



“ शिक्षा मानव को बन्धनों से मुक्त करती है और आज के युग में तो यह लोकतंत्र की भावना का आधार भी है। जन्म तथा अन्य कारणों से उत्पन्न जाति एवं वर्गीय विषमताओं को दूर करते हुए मनुष्य को इन सबसे ऊपर उठाती है।”

–इन्दिरा गांधी

VOLUME I

BLOCK 1: WAVES - AN INTRODUCTION

BLOCK 2: INTERFERENCE

“ Education is a liberating force, and in our age it is also a democratising force, cutting across the barriers of caste and class, smoothing out inequalities imposed by birth and other circumstances.”

– Indira Gandhi

Block

1

WAVES – AN INTRODUCTION

UNIT 1

Wave Motion

9

UNIT 2

Superposition of Waves

35

UNIT 3

Acoustic Wave

65

UNIT 4

Electromagnetic Wave

101

UNIT 5

Polarisation of Light

119

Course Design Committee

Prof. A. K. Ghatak, *Retd.*
IIT Delhi, New Delhi

Prof. Suresh Garg, *Retd.*
School of Sciences
IGNOU, New Delhi
Vice Chancellor,
Usha Martin University

Prof. R.M. Mehra, *Retd.*
Dept. of Electronics,
South Campus,
University of Delhi, Delhi

Dr. Ashok Goyal, *Retd.*
Dept. of Physics, Hansraj College
University of Delhi, Delhi

Dr. Parthasarathy
Dept. of Physics,
Maharaja Agrasen College,
University of Delhi, Delhi

Prof. M.S. Nathawat
Former Director,
School of Sciences,
IGNOU, New Delhi

Prof. Vijayshri
School of Sciences
IGNOU, New Delhi

Prof. Sudip Ranjan Jha
School of Sciences
IGNOU, New Delhi

Prof. Shubha Gokhale
School of Sciences
IGNOU, New Delhi

Dr. Sanjay Gupta
School of Sciences
IGNOU, New Delhi

Dr. Subhalakshmi Lamba
School of Sciences
IGNOU, New Delhi

Block Preparation Team

Prof. S.R. Jha (Units 1-5)
School of Sciences
IGNOU, New Delhi

Units of this block are based on the Oscillations and Waves (BPHE-102) and Optics (PHE-09) courses of the earlier B.Sc. programme of IGNOU.

Course Coordinators: Prof. S.R. Jha, Prof. Sanjay Gupta

Block Production Team

Mr. Rajiv Girdhar
Assistant Registrar
MPDD, IGNOU, New Delhi

Mr. Hemant Kumar Parida
Section Officer
MPDD, IGNOU, New Delhi

Acknowledgement: Shri Gopal Krishan Arora, EDP, SOS for CRC preparation.

October, 2021

© Indira Gandhi National Open University, 2021

ISBN:

Disclaimer: Any materials adapted from web-based resources in this module are being used for educational purposes only and not for commercial purposes.

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the Copyright holder.

Further information on the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110 068 or the official website of IGNOU at www.ignou.ac.in.

Printed and published on behalf of Indira Gandhi National Open University, New Delhi by the Registrar MPDD, IGNOU, New Delhi.

Printed at: Gita Offset Printers Pvt. Ltd., C-90, Okhla Industrial Area, Phase-I, New Delhi-20

CONTENTS

Block and Unit Titles	1
Credit page	2
Contents	3
WAVES AND OPTICS: COURSE INTRODUCTION	5
BLOCK 1: Waves – An Introduction	7
<u>Unit 1 Wave Motion</u>	<u>9</u>
1.1 Introduction	10
1.2 Waves on a String	11
1.2.1 Representation of Wave Motion	13
1.2.2 Relation between Wave Velocity, Frequency and Wavelength	15
1.3 Mathematical Description of Wave Motion	16
1.3.1 Phase of a Wave and Phase Difference	21
1.3.2 Plane Waves and Spherical Waves	22
1.4 Energy Transported by Waves	23
1.5 Intensity of Waves	27
1.6 Summary	29
1.7 Terminal Questions	30
1.8 Solutions and Answers	31
<u>Unit 2 Superposition of Waves</u>	<u>35</u>
2.1 Introduction	36
2.2 Principle of Superposition of Waves	37
2.3 Standing Waves	43
2.3.1 Standing Waves on a String	44
2.4 Beats	53
2.5 Wave Groups and Group Velocity	56
2.6 Summary	60
2.7 Terminal Questions	61
2.8 Solutions and Answers	62
<u>Unit 3 Acoustic Wave</u>	<u>65</u>
3.1 Introduction	66
3.2 Production of Sound Waves	67
3.2.1 Graphical and Mathematical Representation of Sound Waves	69
3.2.2 Sound Waves in Different Media	74
3.2.3 Forced Vibrations and Resonance	77
3.2.4 Musical Instruments	81

3.3 Intensity of Sound	84
3.3.1 Intensity Level: Decibel	84
3.4 Musical Notes and Scale	86
3.4.1 Musical Scales	87
3.5 Acoustics of Buildings	90
3.6 Summary	96
3.7 Terminal Questions	98
3.8 Solutions and Answers	98
<u>Unit 4 Electromagnetic Wave</u>	<u>101</u>
4.1 Introduction	102
4.2 Wave Equation for Electromagnetic Waves	103
4.2.1 Transverse Nature of Electromagnetic Waves	107
4.3 Huygens Principle	111
4.3.1 Wavefronts and their Properties	111
4.4 Summary	113
4.5 Terminal Questions	114
4.6 Solutions and Answers	115
<u>Unit 5 Polarisation of Light</u>	<u>119</u>
5.1 Introduction	120
5.2 What is Polarisation?	121
5.3 Simple States of Polarised Light	123
5.3.1 Linear Polarisation	124
5.3.2 Circular Polarisation	126
5.3.3 Elliptical Polarisation	128
5.4 Principles of Producing Linearly Polarised Light	130
5.4.1 Ideal Polariser: Malus's Law	130
5.4.2 Polarisation by Reflection: Brewster's Law	131
5.4.3 Polarisation by Double Refraction	133
5.4.4 Selective Absorption: Dichroism	137
5.5 Wave Plates: Circular and Elliptical Polarisers	138
5.6 Summary	139
5.7 Terminal Questions	140
5.8 Solutions and Answers	141
Further Readings	223
Table of Physical Constants	224
List of Blocks and Units: BPHCT-137	225
Syllabus: Waves and Optics (BPHCT-137)	226

WAVES AND OPTICS: COURSE INTRODUCTION

Light enables us to see and appreciate the wonders of nature and its creations. Light is so much part of our daily lives that we hardly pay any attention to it. It is almost like air! It's a given. The Earth has been receiving light from the Sun ever since; it is our most important source of energy. Sunlight warms us, causes weather patterns, allows plants to manufacture oxygen and our food from carbon dioxide and water, etc. In a way, light sustains life on our planet. The study of interaction of light with matter constitutes what we call optics.

Though we see objects illuminated by light, we generally cannot see light! Thus, you may wonder as to what light is? Well, we know that light is a form of energy. And, we know that energy can be transferred from one place to another either by moving bodies/particles or by travelling waves. So, the question is: Whether light is made of particles or waves? The evolution of our understanding of the nature of light is perhaps one of the most interesting accounts in the history of science. Since the beginning of modern science in the 16th and 17th century, light has been pictured either as particles or waves. The development in our understanding of the nature of light and its interaction with matter emerged as a result of the fundamental work done by scientists of eminence such as Galileo, Abbe, Newton, Huygens, Young, Fresnel, Fraunhofer, Grimaldi, Arago and Bartholinus. In India, Sir J.C. Bose and Sir C.V. Raman made significant contributions in the study of electromagnetic waves and light.

