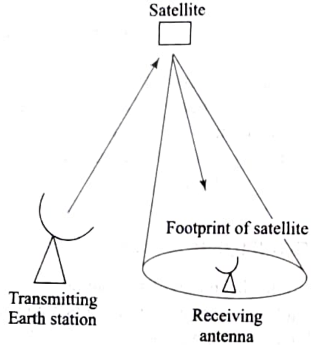


Fig. 11.5 *Schematic of a satellite link.*

where η_p is the polarization mismatch efficiency between the incoming wave and the receiving antenna. If the polarization of the incoming wave is fully matched with the polarization of the receiving antenna, $\eta_p = 1$. If the incoming wave is linearly polarized and the receiving antenna has circular polarization, $\eta_p = 1/2$, and so on.

Substituting for A_R in Eqn (11.8) we get

$$P_R = \frac{P_t G_t}{4\pi d^2} \cdot \frac{\lambda^2 G_R}{4\pi} \cdot \eta_p$$

$$= (P_t G_t) \cdot \left(\frac{\lambda}{4\pi d}\right)^2 \cdot G_R \cdot \eta_p$$

$$= (\text{EIRP}) \cdot (\text{path-loss})^{-1} \cdot G_R \cdot \eta_p \qquad (11.9)$$

The path-loss $(4\pi d/\lambda)^2$ is independent of the transmitting and the receiving systems. The path-loss is inversely proportional to the square of the wavelength and consequently shorter wavelengths (higher frequencies) suffer more path-loss. The path-loss for a geostationary satellite is about 200 dB at 6 GHz.

Typical satellites have EIRP of 50–60 dBW (0 dBW = 1 W), and the receiving antenna for small receving stations have gain of 10–20 dB. The received power, therefore, is in the range of –130 to –140 dBW, for systems with perfect polarization matching.

11.3 TELEPHONE NETWORKS

All transmission in the telephone network is through pair of wires. Virtually, all subscriber loops in the telephone networks are implemented with a single pair of wires. The single pair is used for bidirectional transmission. That is, if the users at both ends of the line talk simultaneously, their signals are superimposed on the wire pair and can be heard at the other end of the line.

Transmission over long distances like between two switching offices, is best implemented with two unidirectional transmissions on different pairs of wires. This is due to the fact that the long distance transmission invariably requires amplifiers which are unidirectional. The long distance transmission, therefore, needs a four-wire system. At some point in long distance connection, hence, it is essential to convert from a two-wire-to-four-wire system. The conversion device is called a 'hybrid'. Figure 11.6 shows a two-wire-to-four-wire and back four-wire-to-two-wire connection using two hybrids.

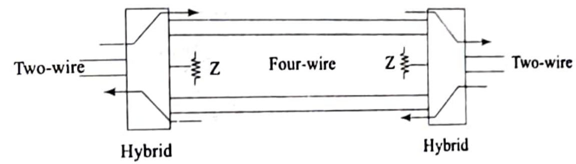


Fig. 11.6 *Two-wire-to-four-wire connections in telephone lines using hybrids.*

In earlier days, the hybrid circuits were implemented with interconnected transformers. In modern days, however, electronic hyrbrids have been developed. The impedance matching network Z is adjusted to get perfect isolation between signals travelling in opposite directions. However, in a switching environment the two wire line has variable impedance and therefore gives impedance mismatch. This impedance mismatch causes reflections on the line resulting in an echo. The amplitude of the echo depends upon the degree of impedance mismatch. Minimization of echoes using adaptive signal processing techniques is an interesting subject in telephone systems.

One more aspect of the telephone lines is the distortion of the voice signals. Invariably, the practical telephone lines have substantial loss in the conductors and relatively less loss in the dielectric separating the conductors. For distortionless transmission, the line must satisfy the condition

$$\frac{R}{L} = \frac{G}{C} \qquad (11.10)$$

where R, L, G, C are the primary constants of the line defined in Chapter 2. Since, for practical lines $G \ll R$, L has to be increased to achieve the condition given by Eqn (11.10). The loading coils were introduced to artificially increase the inductance of the line so as to realize Eqn (11.10) over the length of the line.

11.4 RADAR

The word 'Radar' is an acronym for 'radio detection and ranging'. Radar is an instrument used to detect and locate objects like air crafts, ships, etc. using radio waves. Radar consists of a high power radio transmitter and an extremely sensitive receiver. When the transmitted radio signal strikes the object, some of the energy is reflected back. A highly directional antenna receives the reflected signal and the receiver finds the amplitude and delay of the receiving signal. These two quantities then provide an estimate of the size and distance of the object from the radar (Fig.11.7). The size of the object is characterized by a parameter called the radar cross-section. The radar cross-section is a complex function of the reflectivity of the object and its physical shape and size. Metallic objects have higher radar cross-sections compared to the dielectric objects. The power reflected by an object is equal to the product of the power density of the incident wave and the radar cross-section.

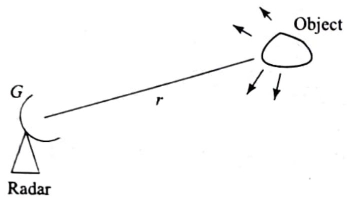


Fig. 11.7 *Basic radar geometry.*

Let the radar be transmitting P_t power with an antenna of directive gain G. Usually the same antenna is used for transmission of high power signal and receiption of the reflected signal. The effective aperture of the antenna while in receiving mode is $A_e (= G\lambda^2/4\pi)$. The power density of the transmitted signal at the object is

$$W = \frac{P_t G}{4\pi r^2} \tag{11.11}$$

where r is the distance of the object from the radar.

The power reflected by the object is

$$P_{ref} = P\sigma = \frac{P_t G\sigma}{4\pi r^2} \tag{11.12}$$

where σ is the radar cross-section of the object. Assuming that the reflected power is uniformly scattered in all directions, the power density of the reflected power at the radar is

$$W_r = \frac{P_{ref}}{4\pi r^2} = \frac{P_t G\sigma}{(4\pi r^2)^2} \tag{11.13}$$

The power received by the radar antenna is

$$P_{rec} = W_r A_e = \frac{P_t G\sigma}{(4\pi r^2)} = \frac{P_t\sigma}{4\pi}\left(\frac{G\lambda}{4\pi r^2}\right)^2 \tag{11.14}$$

For detection of the object, P_{rec} should be greater than the minimum detactable power of the receiver. Since, $r \gg \lambda$, G has to be very large for keeping P_t within acceptable limits. The radars, therefore, use large highly directional antennas like parabolic dishes.

Since, a radar transmits very high power of few kW, the energy transporting medium from the transmitter to the antenna is a metallic waveguide. Both rectangular and circular metallic waveguides find active use in radars.

Radars are used for military applications and navigational purposes. Many radar techniques developed for military applications, now find use in civilian applications. These include weather monitoring, geological search, air traffic control, etc. Radars are also used in remote sensing and planetary exploration.