The two competing theories namely, particle theory and wave theory of light tried to explain the phenomena associated with light such as rectilinear propagation, reflection and refraction. The particle theory (also called corpuscular theory), supported by the great scientist Newton, could explain some of these phenomena. At about the same time, Huygens was effectively extending the wave theory by giving satisfactory explanations of the phenomena of reflection and refraction of light. Newton's high stature in the field of science, in a way, slowed down the progress of wave theory of light.

In the early 19th century, the wave theory got a renewed push due to the experiments performed by Young on interference of light. The series of experiments carried out by Young and his contemporary, Fresnel decisively supported the wave nature of light. The wave theory of light got a further push and a sound theoretical pedestal by Maxwell's electromagnetic theory which established that light is an electromagnetic wave.

The birth of quantum mechanics at the beginning of the 20th century further added to our understanding about the nature of light through the concepts like photon and wave-particle duality. Thus, the modern understanding of the nature of light is that it behaved like waves in its propagation and in phenomena like interference and diffraction and it could also behave as particles as in the photoelectric effect.

While studying this course, you will realise that there is an explosive growth of this subject due to the realisation of some well known physics principles for technological applications. That is why optics occupies a prominent place in present day pure and applied sciences.

This 4-credit course on **Waves and Optics** is presented in the following four blocks:

1. Waves – An Introduction
2. Interference
3. Diffraction
4. Lasers and Optical Fibre

In Block 1 entitled **Waves – An Introduction**, you will study the basic concepts related to wave motion. The understanding about formation and propagation of mechanical waves on a string will give you a background to appreciate the wave nature of light. We have also discussed the principle of superposition of waves which forms the basis of explaining interference and diffraction of light. Though light is a wave phenomenon, it is electromagnetic in nature. Thus, light waves are different from mechanical waves like waves on a string or sound waves. You will study in this block how Maxwell's equations and the consequent wave equations for electric and magnetic fields lead to the conclusion that light is an electromagnetic wave. The phenomenon of polarisation of light, which is an experimental evidence for the transverse nature of light waves, has also been discussed in this block.

In Block 2 entitled **Interference**, you will study how superposition of light from two coherent sources gives rise to observable interference pattern in the form of bright and dark fringes. You will learn how coherent sources can be obtained using the methods of division of wavefront and division of amplitude of the light waves. The phenomenon of interference of light has been put to many useful applications in the form of instrument called optical interferometer. We have discussed the working principle of Michelson interferometer and some of its important applications.

In Block 3 entitled **Diffraction**, you will study Fresnel and Fraunhofer diffraction. The phenomenon of diffraction, which refers to bending of light around the obstacles was first explained by Fresnel on the basis of wave theory and the related concept of half-period element. According to him, in diffraction phenomenon, interference takes place between secondary wavelets from different parts of the same wavefront. You will study Fresnel's explanation of diffraction patterns for some simple obstacles. You will study Fraunhofer diffraction due to a single slit and a double slit and learn how the intensity distribution in these diffraction patterns can be understood on the basis of the wave theory. You will also learn how the analysis of double slit diffraction can be extended to explain the diffraction pattern of a grating – an optical component comprising a large number of slits.

In Block 4 entitled **Lasers and Optical Fibre**, you will learn the working principle of lasers and some of their applications particularly in the field of optical communication. The lasers owe their invention to the quantum theory according to which, light energy consists of minute packets called quanta. You will understand the mechanism of laser operation in terms of concepts like spontaneous and stimulated emission of radiation and population inversion. You will also learn about the characteristic properties of laser light such as high degree of coherence, monochromaticity and directionality. These properties of laser light have been put to use in the fields of basic research as well as technological applications. In fact, the advent of lasers has revolutionised the field of communication technology. The communication – transmission of speech, data, etc. – at optical frequencies is highly reliable and efficient compared to radio and microwave communication. However, for efficient transmission of data at optical frequencies, special type of transmission medium, called optical fibre is required. You will learn in this block about the construction, light transmission mechanism and important characteristics of optical fibre.

We wish you success.

BLOCK 1: WAVES – AN INTRODUCTION

The main focus of the first block of this course is to familiarise you with waves. The purpose of discussing waves in a course which is primarily on optics should not surprise you because light is also a wave phenomenon. The mechanical waves such as waves on a string and electromagnetic waves, both are characterised by some common features such as amplitude, wavelength, frequency and speed. Therefore, a basic understanding of waves is necessary to understand the phenomena associated with light such as interference, diffraction and polarisation.

In **Unit 1** entitled **Wave Motion**, we discuss how waves are formed and how they propagate. For this, we consider transverse waves on a string. You will learn how to represent wave motion graphically as well as mathematically. An important reason to study wave motion is that the waves transport energy. In this unit, you will learn how to compute energy carried by a travelling wave of given amplitude and frequency.

Unit 2 entitled **Superposition of Waves** deals with the superposition principle for waves and its applications. The understanding of the principle of superposition of waves will enable you to analyse interference and diffraction of light, which arise due to simultaneous presence of two or more than two waves at a given point in space and time. Also, you will learn how standing waves are formed on a string when two waves travelling on it in opposite directions superpose.

In **Unit 3** entitled **Acoustic Wave**, you will learn about sound waves. Sound waves are mechanical waves produced by vibrating objects. You will learn what role forced oscillation and resonance play in production and propagation of sound. Human beings can hear sound of a rather wide frequency range: 20 Hz to 20 kHz. In addition, for being heard by human being, the sound wave must have intensity above a threshold value. In view of this, the intensity of sound is expressed on a relative scale called decibel. You will learn about decibel scale in this unit. Also, you will learn the physics behind musical sound and musical scales.

Unit 4 entitled **Electromagnetic Wave** deals with electromagnetic nature of light. You will learn to obtain wave equations for electric and magnetic fields using Maxwell's equations and discover that light is an electromagnetic wave and it is transverse in nature. Though Maxwell gave a sound theoretical basis to the wave model of light, it was proposed by Huygens much earlier on the basis of concepts like wavefront and spherical secondary wavelets. In this unit, you will also learn Huygens principle which is at the core of the wave theory of light.

Unit 5, the last unit of this block, deals with **Polarisation of Light**. Polarisation is an experimental evidence for transverse nature of light (electromagnetic wave). You will learn what is polarisation and why polarisation is a unique characteristic of transverse waves. Polarisation of waves is understood on the basis of the superposition principle for waves. You will learn the conditions under which superposition of waves gives rise to linearly, circularly and elliptically polarised lights. We have also discussed Malus's law and Brewster's law which, respectively, deals with ideal polariser and polarisation due to reflection. You will also study about the phenomena of double refraction and selective absorption which are used for producing linearly polarised light.

We wish you success.



UNIT 1

WAVE MOTION

One of the attractions of sea shore is to see waves. Waves coming from far off distances spread at the shore. Can you imagine how much energy is stored in nature in the form of waves? (Source of picture: wikimedia.org)

Structure

- | | | | |
|-----|--|-----|-----------------------------|
| 1.1 | Introduction | 1.4 | Energy Transported by Waves |
| | Expected Learning Outcomes | 1.5 | Intensity of Waves |
| 1.2 | Waves on a String | 1.6 | Summary |
| | Representation of Wave Motion | 1.7 | Terminal Questions |
| | Relation between Wave Velocity, Frequency and Wavelength | 1.8 | Solutions and Answers |
| 1.3 | Mathematical Description of Wave Motion | | |
| | Phase of a Wave and Phase Difference | | |
| | Plane Waves and Spherical Waves | | |

STUDY GUIDE

This is the first unit of the course on Waves and Optics. The phenomena associated with light such as interference, diffraction and polarisation can be explained on the basis of wave model of light.