On one side the radar research focuses on detecting objects with highest resolution and sensitivity, and on the other, active research is carried out in designing air crafts and ships which have extremely low radar cross sections. Special electromagnetic techniques are used to develop targets which are practically invisible to radar. Of course, it is not possible to design an object which will be invisible at all frequencies. However, the effort is made to make the target invisible over as large a frequency band as possible. The invisible air craft is used as a spy over the enemy territory.

11.5 REMOTE SENSING

In radio remote sensing, the microwave frequencies are used to probe the terrain. A radar is mounted on a moving platform like an aeroplane or a satellite above the earth's surface. Since, the reflectivity of the earth is location and frequency dependent, proper frequncies are to be selected for effective imaging of a terrain. There are two types of radars used in obtaining high quality radio images of the earth's surface.

(i) side looking radar (SLR)
(ii) synthetic aperture radar (SAR).

In a SLR, the reflection from a region illuminated by the radar beam gives the average reflectivity of the region. As the vehicle carrying the radar moves, the beam spot on the terrain also moves and we obtain reflectivity profile of the terrain along a strip as shown in Fig. 11.8.

By multiple scans of the adjacent strips one can then build an image of the terrain. Since, we get the average reflectivity of the terrain over the beam of the antenna, the angular resolution of the SLR images is $\sim \lambda/D$, where λ is the wavelength of the radar and D is the diameter of the antenna. If the height of the radar above the terrain is R, the linear resolution of the SLR image is $R\lambda/D$. For obtaining high resolution, large antennas are to be used. Also the radar vehicle should be at low altitude (small R) to get better linear resolution in the images. The SLR's generally have resolution of few hundred meters.

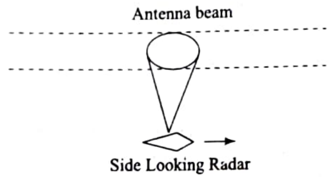


Fig. 11.8 *Side looking radar (SLR).*

In a SAR (Fig. 11.9), the reflected signals for the entire flight of the radar are stored and then coherently processed to form a synthesized antenna beam which is much narrower compared to the beam of the radar antenna. Let the SAR antenna be of diameter D. The beam of the antenna, θ is then $\approx \lambda/D$ (rad). A particular point on the terrain is visible to the antenna till it remains in the beam. If the height of the radar is R, a point on the terrain is visible to the antenna for the flight distance, $R\theta = R\lambda/D$. That is, effectively the synthesized beam corresponds to an aperture length $\approx R\lambda/D$. The synthesized beam width corresponding to this aperture length is $\approx \lambda/(R\lambda/D) \approx D/R$, and the linear resolution on the terrain is $R(D/R) \approx$ D.

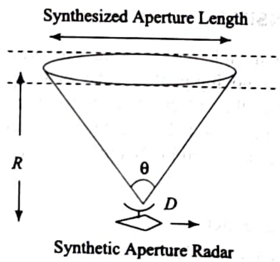


Fig. 11.9 *Synthetic aperture radar (SAR).*

In contrast to the SLR, the linear resolution of SAR is proportional to the size of the antenna. Smaller the antenna better is the resolution in the image. Also, the important thing to note is, the resolution is independent of the height of the vehicle, R. High altitude vehicle can be used for SAR imaging. The high quality SAR images have a resolution as fine as few meters.

A good eletromagnetic modelling of terrain is needed for interpretation of the remote sensing images. Scattering of electromagenetic waves through random media forms the basis of the modelling. Numerical electromagnetic techniques are employed to obtain accurate models of the earth's surface.

11.6 RADIO ASTRONOMY

In radio astronomy, one is interested in imaging the brightness distribution in the sky with as high an angular resolution and as high a sensitivity as possible. The angular resolution of a telescope is of the order of λ/D, where λ is the wavelength of observation and D is the diameter of the telescope. At optical wavelengths, we can obtain an angular resolution of few arc seconds with a telescope size of few meters. At radio frequencies, however, we need a telescope of the size of a few hundred meters to few kms to achieve the same angular resolution. Obviously, building a continuous aperture type of radio telescope of a few kms size is completely out of question. The radio astronomers, therefore, have devised a technique called the aperture synthesis technique using which one can synthesize a large aperture without actually building an aperture of that size. In aperture synthesis technique, instead of measuring the brightness distribution directly, one measures the spatial Fourier spectrum of the brightness distribution, that is, the spatial auto-correlation function (also called the visibility function) of the incoming radiation from the sky. The Fourier inversion of the visibility function then provides the brightness distribution in the sky.

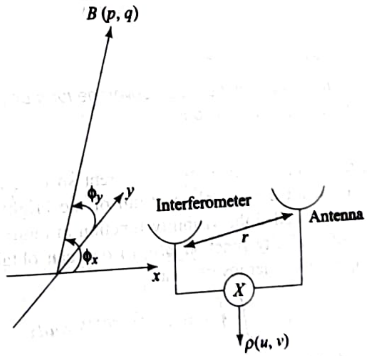


Fig. 11.10 *Basic radio interferometer geometry.*

Let us consider a brightness distribution B(p, q) where p and q are direction cosines of a particular direction (ϕ_x, ϕ_y) in the sky (see Fig. 11.10). Then, we have $p = \cos\phi_x$ and $q = \cos\phi_y$. Let us consider a pair of identical antennas forming a base line $\mathbf{r}(\equiv u\hat{\mathbf{x}} + v\hat{\mathbf{y}})$ in the spatial plane on the ground. Here, u and v denote the normalized projections (normalized with respect to the wavelength) of the $\mathbf{r}$ baseline along the x and y axes respectively, and $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the unit vectors along x and y directions. The voltage outputs of the two antennas are multiplied to obtain the correlation coefficient for the base line $\mathbf{r}$. The correlation, coefficient, due to a small brightness region of size dpdq in the direction (p, q) is then given by

$$d\rho(\mathbf{r}) = B(p, q)dpdqe^{-j2\pi(up+vq)} \qquad (11.15)$$

Now, assuming that radiation from different directions is uncorrelated, the correlation output due to entire brightness distribution in the sky is given as

$$\rho(\mathbf{r}) \equiv \rho(u, v) = \int\int_{sky} B(p, q)e^{-j2\pi(up+vq)}dpdq \qquad (11.16)$$

The integral in Eqn (11.16) is the two dimensional Fourier integral and $\rho(u, v)$, therefore, is the spatial Fourier component corresponding to the spatial frequency (u, v). The output of an antenna pair hence measures one spatial Fourier component of the brightness distribution.

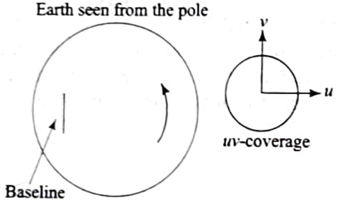


Fig. 11.11 *Earth's rotational synthesis. uv-coverage for a baseline parallel to the equator viewed from the pole.*

Now placing many antenna pairs with different spacing and different orientations, one can obtain the spatial spectrum of the brightness distribution. The spatial spectrum is called the visibility function in radio astronomy. From the measurement of visibility function $\rho(u, v)$ one can obtain the brightness distribution through the Fourier inversion as

$$B(p, q) = \int_u\int_v \rho(u, v)e^{j2\pi(up+vq)}dudv \qquad (11.17)$$

It may be emphasized here that, since the brightness distribution in the sky is stationary, the whole visibility function need not be measured simultaneously. Instead one can measure the visibility coefficients for different antenna baselines sequentially to build the visibility function.

The earth's rotation can be cleverly utilized to realize different interferometer baselines with just one pair of antennas. This is due to the fact that the length and the orientation of an interferometer baseline seen from a particular source in the sky changes as the earth rotates. For example, consider a baseline parallel to the equator (as shown in Fig. 11.11). When this base line is viewed form the celestial pole in the same hemisphere, the length of the baseline remains constant but its orientation changes over 360° during the day. The output of the interferometer from this baseline measured over a day then gives the visibility

function sampled along a ring of radius equal to the length of the baseline in the uv-plane (see Fig. 11.11). This is called the earth's rotational aperture synthesis. By having more interferometer pairs, the spatial Fourier plane can be covered with the earth's rotational synthesis.

The earth's rotational aperture synthesis has played an important role in progress in radio astronomy. Today, we can obtain radio images of the sky with a resolution which is orders of magnitude better than the resolution of the optical images.

11.7 ELECTROMAGNETIC COMPATIBILITY (EMC)

Various sources of electromagnetic emission like the spark plugs, lightning, electric relays, electric motors etc, generate electromagnetic waves that are rich in spectral contents. These waves cause interference or noise to electronic devices and communication receivers. In fact any fast transition in current generates wide range of frequencies. The modern computers switch signals at very fast rate and consequently produce electromagnetic interference.

The electromagnetic interference (EMI) causes malfunctioning of electronic systems and therefore, it is important to design electronic systems which minimally affect other systems. An electronic system which is able to function compatibly with other electronic systems and not produce or be susceptible to interference is called a 'electromagnetically compatible' with its environment. Electromagnetic compatiblity (EMC) is a very important aspect of electronic system design.

EMC is concerned with generation, transmission and receiption of electromagnetic energy. There are three ways to prevent EMI and to make the systems EM compatible.

1. Supress the emission at the source
2. Make the transmission path as inefficient as possible
3. Make receiver less susceptible to emission.

The first line of defense is to avoid fast switching of the signals. Even in digital systems, the signals should be switched smoothly so as to reduce the frequency content of the EMI. Reducing the high frequencies in the EMI inherently reduces the coupling of the interference.

The second line of defense would be to shield the EMI source from the receiver. The shielding techniques essentially fall in the domain of electromagnetic waves.

The term 'shield' usually refers to a metallic enclosure that completely encloses an electronic product or a portion of it. There are two purposes of a shield. The first, is to prevent the emission of the product from radiating outside the boundary of the product (see Fig. 11.12(a)). The second purpose of the shield is to prevent radiated emission external to the product from coupling to the electronics of the product (Fig. 11.12(b)). Conceptually, a shield is a barrier to

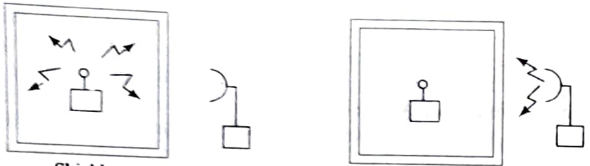


Fig. 11.12 *Electromagnetic shielding.*

the transmission of electromagnetic fields. The skin-effect is exploited to shield electronic components from each other. A good electromagnetic modelling is needed to predict the EMI and to make a system electromagnetically compatible.

11.8 SUMMARY

In this chapter, we have highlighted some of the prime applications of electromagnetic waves. The concepts of electromagnetic waves find many more applications in modern electronics and communication systems. As the speed of the digital circuits is increasing, the concepts of electromagnetic waves are becoming more relavant in system design. The concepts developed in this book, therefore, prepare students for handling high frequency electronics and communication systems.

Review Questions

11.1 What is an optical fiber?

11.2 What are the different types of optical fibers?

11.3 What is dispersion on optical fiber?

11.4 What is the difference between intra- and inter-modal dispersions? Which type of dispersion is present in single mode optical fiber?

11.5 What is V-number of an optical fiber?

11.6 What should be the V-number for single mode operation?

11.7 What is b-V diagram and what is its usage?

11.8 If the b-V diagram is linear, what will be the dispersion?

11.9 What are the conditions for laser oscillations?

11.10 How do we change a randomly polarized light into a lineraly polarized light?

11.11 What is EIRP of a satellite?

11.12 What is the 'path-loss' in a satellite link? How does it vary with frequency?

11.13 What is a radar? How does it work?

11.14 What is a radar cross-section? What does it depend upon?

11.15 What is SLR?

11.16 How does the resolution of a SLR image vary with frequency and antenna size?

11.17 What is SAR?

11.18 How does the SAR resolution vary with antenna size and frequency?

11.19 Why is SAR resolution independent of distance?

11.20 What is apatial correlation function and how is it related to the brightness distribution in the sky?

11.21 What is aperture synthesis technique?

11.22 What is earth's rotational aperture synthesis?

11.23 What is the angular resolution of a radio telescope?

11.24 What is EMC?

11.25 Why is EMC important in modern electronic system design?

Appendix **A**

Units and Symbols

Table A.1 MKSA (Rationalized) Units

Quantity	Typical Symbols	Units	Abbreviations
Length	l,r,R	Meter	m
Mass	m	kilogram	kg
Time	t	Second	s
Current	I,i	Ampere	A

Table A.2 Units for electromagnetic quantities

Quantity	Typ. Symbol	MKS units
Admittance	Y	Siemens (S), Mho (℧)
Attenuation	α	Neper/meter (Np/m)
conductance	G	Siemens (S), Mho (℧)
Capacitance	C	Farad (F)
Charge	Q,q	Coulomb(C)
Charge density (volume)	ρ	Coulomb/meter^3(C/m^3)
Charge density (surface)	ρ_s	Coulomb/meter^2(C/m^2)
Charge density (line)	ρ_l	Coulomb/meter (C/m)
Conductivity	σ	Mho/meter (℧/m)
Energy (work)	W	Joule (J)
Current	I	Ampere (A)
Current density(volume)	**J**	Ampere/meter^2(A/m^2)
Current density(surface)	$\mathbf{J}_s$	Amper/meter(A/m)
Electric flux density	**D**	Coulomb/meter^2(C/m^2)
Electric field intensity	**E**	Volt/meter (V/m)
Electrical Potential	V	Volt
Energy (work)	W	Joule(J)
Energy density	W	Joule/meter^3(J/m^3)
Electromotive force	V	Volt (V)
Force	**F**	Newton (N)
Frequency	f	Hertz (Hz) (/sec)
Impedance	Z	Ohm (Ω)
Inductance	L	Henry (H)
Magnetic field Intensity	**H**	Ampere/meter (A/m)
Magnetic flux	Φ	Weber (Wb)
Magnetic flux density	**B**	Weber/meter^2(Wb/m^2)
Magnetic vector potential	**A**	Weber/meter (Wb/m)
Magnetization	**M**	Ampere/meter (A/m)
Permeability	μ	Henry/meter (H/m)
Permittivity	ϵ	Farad/meter (F/m)
Phase	ϕ	radian
Phase constant	β	radian/meter (/m)
Power	P	watt (W) or Joule/second (J/s)
Propagation constant	γ	meter^{-1} (/m)
Radiation intensity	U	Watt/steradian (W/Str)
Reactance	X	Ohm (Ω)
Reluctance		Henry^{-1}(H^{-1})
Resistance	R	Ohm (Ω)
Susceptance	B	Siemens (S), Mho (℧)