Therefore, we begin this course with a discussion about formation and propagation of waves and its mathematical description. We have considered mechanical waves such as waves on a string to explain wave motion. However, the concepts that you study in this unit apply to all kinds of physical waves: light and all other electromagnetic waves such as radio waves, microwaves, infrared, ultraviolet, X-rays, gamma rays, and sound waves, waves on the water surface, etc. Although waves and related phenomena appear obvious, the scientific understanding of the underlying concepts requires careful study.

Many concepts of this unit are familiar to you from your +2 level physics course. You have also studied about waves in Unit 19 of the first semester course entitled Mechanics (BHPCT-131). Therefore, you will find it easy to comprehend the concepts discussed in this unit. It will be good if you revise Unit 19 as we will not go into all the details here but focus only on the main points. While studying this unit, you should focus on the following important aspects associated with waves: i) how oscillatory motion leads to wave motion; ii) what are the parameters which specify a wave and its propagation; and iii) how is energy carried by waves. The level of mathematical treatment in this unit is fairly simple. You are advised to solve all the SAQs, and TQs to master your understanding of the concepts discussed in this unit.

“All truths are easy to understand once they are discovered; the point is to discover them.”

Galileo Galilei

1.1 INTRODUCTION

The study of wave motion is interesting because waves are present all around us. For example, if you have ever stood at the sea shore or dropped pebbles in a still pond, you must have seen **water waves**. What we *hear* and what we see in the world around us depends on waves. When we *speak*, our *vocal cords* inside our throat *vibrate*. Their vibrations cause the surrounding air molecules to vibrate and the effect is heard as **sound**. When this sound reaches the ears of other persons near us, their ear drums vibrate and cause the sensation of hearing in them. From your school physics course, you know that sound is a form of energy and it is carried by **sound waves**, which enable us to hear what others speak.

Visible light enables us to see. You know from the second semester course BHPCT-133 entitled Electricity and Magnetism that it is an **electromagnetic wave**. You also know about other electromagnetic waves such as radio waves, microwaves and X-rays. Most modern communication technologies such as radio, television, telephone, fax, internet, etc. are based on transmission and receipt of signals in the form of radio waves and microwaves. X-rays are used in medical diagnosis, e.g., for taking images of bones to diagnose fractures. At the microscopic level, we learn about **matter waves** to understand the nature of atoms, molecules, electrons, protons and other elementary particles.

All these examples should convince you that understanding the physics of wave motion is of fundamental importance. The waves mentioned above can be broadly categorised into three main types: **mechanical waves**, **electromagnetic waves** and **matter waves**. In this unit, we shall discuss mechanical waves on a string. However, the basic ideas we develop in the unit apply to waves of all kinds including light (electromagnetic wave).

We begin our study of wave motion by describing briefly, in Sec. 1.2, **how waves are formed and how they propagate**. For this, we consider transverse waves on a string which you have studied in Unit 19 of BHPCT-131. In this section, you will also learn how to depict wave motion graphically. In Sec. 1.3, you will revise the mathematical description of waves. The two types of waves, namely, plane waves and spherical waves have also been discussed qualitatively in this section. One of the important reasons to study wave motion is that waves carry energy from one place to another. In Sec. 1.4, you will learn how to obtain the expression of **energy transported by travelling waves**. Finally, in Sec. 1.5, we introduce the concept of **intensity** of a wave.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ explain the propagation of transverse waves on a string;
- ❖ represent waves on vibration graph and waveform graph;
- ❖ write the mathematical expression for a wave;
- ❖ describe plane waves and spherical waves;
- ❖ derive an expression for the energy carried by waves; and
- ❖ explain the concept of intensity of a wave and show that it depends on the amplitude of the wave.

1.2 WAVES ON A STRING

We begin our discussion on the phenomena of formation of waves and their propagation by considering wave formation on a stretched string. Before we do that, you should perform the following activity which will help you understand how waves are generated on a string.

Activity

Take a long thin elastic string and fix one of its ends to a distant wall as shown in Fig. 1.1a. Hold the other end of the string with your hand so that the string is stretched and taut. Now quickly move your hand up and down, once. What do you observe? A disturbance marked A in Fig. 1.1c travels along the length of the string. An isolated disturbance like A is called a **pulse** and it is generated due to one quick up and down motion of the portion of the string held in your hand.

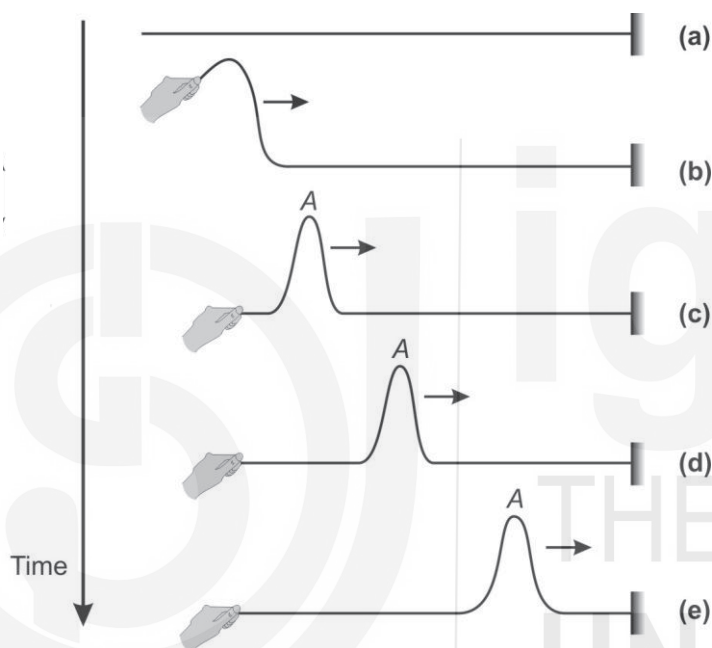


Fig. 1.1: a) An string fixed to a rigid wall; b) one quick up and down motion of the portion of the string held by hand; c) pulse A is generated; d) and e) show that the pulse A moves along the string as time passes.

The pulse generated on the string is also called **disturbance**. The word **disturbance** is used as a general term which refers to the deformation in the shape of the string with respect to its undisturbed horizontal configuration as shown in Fig. 1.1a.

In the above activity, if you keep moving the end of string held in your hand up and down in a periodic manner, you will observe that a wave is generated on the string as shown in Fig. 1.2a. You have learnt in Sec. 19.2 that **the periodic motion of the string in hand at one end gives rise to a wave moving along the string**. So we will not repeat the discussion here.

Note that Figs. 1.2a to 1.2e show five **snapshots** (photographs) of a wave on the string beginning at the instant $t = 0$. These have been taken at the intervals of $T/4$ up to the instant $t = T$. Here T is the **time period**.