Table A.3 List of Prefixes (Multipliers) Used With Units

Prefix	Symbol	Magnitude
Tera	T	10^{12}
Giga	G	10^{9}
Mega	M	10^{6}
Kilo	k	10^{3}
Hecto	h	10^{2}
Deca	da	10^{1}
Deci	d	10^{-1}
Centi	c	10^{-2}
Milli	m	10^{-3}
Micro	μ	10^{-6}
Nano	n	10^{-9}
pico	p	10^{-12}
Femto	m	10^{-15}
Atto	a	10^{-18}

Appendix **B**

Coordinate Systems

There are three coordinate systems generally used in the analysis of electromagnetic problems namely Cartesian, cylindrical and spherical coordinate systems. Depending upon the geometry of the space under investigation, a suitable coordinate system is employed. All coordinate systems uniquely represent a point in the three-dimensional space. All coordinate systems have a reference point called the origin and all the spatial locations are measured with respect to the origin. If the three-dimensional space is assumed to be like a rectangular box, the appropriate coordinate system is the Cartesian coordinate system. The coordinate axes are along the three edges of the rectangular box and are denoted by x, y, z as shown in Fig. B.1.

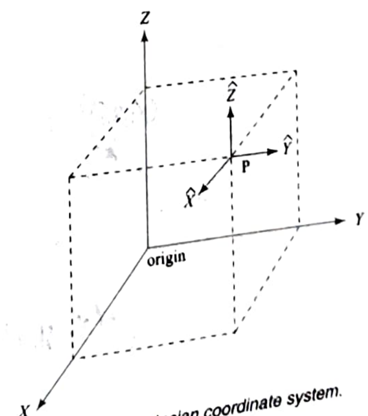


Fig. B.1 *Cartesian coordinate system.*

If the space is assumed to be like a cylinder the appropriate coordinate system is the cylindrical coordinate system (ρ, ϕ, z) as shown in Fig B.2. The three axes ρ, ϕ and z are marked at some point P in Fig. B.2. The z-direction is same at every point in the space but the physical orientation of ρ and ϕ axes changes with the location.

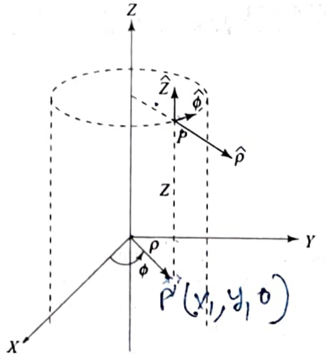


Fig. B.2 *Cylindrical coordinate system.*

If the three dimensional space is assumed to be like a sphere, the appropriate coordinate system is the spherical coordinate system (r, θ, ϕ) as shown in Fig B.3. Here, all the three axes r, θ, ϕ change their orientation with location.

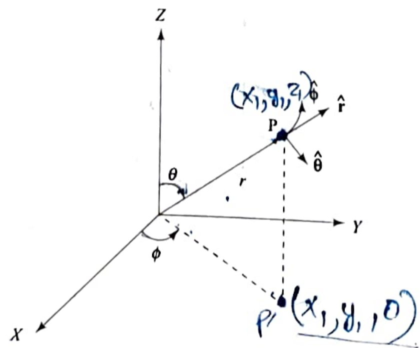


Fig. B.3 *Spherical coordinate system.*

In all coordinate systems the positive direction of any axis is that direction in which the quantity increases. In the cylindrical and spherical coordinate systems, for example, the directions radially outward from the vertical line and the origin respectively, indicate the positive direction of ρ-axis and r-axis uniquely. This type of uniqueness may not be true for all axes. For example in Fig. B.1 or Fig. B.2 the direction of z-axis can be taken positive upward or downward. There is no specific reason for drawing the directions the way they are written in Figs B.1 to B.3. Similarly, there is no specific reason for measuring angle ϕ in the anti-clock wise direction. By reversing the direction of the axes the nature of the particular coordinate system is not changed but the representation of a vector would change. It is, therefore, essential that we define a rule by which one can uniquely get the orientations of the coordinate axes with respect to each other. The rule of right hand (RH) screw or left hand (LH) screw is used for this purpose. Generally, the systems which follow the right hand rule are used (although there is no specific reason for it, and one could have developed the whole analysis using the left hand rule as well) for analysis.

In the RH-system if the fingers of your right hand are oriented in the direction of rotation from one axis to another adjacent axis in the right order, the thumb of your hand should point in the positive direction of the third axis. In Fig B.4 let us look at the cartesian coordinate system. If we align our fingers of right hand along the arrow, (rotation from x-axis to y-axis) the thumb will point upwards. The positive direction of z-axis is, therefore, correctly marked in Fig. B.4(a), whereas its direction is reversed in Fig. B.4(b).

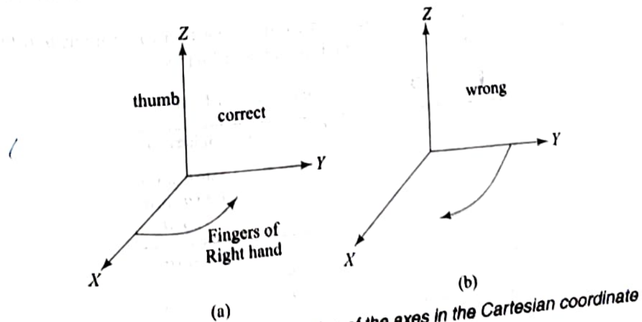


Fig. B.4 *Correct and wrong orientation of the axes in the Cartesian coordinate system.*

Similarly, the axes are correct marked in Fig. B.5(a) but they are wrongly marked in Fig. B.5(b). Every time, it is essential to verify the correctness of the direction of the coordinate axes and one should develop a habit of drawing correct orientation of the coordinate axes to avoid any error in further vector calculations.

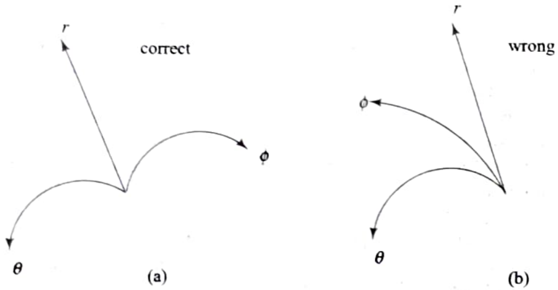


Fig. B.5 *Correct and wrong orientation of the axes in the spherical coordinate system.*

A vector **A** can be decomposed into the three components along the three axes in any coordinate system. Therefore, we can write:

$$\mathbf{A} = A_x\hat{\mathbf{x}} + A_y\hat{\mathbf{y}} + A_z\hat{\mathbf{z}}$$

or

$$\mathbf{A} = A_\rho\hat{\rho} + A_\phi\hat{\phi} + A_z\hat{\mathbf{z}}$$

or

$$\mathbf{A} = A_r\hat{r} + A_\theta\hat{\theta} + A_\phi\hat{\phi} \tag{B.1}$$

where $\hat{\mathbf{x}}$ represents the unit vector along the x-axis, $\hat{\mathbf{r}}$ represents the unit vector along the r-axis, and so on.

Table B.1 gives coordinate conversion from one system to another, and the relation between vector components in different coordinates.

Table B.1 Coordinate Conversion Table

=	Cartesian	Cylindrical	Spherical
x	x	$\rho\cos\phi$	$r\sin\theta\sin\phi$
y	y	$\rho\sin\phi$	$r\sin\theta\sin\theta$
z	z	z	$r\cos\theta$
ρ	$\sqrt{x^2+y^2}$	ρ	$r\sin\theta$
ϕ	$\tan^{-1}(\frac{y}{x})$	ϕ	ϕ
z	z	z	$r\cos\theta$
r	$\sqrt{x^2+y^2+z^2}$	$\frac{\rho}{\sin\theta}$	r
θ	$\cos^{-1}\left\{\frac{z}{\sqrt{x^2+y^2+z^2}}\right\}$	$\tan^{-1}(\frac{\rho}{z})$	θ
ϕ	$\tan^{-1}(\frac{y}{x})$	ϕ	ϕ

Table B.2 Relationships Between Vector Components in the Cartesian, Cylindrical, and Spherical Coordinate Systems

=	Cartesian	Cylindrical	Spherical
A_x	A_x	$A_\rho\cos\phi - A_\phi\sin\phi$	$A_r\sin\theta\cos\phi$ $+A_\theta\cos\theta\cos\phi$ $-A_\phi\sin\phi$
A_y	A_y	$A_\rho\sin\phi + A_\phi\cos\phi$	$A_r\sin\theta\sin\phi$ $+A_\theta\cos\theta\sin\phi$ $-A_\phi\cos\phi$
A_z	A_z	A_z	$A_r\cos\theta + A_\theta\sin\theta$
A_ρ	$A_x\cos\phi + A_y\sin\phi$	A_ρ	$A_r\sin\theta + A_\theta\cos\theta$
A_ϕ	$-A_x\sin\phi + A_y\cos\phi$	A_ϕ	A_ϕ
A_z	A_z	A_z	$A_r\cos\theta - A_\theta\sin\theta$
A_r	$A_x\sin\theta\cos\phi + A_y\sin\theta\sin\phi$ $+A_z\cos\theta$	$A_\rho\sin\theta + A_z\cos\theta$	A_r
A_θ	$A_x\cos\theta\cos\phi + A_y\cos\theta\sin\phi$ $-A_z\sin\theta$	$A_\rho\cos\theta - A_z\sin\theta$	A_θ
A_ϕ	$-A_x\sin\phi + A_y\cos\phi$	A_ϕ	A_ϕ

Appendix **C**

Vector Calculus

The three coordinate systems utilized in this book are rectangular coordinate system, circular cylindrical coordinate system and spherical coordinate system. These were defined in Appendix B.

We know that the intersection of two surfaces is a line; intersection of three surfaces is a point. Thus, the coordinates of a point may be given by stating three parameters, each of which defines a coordinate surface.

In a rectangular coordinate system, the three planes $x = a$, $y = b$ and $z = c$, intersect at a point designated by the coordinates (a,b,c) (Fig. B.1). The elements of length in the three coordinate directions are dx, dy, and dz, the elements of the area in three planes are $dxdy$, $dydz$, and $dzdx$, and the element of volume is $dxdydz$.

In a cylindrical coordinate system, the coordinate surfaces are, a set of circular cylinders ($\rho = \rho_1$), a set of planes all passing through the axis ($\phi = \phi_1$), and a set of planes normal to the axis ($z = z_1$). Coordinates of a particular point are given as (ρ_1, ϕ_1, z_1) (Fig. B.2). The ρ, ϕ and z are known respectively as the radius, azimuthal angle, and the distance along the axis. Elements of length in three coordinate directions are $d\rho$, $\rho d\phi$, and dz, elements of area are $\rho\, d\rho\, d\phi$, $\rho\, d\phi\, dz$ and $d\rho\, dz$, and the element of volume is $\rho d\rho d\phi dz$.

In a spherical coordinate system the surfaces are, a set of spheres (radius r from the origin = r_1), a set of cones about the z-axis ($\theta = \theta_1$), and a set of planes passing through the polar axis ($\phi = \phi_1$). The intersection of a sphere of radius, r_1, cone of half angle θ_1, and plane passing through the z-axis and making angle ϕ_1 with the x-axis, gives a point whose coordinates are said to be (r_1, θ_1,ϕ_1) (Fig. B.3). Elements of distance are dr, $rd\theta$, and $r \sin\theta d\phi$, elements of area are $rdr\, d\theta$, $r^2 \sin\theta d\theta d\phi$, and $rdr \sin\theta d\phi$, and the element of volume is $r^2 \sin\theta dr d\theta d\phi$.

C.1 GENERAL CURVILINEAR COORDINATES

The three coordinate systems utilized in this book are orthogonal coordinate systems in that the lines of intersection of the coordinate surfaces are at right angles to one another at any given point. It is possible to develop general expressions for divergence, curl, and other vector operations for such systems.

Suppose, that a point in space is thus defined in any orthogonal system by the coordinate surfaces q_1, q_2, q_3. These, then intersect at right angles and a set of three unit vectors, $\hat{\mathbf{a}}_1$, $\hat{\mathbf{a}}_2$, $\hat{\mathbf{a}}_3$ may be placed at this point. These should point in the direction of increasing coordinates. The three coordinates need not necessarily express directly a distance (consider, for example, the angles of spherical coordinates) so that the differential elements of distance must be expressed as

$$dl_1 = h_1 dq_1, dl_2 = h_2 dq_2, dl_3 = h_3 dq_3 \tag{C.1}$$

where h_1, h_2, h_3 are the scale factors and in general may be functions of all three coordinates q_1, q_2, q_3. Table C.1 gives h_1, h_2, h_3 for the three coordinate systems.