In Fig. 1.2a showing the waveform at the instant $t = 0$, the point marked A on the string is the maximum value of the disturbance (displacement of the element of the string from its equilibrium position). It is called the **crest** of the wave. Now, let a ribbon, shown by mark X, be tied at the point of the string at the position x_1 in such a manner that it coincides with the position of crest A at the instant $t = 0$. Can you see **what happens to the ribbon and the crest A as free end of the string is moved up and down?**

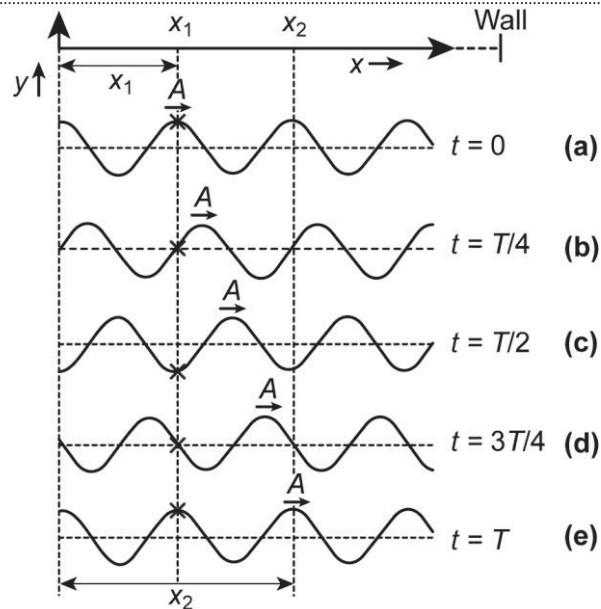


Fig. 1.2: Periodic motion of the hand holding the free end of the string generates a disturbance with a sinusoidal profile. Parts (a) to (e) of the Figure depict how the disturbance generated by the hand travels along the string to form a wave.

The **ribbon moves up and down** just like the hand but the crest **A** moves from x_1 to x_2 , and so on. Thus, we can conclude that as a wave (disturbance) propagates in a medium (string), two distinct motions are taking place:

- motion of the disturbance (represented by crest A), and
- the oscillatory motion of the particles (represented by the ribbon) of the medium.

We have taken the example of mechanical waves on a string for introducing the concept of wave motion. You know that **mechanical waves can exist only in material medium such as strings, water, air, rocks, etc.** Of course, sound waves are the most familiar example of mechanical waves in air.

Mechanical waves are of two types: **transverse** and **longitudinal**. The wave travelling on a string is an example of transverse wave and sound wave in air is an example of longitudinal wave.

Electromagnetic waves are also transverse in nature.

The periodic oscillatory motion of the end of the string held by hand **forces** the particles or elements of the string (medium) to execute oscillatory motion. Such oscillatory motion of the particles of the medium is called forced oscillations.

In transverse waves, the particles of the medium (such as string) **oscillate perpendicular to the direction in which the wave travels**. Waves propagating on the strings of musical instruments (such as *ektara*, *sarangi*, *sitar*, *vina* and violin) are transverse waves.

On the other hand, in a longitudinal wave, particles of the medium oscillate along the direction of propagation of the waves. Mechanical waves in a gaseous medium are always longitudinal but those in liquids and solids may be both longitudinal and transverse.

Waves produced due to *continuous* and *periodic forced oscillations* in an open medium (medium without boundaries) are **travelling periodic waves**. If the oscillations of the driving force such as the hand holding the free end of the string are **simple harmonic**, the travelling waves are called **harmonic waves**; these waves propagate with the frequency of the periodic driving force. In this

unit we shall confine our discussions only to **harmonic waves in one-dimension**.

You may like to quickly check your understanding of waves answering a simple SAQ.

SAQ 1 – Characteristics of mechanical waves

Fill in the blanks:

- A single disturbance that moves through a medium is called a
- When the particles of the medium are vibrating along a direction perpendicular to the direction of wave propagation, then the wave is a wave.
- If a one dimensional longitudinal wave is propagating from west to east direction, the particles of the medium would be oscillating in direction.

We hope that you understand the role of the oscillatory motion in the generation of mechanical waves. The next step is to describe wave motion. So, we quickly revisit the concepts discussed in Sec. 19.3 of Unit 19 of the first semester course BPHCT-131 to represent waves graphically and mathematically. We begin by revising how to represent wave graphically. **For simplicity, we consider mechanical waves on a string**, though the results will apply to other types of one-dimensional (1-D) waves as well.

1.2.1 Representation of Wave Motion

To keep the discussion simple, we assume that

- The waves are harmonic, i.e., their source (driving force) executes simple harmonic motion.
- The medium is uniform and the wave propagates at a constant speed.

Under these assumptions, the wave propagates with the same time period as the period of the oscillating source. That is, the frequency, f of the wave is equal to the frequency of the oscillating source (or the particles of the medium). Since frequency is reciprocal of time period, $f = 1/T$. The

amplitude of the wave is given by the **maximum displacement** of the particles of the medium from their respective mean positions.

Further, the particles of the medium located at two consecutive crests or troughs of a wave are in the same **state of motion; that is, their displacements and velocities are the same**. Any pair of two such particles of the medium are said to be in the same **phase**. The distance between any two consecutive particles on a waveform in the same phase is called the **wavelength** of the wave.

On the basis of above information, we can now represent a sinusoidal wave graphically. Let us take the example of sinusoidal waves on a string (Fig. 1.2) and represent them graphically. For depicting waves, we show the displacement of the particle(s) of the medium on two types of graphs:

Vibration Graph: In this graph, we plot the displacement, y of a particle of the medium located at a fixed position, x as time passes. This graph, shown in

Fig. 1.3, shows the wave behaviour at a **single location** along the wave's path as time passes. We can obtain it by fixing a slit perpendicular to the direction of wave propagation at a point in space along the path of the wave and observe wave motion as time passes. It will be a sinusoidal graph.

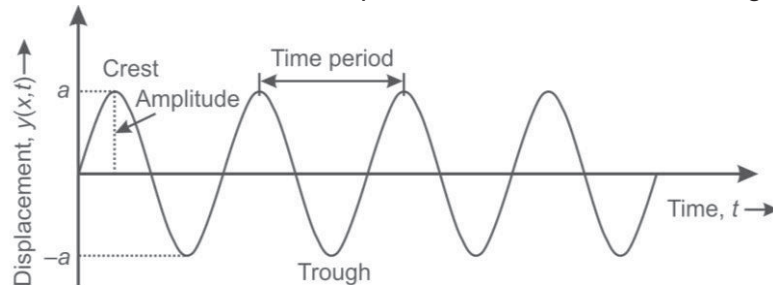


Fig. 1.3: The vibration graph of a wave showing its motion at a *given position* as a function of time.

The vibration graph **represents the motion of a wave with time at a given location in space**. In other words, the vibration graph represents the vibratory/oscillatory motion of a particle of the medium located at a fixed point as wave propagates. You know that such simple harmonic motion of the particles of the medium is represented by

$$y(t) = a \sin \omega t \quad (1.1)$$

Here, a is the **amplitude** of the wave (which is the same as the amplitude of oscillation of the particles of the medium) and ω , its **angular frequency**. It is related to the **frequency**, f by $\omega = 2\pi f$ and to the time period, T by $\omega = 2\pi/T$. Thus, Eq. (1.1) can also be written as

$$y(t) = a \sin 2\pi(t/T) \quad (1.2)$$

The **vibration graphs** inform us of the wave's *shape*, *amplitude*, and *time period* as shown in Fig. 1.3.

Waveform Graph: In this graph, we keep the time fixed and plot the displacement, y of the particles located at different positions, x . The waveform graph (Fig. 1.4) is the same as the snapshot of a wave at any instant. It displays wave behaviour at **different locations at a single moment in time** and indicates the wave's *shape*, *amplitude* and *wavelength* (Fig. 1.4).