Table C.1

Coordinate system	h_1	h_2	h_3
Cartesian system	1	1	1
Cylindrical system	1	ρ	1
Spherical system	1	r	$r \sin\theta$

C.1.1 Scalar and Vector Products

Let there be two vectors $\mathbf{A}(\equiv A_1\hat{\mathbf{a}}_1 + A_2\hat{\mathbf{a}}_2 + A_3\hat{\mathbf{a}}_3)$ and $\mathbf{B}(\equiv B_1\hat{\mathbf{a}}_1 + B_2\hat{\mathbf{a}}_2 + B_3\hat{\mathbf{a}}_3)$. Then we get, the scalar or dot product as,

$$\mathbf{A} \cdot \mathbf{B} = A_1B_1 + A_2B_2 + A_3B_3, \tag{C.2}$$

and the vector or cross product as,

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \hat{\mathbf{a}}_1 & \hat{\mathbf{a}}_2 & \hat{\mathbf{a}}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} \tag{C.3}$$

C.1.2 Differential Operator ∇

A very important differential operator which is used in the analysis of the electromagnetics problems is the operator ∇, pronounced as 'del'. This operator is a vector operator and it has the dimension of length-inverse (L^{-1}). In all algebraic manipulations the ∇ can be treated as a normal vector. The operator ∇ can operate on a scalar like the multiplication of a scalar and a vector, or it can operate on a vector like the dot or cross products of the two vectors.

C.1.3 Gradient

The gradients of any scalar function ψ is a vector whose component in any direction is given by the spatial rate of change of ψ along that direction. Thus we have,

$$\nabla\psi = \hat{\mathbf{a}}_1 \frac{\partial\psi}{h_1\partial q_1} + \hat{\mathbf{a}}_2 \frac{\partial\psi}{h_2\partial q_2} + \hat{\mathbf{a}}_3 \frac{\partial\psi}{h_3\partial q_3} \tag{C.4}$$

Cartesian system: $(q_1, q_2, q_3) \equiv (x, y, z)$, $h_1 = 1$, $h_2 = 1$ and $h_3 = 1$ (see Table C.1).

$$\nabla\psi = \hat{\mathbf{x}}\frac{\partial\psi}{\partial x} + \hat{\mathbf{y}}\frac{\partial\psi}{\partial y} + \hat{\mathbf{z}}\frac{\partial\psi}{\partial z} \tag{C.5}$$

Cylindrical system: $(q_1, q_2, q_3) \equiv (\rho, \phi, z)$, $h_1 = 1$, $h_2 = r$ and $h_3 = 1$ (see Table C.1).

$$\nabla\psi = \hat{\rho}\frac{\partial\psi}{\partial\rho} + \hat{\phi}\frac{1}{\rho}\frac{\partial\psi}{\partial\phi} + \hat{\mathbf{z}}\frac{\partial\psi}{\partial z} \tag{C.6}$$

Spherical system: $(q_1, q_2, q_3) \equiv (r, \theta, \phi)$, $h_1 = 1$, $h_2 = r$ and $h_3 = r\sin\theta$ (see Table C.1).

$$\nabla\psi = \hat{\mathbf{r}}\frac{\partial\psi}{\partial r} + \hat{\theta}\frac{1}{r}\frac{\partial\psi}{\partial\theta} + \hat{\phi}\frac{1}{r\sin\theta}\frac{\partial\psi}{\partial\phi} \tag{C.7}$$

C.1.4 Divergence

When the operator ∇ operates on a vector $\mathbf{A}(\equiv A_1\hat{\mathbf{a}}_1 + A_2\hat{\mathbf{a}}_2 + A_3\hat{\mathbf{a}}_3)$ like a dot product the outcome is a scalar quantity and is called the divergence of vector $\mathbf{A}$. Physically, this represents net outward flux per unit volume of the quantity represented by the vector $\mathbf{A}$.

$$\nabla\cdot\mathbf{A} = \frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial q_1}(h_2h_3A_1) + \frac{\partial}{\partial q_2}(h_1h_3A_2) + \frac{\partial}{\partial q_3}(h_2h_1A_3)\right] \tag{C.8}$$

Cartesian system: $(q_1, q_2, q_3) \equiv (x, y, z)$, $h_1 = 1$, $h_2 = 1$ and $h_3 = 1$.

$$\nabla\cdot\mathbf{A} = \hat{\mathbf{x}}\frac{\partial A_x}{\partial x} + \hat{\mathbf{y}}\frac{\partial A_y}{\partial y} + \hat{\mathbf{z}}\frac{\partial A_z}{\partial z} \tag{C.9}$$

Cylindrical system: $(q_1, q_2, q_3) \equiv (\rho, \phi, z)$, $h_1 = 1$, $h_2 = r$ and $h_3 = 1$.

$$\nabla\cdot\mathbf{A} = \frac{1}{\rho}\left\{\frac{\partial}{\partial\rho}(\rho A_\rho) + \frac{\partial}{\partial\phi}A_\phi + \frac{\partial}{\partial z}(\rho A_z)\right\} \tag{C.10}$$

$$= \frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho A_\rho) + \frac{1}{\rho}\frac{\partial}{\partial\phi}A_\phi + \frac{\partial}{\partial z}(A_z) \tag{C.11}$$

Spherical system: $(q_1, q_2, q_3) \equiv (r, \theta, \phi)$, $h_1 = 1$, $h_2 = r$ and $h_3 = r\sin\theta$.

$$\nabla\cdot\mathbf{A} = \frac{1}{r^2\sin\theta}\left\{\frac{\partial}{\partial r}(r^2\sin\theta A_r) + \frac{\partial}{\partial\theta}(r\sin\theta A_\theta) + \frac{\partial}{\partial\phi}(rA_\phi)\right\} \tag{C.12}$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}(r^2A_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta A_\theta) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}A_\phi \tag{C.13}$$

C.1.5 Curl

When a vector function $\mathbf{A}$ is operated by ∇ like a cross product, the outcome is what is called the curl of $\mathbf{A}$. The $curl(\mathbf{A})$ is a vector quantity and it is a measure of rotation created by the vector $\mathbf{A}$ per unit area.

$$\nabla\times\mathbf{A} = \frac{1}{h_1h_2h_3}\begin{vmatrix} h_1\hat{\mathbf{a}}_1 & h_2\hat{\mathbf{a}}_2 & h_3\hat{\mathbf{a}}_3 \\ \partial/\partial q_1 & \partial/\partial q_2 & \partial/\partial q_3 \\ h_1A_1 & h_2A_2 & h_3A_3 \end{vmatrix} \tag{C.14}$$

Cartesian system: $(q_1, q_2, q_3) \equiv (x, y, z)$, $h_1 = 1$, $h_2 = 1$ and $h_3 = 1$.

$$\nabla\times\mathbf{A} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ A_x & A_y & A_z \end{vmatrix} \tag{C.15}$$

Cylindrical system: $(q_1, q_2, q_3) \equiv (\rho, \phi, z)$, $h_1 = 1$, $h_2 = \rho$ and $h_3 = 1$.

$$\nabla\times\mathbf{A} = \frac{1}{\rho}\begin{vmatrix} \hat{\rho} & \rho\hat{\phi} & \hat{\mathbf{z}} \\ \partial/\partial\rho & \partial/\partial\phi & \partial/\partial z \\ A_\rho & \rho A_\phi & A_z \end{vmatrix} \tag{C.16}$$

Spherical system: $(q_1, q_2, q_3) \equiv (r, \theta, \phi)$, $h_1 = 1$, $h_2 = r$ and $h_3 = r\sin\theta$.

$$\nabla\times\mathbf{A} = \frac{1}{r^2\sin\theta}\begin{vmatrix} \hat{\mathbf{r}} & r\hat{\theta} & r\sin\theta\hat{\phi} \\ \partial/\partial r & \partial/\partial\theta & \partial/\partial\phi \\ A_r & rA_\theta & r\sin\theta A_\phi \end{vmatrix} \tag{C.17}$$

C.1.6 Laplacian

The Laplacian operator ∇^2 is a scalar operator which can operate on a scalar or a vector. If it operates on a scalar, the outcome is a scalar quantity. Whereas, if it operates on a vector, the outcome is a vector quantity.

The Laplacian of a scalar, which is defined as the divergence of the gradient of the scalar may be found by combining Eqns (C.4) and (C.8):

$$\nabla^2\psi = \nabla \cdot \nabla\psi \tag{C.18}$$

$$= \frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2h_3}{h_1}\frac{\partial\psi}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_3h_1}{h_2}\frac{\partial\psi}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1h_2}{h_3}\frac{\partial\psi}{\partial q_3}\right)\right] \tag{C.19}$$

Cartesian system: $(q_1, q_2, q_3) \equiv (x, y, z)$, $h_1 = 1$, $h_2 = 1$ and $h_3 = 1$.

$$\nabla^2\psi = \nabla \cdot \nabla\psi \tag{C.20}$$

$$= \frac{\partial}{\partial x}\left(\frac{\partial\psi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial\psi}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{\partial\psi}{\partial z}\right) \tag{C.21}$$

$$= \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} \tag{C.22}$$

Cylindrical system: $(q_1, q_2, q_3) \equiv (\rho, \phi, z)$, $h_1 = 1$, $h_2 = \rho$ and $h_3 = 1$.

$$\nabla^2\psi = \frac{1}{\rho}\left\{\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) + \frac{\partial}{\partial\phi}\left(\frac{1}{\rho}\frac{\partial\psi}{\partial\phi}\right) + \frac{\partial}{\partial z}\left(\rho\frac{\partial\psi}{\partial z}\right)\right\} \tag{C.23}$$

$$= \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\psi}{\partial\phi^2} + \frac{\partial^2\psi}{\partial z^2} \tag{C.24}$$

Spherical system: $(q_1, q_2, q_3) \equiv (r, \theta, \phi)$, $h_1 = 1$, $h_2 = r$ and $h_3 = r\sin\theta$.

$$\nabla^2\psi = \frac{1}{r^2\sin\theta}\left\{\frac{\partial}{\partial r}\left(r^2\sin\theta\frac{\partial\psi}{\partial r}\right) + \frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{\partial}{\partial\phi}\left(\frac{1}{\sin\theta}\frac{\partial\psi}{\partial\phi}\right)\right\} \tag{C.25}$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2} \tag{C.26}$$

C.1.7 Laplacian of a Vector

For the Laplacian of a vector in a system of coordinates other than rectangular, it is convenient to use the vector identity

$$\nabla^2\mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla \times \nabla \times \mathbf{A} \tag{C.27}$$

Each of the operations on the right has been defined earlier.

C.2 VECTOR IDENTITIES

$$(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} = (\mathbf{B} \times \mathbf{C}) \cdot \mathbf{A} = (\mathbf{C} \times \mathbf{A}) \cdot \mathbf{B}$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}$$

$$\nabla \cdot (V\mathbf{A}) = \mathbf{A} \cdot \nabla V + V\nabla \cdot \mathbf{A}$$

$$\nabla \times (V\mathbf{A}) = \nabla V \times \mathbf{A} + V\nabla \times \mathbf{A}$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot \nabla \times \mathbf{A} - \mathbf{A} \cdot \nabla \times \mathbf{B}$$

$$\nabla(\mathbf{A} \cdot \mathbf{B}) = (\mathbf{A} \cdot \nabla)\mathbf{B} + (\mathbf{B} \cdot \nabla)\mathbf{A} + \mathbf{A} \times (\nabla \times \mathbf{B}) + \mathbf{B} \times (\nabla \times \mathbf{A})$$

$$\nabla \cdot \nabla \equiv \nabla^2$$

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0$$

$$\nabla \times (\nabla V) = 0$$

$$\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2\mathbf{A}$$

C.3 INTEGRAL THEOREMS

In vector calculus there are two very important theorems which are useful in converting line integrals to surface integrals and the surface integrals to the volume integrals. These theorems are called the Stokes theorem and the Divergence theorem.

C.3.1 Divergence Theorem

Consider a closed surface S in a vector field **A** as shown in Fig. C.1.

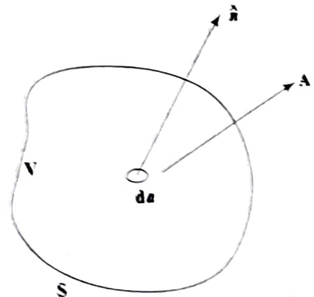


Fig. C.1 *Closed surface with outward normal.*

Let the volume enclosed by this surface be given by V. Then, according to the

Divergence theorem.

$$\oint_S \mathbf{A}\cdot d\mathbf{a} = \iiint_V (\nabla\cdot\mathbf{A})\,dv \qquad \text{(C.28)}$$

Here, the sign $\oint$ indicates the integral over a closed surface. The elemental area $d\mathbf{a} = da\,\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ represents outward unit normal to the surface.

C.3.2 Stokes Theorem

Consider a closed curve C enclosing an area S in the vector field **A** (Fig. C.2). The Stokes theorem then states that

$$\oint_C \mathbf{A}\cdot d\mathbf{l} = \iint_S (\nabla\times\mathbf{A})\cdot d\mathbf{a} \qquad \text{(C.29)}$$

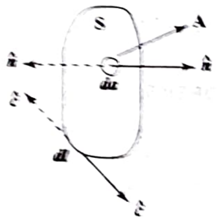


Fig. C.2 Closed loop.