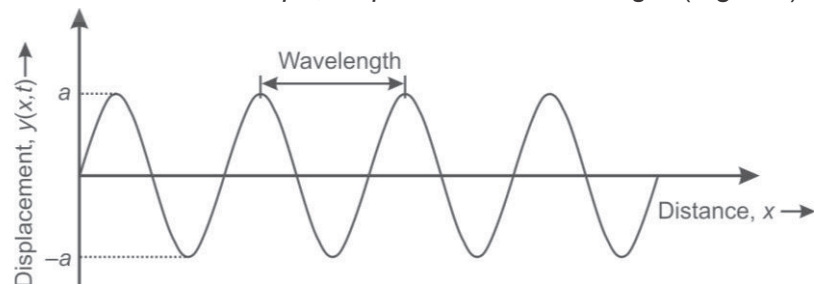


Fig. 1.4: The waveform graph of a wave showing displacement of the particles of the medium as a function of position at a *given time*; this is like the snapshot of the wave at any instant of time.

Further, note that the waveform graph is also sinusoidal and **represents the wave at a given instant of time**. Thus, we can write the equation representing this graph as

$$y(x) = a \sin kx \quad (1.3)$$

where k is a constant and it is called **wave number** of the wave. The wave number, k is related to the wavelength, λ of the wave. To discover the

relation, recall that the wavelength, λ is the distance between any two consecutive points, along the distance axis, which are in the same state of motion (that is, in the same phase), as shown in Fig. 1.4. Thus, the displacement, $y(x)$ is the same at both ends of this wavelength, i.e., $y(x = x_1) = y(x = x_1 + \lambda)$. So, using Eq. (1.3), we can write,

$$a \sin k x_1 = a \sin k (x_1 + \lambda) = a \sin (kx_1 + k\lambda)$$

Since $\sin(\theta + 2\pi) = \sin \theta$, where θ is any variable, the above equality will be satisfied only when

$$k\lambda = 2\pi \quad \text{or} \quad k = \frac{2\pi}{\lambda} \quad (1.4)$$

Eq. (1.4) gives the relation between k and λ .

Using Eq. (1.4) in Eq. (1.3), we can write

$$y(x) = a \sin 2\pi(x/\lambda) \quad (1.5)$$

In addition to the wave parameters such as amplitude, wavelength, frequency, and phase defined so far, we also need to know the velocity of a wave and how it is related to the wave parameters. Let us find out.

1.2.2 Relation between Wave Velocity, Frequency and Wavelength

We can establish the relation between wave velocity, frequency and wavelength on the basis of their definitions. Recall that a wave travels a distance equal to one wavelength in one time period. Therefore, the wave velocity is given by

$$v = \frac{\text{Wavelength}}{\text{Time Period}} = \frac{\lambda}{T}$$

Since frequency is reciprocal of time period, ($f = 1/T$), we can write

$$v = f \lambda \quad (1.6)$$

That is, the velocity of wave is equal to the product of its frequency and its wavelength.

Note that we have obtained the expression for wave velocity [Eq. (1.6)] for a wave on a string. **But, this equation holds for all kinds of waves, whether transverse or longitudinal, mechanical or electromagnetic.**

Before studying further, you may like to answer an SAQ

SAQ 2 – Calculating wave parameter

A transverse wave is generated on a string due to the periodic motion of one of its ends with the other end fixed to a distant rigid support. If the frequency of the periodic motion of the free end of the string is 3.0 Hz and the wave speed is 6.0 ms^{-1} , calculate the distance between two adjacent crests of the wave.

The vibration graph and waveform graph represent a wave at fixed position and at a fixed time, respectively. But, as we know, a travelling wave varies with time as well as position. Thus, neither Eq. (1.2) nor Eq. (1.5) can be taken as a complete mathematical description of a travelling wave. Let us now learn how to represent a wave mathematically.

Note that the wave number k is defined as $2\pi / \lambda$. Actually, the term wave number should refer to the number of waves in one metre. That is, it should be a quantity given by $(1/\lambda)$. And $k (= 2\pi/\lambda)$ should be called angular wave number. But, we will follow the standard convention and call k the wave number. You should be mindful of the factor 2π used in defining angular frequency ω and wave number k .

You know that velocity is a vector quantity which contains information about the magnitude of the velocity (that is, speed) and the direction of motion. For one-dimensional waves, it is one and the same whether we use the term velocity or speed provided we assume that the wave is propagating along the positive x-direction.

1.3 MATHEMATICAL DESCRIPTION OF WAVE MOTION

Refer to Fig. 1.5 which shows two snapshots of a wave travelling along the positive x -direction. Fig. 1.5a depicts the waveform at time $t = 0$ and Fig. 1.5b depicts the waveform at a later time $t_1(> 0)$. Note that in time t_1 , the entire waveform has moved in the positive x -direction. Let us assume that the wave travels with velocity v .

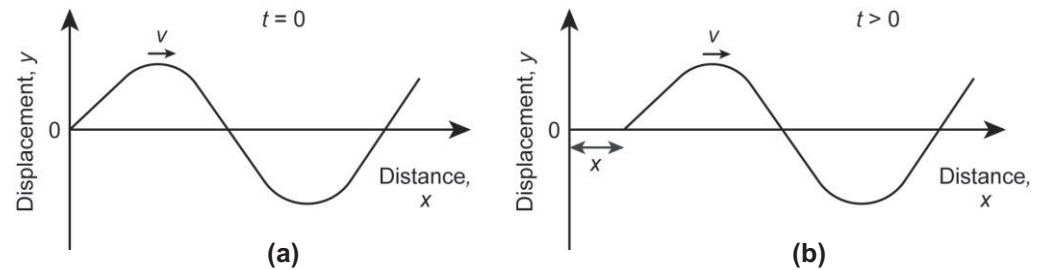


Fig. 1.5: Snapshots of a wave travelling in positive x -direction at a) $t = 0$; b) $t > 0$.

As you know, in wave motion, the entire waveform moves. Thus, as evident from Fig. 1.5, the disturbance at $x = 0$ (that is, the displacement y of the particle of the medium at $x = 0$) will travel to a point $x (> 0)$ at a later time, because the disturbance takes time to travel to that point. The time taken by the disturbance depends on how far the point is from $x = 0$ and what is the velocity of propagation of the disturbance (or the wave).

Suppose the disturbance at $x = 0$ (Fig. 1.5a) takes time Δt in travelling to point $x (> 0)$ (Fig. 1.5b). Thus, the displacement of the particle at x at time t will be the same as the displacement of the particle at $x = 0$ at time $(t - \Delta t)$ where, Δt is the time taken by the displacement to move from $x = 0$ to $x = x$.

Now, from Eq. (1.2), you know that the displacement of a particle at point x is given as

$$y(x, t) = a \sin 2\pi(t/T)$$

So, to describe the motion of the disturbance or the wave, we replace t by $(t - \Delta t)$ in the above equation and get

$$y(x, t) = a \sin 2\pi\left(\frac{t - \Delta t}{T}\right) \quad (1.7)$$

To find the time Δt , we note that the disturbance travels with velocity v . So, we can write

$$v = \frac{\Delta x}{\Delta t} = \frac{x - 0}{\Delta t}$$

$$\text{So,} \quad \Delta t = \frac{x}{v} \quad (1.8)$$

Substituting Eq. (1.8) in Eq. (1.7), we get

$$\begin{aligned} y(x, t) &= a \sin 2\pi\left(\frac{t}{T} - \frac{x}{vT}\right) \\ &= a \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) \end{aligned} \quad (1.9a)$$

It is possible to write Eq. (1.9a) in a few other equivalent forms. To show, we note that $v = \lambda / T$. Then we can write Eq. (1.9a) in terms of v and T as:

$$y(x, t) = a \sin \left[\frac{2\pi}{T} \left(t - \frac{x}{v} \right) \right] \quad (1.9b)$$

We can also rewrite this equation as:

$$y(x, t) = a \sin \left[\frac{2\pi}{\lambda} (vt - x) \right] \quad (1.9c)$$

Any one of the Eqs. (1.9a to c) gives a complete mathematical description of a one-dimensional (1-D) wave travelling along the positive x-direction. Note that Eqs. (1.9a), (1.9b) and (1.9c) are equivalent representations of a 1-D wave. In terms of angular frequency $\omega (= 2\pi / T)$ and wave number $k (= 2\pi / \lambda)$, we can write Eq. (1.9a) as

$$y(x, t) = a \sin (\omega t - kx) \quad (1.9d)$$

The simple way in which ω and k enter the mathematical expression for the wave explains why these quantities are so often used in the description of wave motion. Note that a wave described by either of the Eq. (1.9a to d) has a single constant frequency and signifies a **monochromatic** wave.

Eqs. (1.9a to d) provide us **equivalent descriptions** of a wave moving in the positive x-direction; which of these equations we use is a matter of convenience and depends on the specific situation.

It is important to mention here that these equations describe all types of one-dimensional sinusoidal waves travelling in the positive x-direction. They represent transverse waves arising due to the vertical oscillations of particles of a string as well as longitudinal waves arising due to the horizontal or longitudinal oscillations of the particles of the medium. The only difference is in **what the quantity $y(x, t)$ represents**: It represents vertical displacement in the case of waves on strings and horizontal displacement in the case of longitudinal waves.

You may also like to know: **How do we describe a wave travelling in the negative x-direction? A wave travelling in the negative x-direction** is represented by any one of the following equivalent expressions:

$$y(x, t) = a \sin \left[\frac{2\pi}{\lambda} (vt + x) \right] \quad (1.10a)$$

$$y(x, t) = a \sin \left[\frac{2\pi}{T} \left(t + \frac{x}{v} \right) \right] \quad (1.10b)$$

$$y(x, t) = a \sin \left[2\pi \left(\frac{t}{T} + \frac{x}{\lambda} \right) \right] \quad (1.10c)$$

$$y(x, t) = a \sin (\omega t + kx) \quad (1.10d)$$

Note that for a longitudinal wave, both the position, x and displacement, $y(x, t)$ are along the same direction. They, however, refer to two different quantities: x refers to the position of a particle of the medium and $y(x, t)$ refers to the displacement of the particle located at x , with respect to its equilibrium position.

On comparing Eqs. (1.9b) and (1.9d) or Eqs. (1.10b) and (1.10d), we obtain the expression for wave velocity in terms of ω and k as:

$$v = \omega / k \quad (1.11)$$

So far, you have learnt the graphical and mathematical descriptions of one dimensional wave. Now, you would like to know: Is there an equation which governs the propagation of waves similar to Newton's laws of motion which govern the motion of objects? Yes, there is such an equation called the wave equation. Let us learn it now.

Wave Equation

You will learn to derive wave equation if you pursue higher studies in physics as it is beyond the scope of this course.

The wave equation governing the propagation of the transverse wave on a string is obtained by applying Newton's second law of motion to a small segment of the string. Here, we shall write the wave equation without deriving it and discuss its physical significant qualitatively (Read the margin remark). The wave equation for 1-D wave is given as:

$$\frac{\partial^2 y}{\partial t^2} = v^2 \frac{\partial^2 y}{\partial x^2} \quad (1.12)$$

To understand what the wave equation conveys physically, let us consider the transverse wave on a string. The LHS of Eq. (1.12) represents the acceleration of an element of the string. (Note that $y(x, t)$ gives the displacement of an element of the string, located at x , at time t .) The use of partial derivative is due to the fact that the displacement of the particles of the medium is a function of x and t both. On the RHS of Eq. (1.12), v represents the velocity of the wave and the term $(\partial^2 y / \partial x^2)$ gives a measure of the amount of bending in the string. The wave equation [Eq. (1.12)] governs the motion of all kinds of 1-D waves.

Note that $(\partial y / \partial x)$ gives the slope of the string at point x . Therefore,

$$\frac{\partial^2 y}{\partial x^2} = \left(\frac{\partial}{\partial x} \left(\frac{\partial y}{\partial x} \right) \right)$$

gives how the slope varies with position, x .

Some important points related to wave equation are given below:

1. The wave equation gives the velocity of the wave in terms of the properties of the medium. It is found that, for transverse waves on a string, the velocity of wave is given by

$$v = \sqrt{T / \rho} \quad (1.13)$$

where T is the tension in the string and ρ , its linear mass density.

2. Although Eq. (1.12) gives the wave equation in terms of the parameters associated with 1-D transverse wave propagating on a string, similar wave equation is associated with all kinds of waves. In the wave equation for sound waves in a fluid, the wave function, that is, the displacement $y(x, t)$ represents the longitudinal displacement of the particles of the fluid. Similarly, we have wave equation for light (electromagnetic wave) wherein the wave function represents the oscillating electric and magnetic fields. So, in general, we can write the 1-D wave equation as

$$\frac{\partial^2 \psi}{\partial t^2} = v^2 \frac{\partial^2 \psi}{\partial x^2} \quad (1.14)$$

where $\psi(x, t)$ represents **the physical quantity that oscillates as the wave passes**. For example, in case of transverse wave on a string, $\psi(x, t)$ is $y(x, t)$, the displacement of an element of the string, located at x , at time t . As time passes, this element of the string oscillates about its mean position.

3. Another important aspect of the wave equation is that it enables us to determine whether or not a given wave function (that is, the mathematical

Note that we have used the symbol T for time-period of the wave and also for the tension in the string. You should be mindful of the context to know what does the symbol T refers to.

representation of a wave) represents a progressive wave. That is, **any function which represents a wave must satisfy the wave equation**. To elaborate this, let us take the example of transverse wave on a string.

You have seen that Eqs. (1.9) and (1.10) represent transverse waves on a string propagating along positive x -direction and negative x -direction respectively. The wave functions represented by Eqs. (1.9) and (1.10) must satisfy the wave equation for 1-D transverse wave on the string given by Eq. (1.12). By taking derivatives ($\partial^2 y / \partial t^2$) and ($\partial^2 y / \partial x^2$) of Eqs. (1.9) and (1.10), you can show that it is indeed true. So, if we are given a function and asked to ascertain whether or not it represents a travelling wave on a string, we can check that by requiring the function to satisfy the wave equation given by Eq. (1.12).

To concretise your understanding of wave equation and wave function representing a wave, solve the following SAQ.

SAQ 3 – Mathematical representation of a wave and wave equation

Show that the following functions satisfy the wave equation given by Eq. (1.12):

- a) $y(x, t) = (x - vt)^2$
- b) $y(x, t) = a \sin(kx - \omega t)$
- c) $y(x, t) = a \sin(kx + \omega t)$
- d) $y(x, t) = ae^{i(kx - \omega t)}$

On solving SAQ 3, you know that a wave can be represented by a variety of mathematical functions. The only condition is that the function must satisfy the wave equation. Apart from Eq. (1.9), two more commonly used mathematical representations of 1-D wave travelling in the positive x -direction are

$$y(x, t) = a \sin(kx - \omega t) \quad (1.15a)$$

and $y(x, t) = a \cos(kx - \omega t) \quad (1.15b)$

Similarly, the waves travelling in the negative x -direction can be represented as

$$y(x, t) = a \sin(kx + \omega t) \quad (1.15c)$$

and $y(x, t) = a \cos(kx + \omega t) \quad (1.15d)$

You should convince yourself that the wave functions given by Eq. (1.15) represent progressive waves because they satisfy wave equation given by Eq. (1.12).

However, the particular choice of the mathematical representation of a wave is determined by the convenience of mathematical treatment and physical situation under consideration.