The sign $\oint$ indicates line integral around a closed path and $\mathbf{dl} = dl\,\hat{\mathbf{c}}$ and $d\mathbf{a} = da\,\hat{\mathbf{n}}$. $\hat{\mathbf{c}}$ is the unit vector along the line segment dl and $\hat{\mathbf{n}}$ is the unit vector normal to the surface. However, it is not immediately clear which of two directions (indicated by thick and dotted arrows) are to be taken in Fig. C.2. In other words the question is, which are the correct positive directions for the vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{c}}$? This type of confusion was not there in Fig. C.1 as for a closed surface the outward direction has a unique meaning. The answer essentially lies in our choice of right handed coordinate system. The directions of $\hat{\mathbf{c}}$ and $\hat{\mathbf{n}}$ should be consistent with the right handed system. If the line contour is traced in the anticlockwise direction (thick arrow for $\hat{\mathbf{c}}$) then the direction of $\hat{\mathbf{n}}$ vector will be pointing outward the plane of the paper. One can, therefore, choose $\hat{\mathbf{c}}$ and $\hat{\mathbf{n}}$ given either by the thick arrows or by the dotted arrows. Thick arrow for one unit vector and dotted for the other will give erroneous answer.

Appendix **D**

Bessel Functions

Bessel and Modified Bessel functions have been used in Chapter 7. In working with electromagnetic problems series solutions to the differential equations are often required. These series solutions are denoted by the functions which have an order and an argument. The functions of different orders and their derivatives are related through recurrence relations. The recurrence relations come handy in algebraically simplifying expressions. Here, we give the recurrence relations for the Bessel functions.

D.1 BESSEL FUNCTIONS

Bessel functions are solutions of the following differential equation, which is known as the Bessel's equation

$$\frac{d^2\psi}{dz^2} + \frac{1}{z}\frac{d\psi}{dz} + \left(1 - \frac{\nu^2}{z^2}\right)\psi = 0 \qquad \text{(D.1)}$$

The *order* of the equation is given by the value of ν. In general, ν can be non-integral however here we take ν to be an integer.

Being a second-order differential equation, the Bessel's equation has two independent solutions. First solution of the equation is obtained in series form as

$$J_\nu(z) = \sum_{m=0}^{\infty} \frac{(-1)^m z^{\nu+2m}}{m!(m+\nu)!2^{\nu+2m}} \qquad \text{(D.2)}$$

Where $J_\nu(z)$ is Bessel function of the *first kind*, of order ν.

The second independent solution of the Bessel's equation is defined by

$$N_\nu(z) = \frac{J_\nu(z)\cos\nu\pi - J_{-\nu}(z)}{\sin\nu\pi} \qquad \text{(D.3)}$$

Where $N_\nu(z)$ is known as a Neumann function or, more commonly, as a Bessel function of *second kind*, of order ν. A complete solution of Eqn (D.1) is

$$\psi = AJ_\nu(z) + BN_\nu(z) \tag{D.4}$$

Bessel functions of the second kind become infinite at $z = 0$ and cannot be used to represent physical fields except in those problems in which the region $z = 0$ is excluded.

D.1.1 Asymptotic Form of Bessel Functions

For $z \ll 1$, the J and N functions are given approximately by the expressions

$$J_\nu(z) \approx \frac{z^\nu}{2^\nu \nu!} \tag{D.5}$$

$$N_\nu(z) \approx \frac{2^\nu(\nu-1)!}{\pi z^\nu} \tag{D.6}$$

and in particular

$$J_0(z) \approx 1 \tag{D.7}$$

$$N_0(z) \approx \frac{2}{\pi}(\ln z + C - \ln 2) \tag{D.8}$$

Where, C is constant. On the other hand for $z >> 1$ the asympotic expressions are

$$J_\nu(z) \approx \sqrt{\frac{2}{\pi z}} \cos\left(z - \frac{\nu\pi}{2} - \frac{\pi}{4}\right) \tag{D.9}$$

$$N_\nu(z) \approx \sqrt{\frac{2}{\pi z}} \sin\left(z - \frac{\nu\pi}{2} - \frac{\pi}{4}\right) \tag{D.10}$$

D.1.2 Differentiation and Integration of Bessel Functions

Using the series definition for $J_0(z)$, and differentiating term by term, we can show that

$$\frac{d}{dz}[J_0(z)] = -J_1(z) \tag{D.11}$$

Similarly, it can be shown that in general the following relations are true: In the expressions listed here $Z_\nu(z)$ may denote $J_\nu(z)$, or $N_\nu(z)$. Also, we denote

$$Z'_\nu(z) = \frac{d}{dz}Z_\nu \tag{D.12}$$

$$Z'_0(z) = -Z_1(z) \tag{D.13}$$

$$Z'_1(z) = Z_0(z) - \frac{1}{z}Z_1(z) \tag{D.14}$$

$$zZ'_\nu(z) = \nu Z_\nu(z) - zZ_{\nu+1}(z) \tag{D.15}$$

$$= zZ_{\nu-1}(z) - \nu Z_\nu(z) \tag{D.16}$$

$$\frac{d}{dz}[z^\nu Z_\nu(z)] = z^\nu Z_{\nu-1}(z) \tag{D.17}$$

$$\frac{d}{dz}[z^{-\nu} Z_\nu(z)] = -z^{-\nu} Z_{\nu+1}(z) \tag{D.18}$$

We also have a useful recurrence relation

$$2\nu Z_\nu(z) = z[Z_{\nu-1}(z) + Z_{\nu+1}(z)] \tag{D.19}$$

The integrals of the Bessel functions are

$$\int Z_1(z)dz = -Z_0(z) \tag{D.20}$$

$$\int z^\nu Z_{\nu-1}(z)dz = z^\nu Z_\nu(z) \tag{D.21}$$

$$\int z^{-\nu} Z_{\nu+1}(z)dz = -z^{-\nu} Z_\nu(z) \tag{D.22}$$

D.2 MODIFIED BESSEL FUNCTIONS

The modified Bessel's equation of order ν is

$$\frac{d^2\psi}{dz^2} + \frac{1}{z}\frac{d\psi}{dz} - (1 + \frac{\nu^2}{z^2})\psi = 0 \tag{D.23}$$

where ν in general, is any number. However, here we assume ν to be integer. The two series solutions to the equation are denoted by $I_\nu(z)$ and $K_\nu(z)$.

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{z^{\nu+2m}}{2^{\nu+2m} m!(\nu+m)!} \tag{D.24}$$

$$K_\nu(z) = \frac{2}{\cos\nu\pi}\left(\frac{\partial I_{-\nu}}{\partial\nu} - \frac{\partial I_\nu}{\partial\nu}\right) \tag{D.25}$$

D.2.1 Asymptotic Form of Modified Bessel Functions

For $z << 1$, the I and K functions are given approximately by

$$I_\nu(z) \approx \frac{z^\nu}{2^\nu \nu!} \tag{D.26}$$

$$K_\nu(z) \approx \frac{2^{\nu-1}(\nu-1)!}{z^\nu} \tag{D.27}$$