Now, you should go through the following example to learn to calculate wave parameters from a given expression for a wave.

EXAMPLE 1.1: CALCULATING WAVE PARAMETERS

The expression for a 1-D wave travelling along the positive x-direction is given as

$$y(x, t) = 0.04 \sin[\pi(x - 4t)] \text{ m}$$

Calculate the amplitude, wave number, wavelength, angular frequency and frequency of the wave.

SOLUTION ■ Comparing the given expression for the wave with

$$y(x, t) = a \sin(kx - \omega t)$$

We get,

$$\text{amplitude, } a = 0.04 \text{ m}$$

$$\text{wave number, } k = \pi \text{ m}^{-1}$$

$$\text{So, wavelength, } \lambda = \frac{2\pi}{k} = \frac{2\pi}{\pi} = 2.0 \text{ m}$$

$$\text{angular frequency, } \omega = 4 \pi \text{ s}^{-1}$$

$$\text{So, frequency, } f = \frac{\omega}{2\pi} = \frac{4\pi}{2\pi} \text{ s}^{-1} = 2.0 \text{ s}^{-1}.$$

You may now like to solve an SAQ.

SAQ 4 – Calculating parameters of a wave

A 1-D sinusoidal wave is propagating along the positive x-direction. The displacement at two points $x_1 = 0$ and $x_2 = 2.0 \text{ cm}$ is given by the following expressions:

$$y(x_1, t) = (0.02 \text{ cm}) \sin[(3\pi \text{ s}^{-1}) t]$$

$$y(x_2, t) = (0.02 \text{ cm}) \sin\left[(3\pi \text{ s}^{-1}) t + \frac{\pi}{2}\right]$$

Determine the amplitude, frequency, wavelength, direction of propagation and speed of the wave.

Before proceeding further, let us recall the important points discussed so far.

Recap

- When a mechanical wave propagates in a medium, two distinct motions take place: (i) motion of the disturbance, and (ii) oscillatory motion of the particles of the medium.
- A sinusoidal 1-D wave travelling in the positive x-direction can be represented by

$$y(x, t) = a \sin(\omega t - kx)$$

$$y(x, t) = a \sin(kx - \omega t)$$
- A sinusoidal 1-D wave travelling in the negative x-direction can be represented by

$$y(x, t) = a \sin(\omega t + kx)$$

$$y(x, t) = a \sin(kx + \omega t)$$
- The wave equation for 1-D wave is given as

$$\frac{\partial^2 y}{\partial t^2} = v^2 \frac{\partial^2 y}{\partial x^2}$$

So far, you have learnt that waves propagate in a medium when the particles of the medium oscillate about their respective mean positions. You have also learnt about parameters such as the **amplitude, time period, frequency and wavelength of the wave** which characterise a wave. To complete the mathematical description of a wave, we also need to know the **phase** of a wave. You have learnt about phase of a wave and phase difference in detail in Sec. 19.3.4 of Unit 19, BPHCT-131. We, therefore, briefly discuss it here for the sake of completeness.

1.3.1 Phase of a Wave and Phase Difference

Recall from Sec. 1.2 that two particles on a waveform graph of a wave are said to be **in the same state of motion** if their displacements and velocities are the same. Thus, even though all particles of the medium execute simple harmonic motion and their displacements show the same sinusoidal variation in time, their **states of motion** may be different at any given instant of time. We denote this **difference in the states of motion of particles of the medium** in terms of the **phase angle** or simply **phase**.

Mathematically, **the argument of the sine (or cosine) function [Eq. (1.9)] representing a periodic travelling wave** represents the **phase** of the wave. We denote it by the symbol $\phi(x, t)$. Thus, the phase of a sinusoidal wave at position, x and time, t represented by Eq. (1.9d) is the argument of the sine function given by

$$\phi(x, t) = \omega t - kx \quad (1.16)$$

Note that the phase of a wave is an angle and is measured in degrees or in radians; **a phase difference of 360° corresponds to one wavelength.**

Eq. (1.16) shows that the phase of a wave changes with both space and time. Further, from the definition of the phase, it follows that **all points on the waveform separated by one wavelength or its integer multiples are in the same phase.**

To obtain the expression for the phase difference between two arbitrary points on a wave, let us consider any two points located at positions say, x_1 and x_2 , respectively, on a waveform graph (Fig. 1.4). Mathematically, the phases ϕ_1 and ϕ_2 of the particles at positions x_1 and x_2 on the waveform at a fixed time t can be written as [Eq. (1.16)]:

$$\phi_1 = \omega t - kx_1$$

and
$$\phi_2 = \omega t - kx_2$$

Hence, the phase difference, $\Delta\phi$ between the particles at two positions x_1 and x_2 on the waveform is given by

$$\begin{aligned} \Delta\phi(x, t) &= (\omega t - kx_2) - (\omega t - kx_1) \\ &= k(x_1 - x_2) \\ &= \frac{2\pi}{\lambda}(x_1 - x_2) \end{aligned} \quad (1.17)$$

We can extend this discussion of phase of different points on a single waveform with respect to some reference point to a situation where more than one wave is travelling in space. **Two waves are said to be in phase when the corresponding points of each wave reach their respective maximum or minimum displacements at the same time.** Thus, if the crests and

troughs of the two waves coincide, they are said to be **in phase** (Fig. 1.6a). However, if the crest of one wave coincides with the trough of the other wave, as shown in Fig. 1.6b, their phases are said to differ by 180° and the waves are said to have **opposite phase**. The **phase difference** between two waves can vary from 0 to 360° . Fig. 1.6c shows two out of phase waves with an arbitrary phase difference. These ideas about phase difference between two waves play an important role in deciding the outcome of their superposition. You will learn about superposition of waves in the next unit.

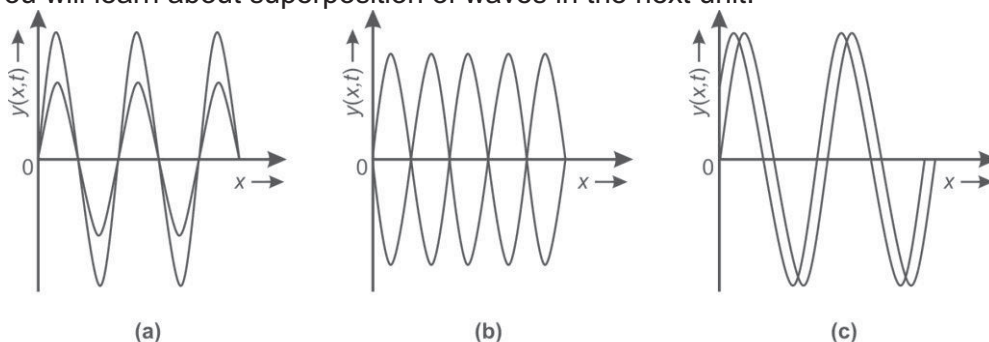


Fig. 1.6: Two waves A and B are a) in-phase; b) completely out of phase; c) out of phase by an arbitrary angle.

Till now, we have discussed wave phenomena by considering one dimensional wave on a string. But, the waves we encounter in nature are generally two-dimensional or three-dimensional. For example, waves on water surface are two-dimensional and sound waves in air are three-dimensional. Such waves are represented as plane waves and spherical waves using the concept of wavefront. Let us now learn about it.

1.3.2 Plane Waves and Spherical Waves

Let us first consider waves on a water surface. You know that if you drop a pebble on the surface of water, waves propagate outward from the point where the pebble touched the water surface. The wave spreads in the form of concentric circles. This is an example of two-dimensional (2-D) wave. The radius of each circular crest (or trough) increases as the wave propagates.

Note that all the particles on a given crest or trough will be in phase because all these particles are at the same distance from the source (pebble touching the water surface). An expanding circular crest (or trough) is called a **wavefront**. A wavefront is defined as the locus of all points which are in phase. So, two-dimensional waves are characterised by **circular wavefronts**.

Now, let us imagine a point source which is generating waves in the surrounding medium. A point source of sound is a good example of such a source. Can you guess what would be the nature of the wavefront? The wavefronts of the waves will be spherical with the point source as centre. A few sections of concentric spherical surfaces representing the spherical wavefronts is shown in Fig. 1.7a.

Thus, the waves emanating from a point source in a uniform medium are called **spherical waves**. Spherical waves are characterised by spherical wavefronts. Further, the lines that are directed radially outward from the source are called **rays**. So, rays are lines perpendicular to the wavefront and directed radially outward from the source. Rays indicate the direction of propagation of waves. The concept of wavefront and rays is very useful in explaining many phenomena such as interference, diffraction and polarisation related to light waves.

Recall that the particles of the medium through which waves propagate are said to be in phase if their state of motion is same.

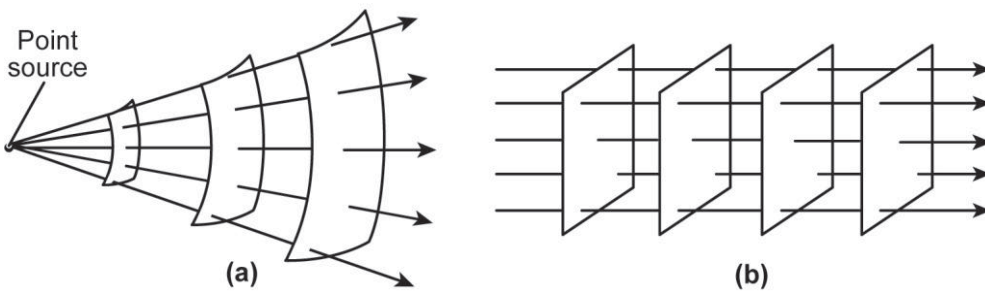


Fig. 1.7: a) Spherical; b) plane wavefronts for waves emanating from a point source.

Now, refer to Fig. 1.7 again. Suppose, the point source is at a very large distance from the point of observation. Will the nature of wavefront change? Yes; the spherical wavefronts will become nearly planar. So, at large distances from the source, a section of a plane can be used to represent the wavefront as shown in Fig. 1.7b.

Wave from a point source can, therefore, be considered as **plane waves** at large distances. **In case of plane waves, rays are straight lines parallel to each other.**

One of the important characteristics of waves is that it transports energy. You would, therefore, like to know how much energy is transported by a wave of frequency f and wavelength λ , travelling with the speed v ? We now seek the answer to this question.

1.4 ENERGY TRANSPORTED BY WAVES

To obtain an expression for the energy transported by a travelling wave, let us consider a taut string with one end fastened to a rigid wall and the other end attached to a vertical oscillator, as shown in Fig. 1.8a. When the oscillator is at rest, each element of the string will be in its respective equilibrium position. So, in this condition, the string has no kinetic energy and, its elastic potential energy is also zero.

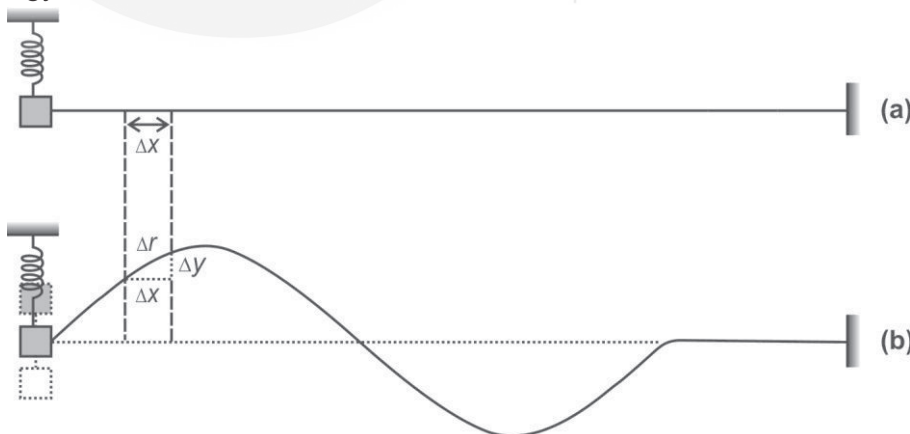


Fig. 1.8: a) A stationary string whose one end is fixed and the other end is attached to a vertical oscillator; b) snapshot of the transverse wave on the string generated by the vertical oscillator attached to its free end.

When a periodic force is applied to the free end of the string due to vertical oscillations of the attached oscillator, a sinusoidal transverse wave is generated on the string. As the wave propagates along the string, each

element of the string is set into motion and gains kinetic energy. Also, since the string is under the influence of constant force (tension), it will have potential energy due to stretching, which is essentially elastic potential energy. As the sinusoidal wave propagates on the string, the energy contained in one wavelength of the wave is transported along the propagation direction.

To calculate the energy contained in the wave, let us consider a small element dx of the string as shown in Fig. 1.8. Note that as the wave propagates, the length of the element dx of the stationary string (Fig. 1.8a) stretches to dr (Fig. 1.8b) due to tension T in the string. But the mass of the stretched element dr is the same as the mass ρdx of the element dx , where ρ is the mass per unit length of the string.

Thus, the kinetic energy of the element Δx can be written as

$$d(K.E.) = \frac{1}{2}mv^2 = \frac{1}{2}(\rho\Delta x)\left(\frac{dy}{dt}\right)^2 \quad (1.18)$$

where $\rho\Delta x$ is the mass of the element Δx and (dy/dt) is its velocity. Actually, the function, $y(x, t)$ describing the wave is a function of two variables namely position x and time t . Therefore, strictly speaking, we should take a partial derivative, $\frac{\partial y(x, t)}{\partial t}$, of $y(x, t)$ with respect to the variable t keeping x constant to determine the velocity v of the element Δx .

Since the sinusoidal wave is represented by Eq. (1.9d):

$$y(x, t) = a \sin(\omega t - kx)$$

we can write

$$\frac{\partial y}{\partial t} = a \omega \cos(\omega t - kx)$$

Thus, on substituting the value of $\partial y / \partial t$ in Eq. (1.18), we get the kinetic energy of the element Δx as

$$d(K.E.) = \frac{1}{2}(\rho\Delta x) [a^2 \omega^2 \cos^2(\omega t - kx)] \quad (1.19)$$

Now, to calculate the potential energy which arises due to elastic potential energy of the string, we note that **as the wave propagates along the string, the string is stretched and acquires elastic potential energy**. From Fig. 1.8b, you may note that the length of the element increases by an amount $(\Delta r - \Delta x)$. This stretching is caused due to the tension T in the string which is a constant force. So, the elastic potential energy acquired by the element of length Δx is:

$$d(P.E.) = \text{Work done} = \text{Force} \times (\text{Increase in the length of the string}) = T(\Delta r - \Delta x)$$

From Fig. 1.8b, we note that

$$\Delta r = \sqrt{(\Delta x)^2 + (\Delta y)^2}$$

$$\text{So, } d(P.E.) = T [\{ (\Delta x)^2 + (\Delta y)^2 \}^{1/2} - \Delta x]$$

$$= T \Delta x \left[\left(1 + \left(\frac{\partial y}{\partial x} \right)^2 \right)^{1/2} - 1 \right] \text{ because } \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \cong \frac{dy}{dx} = \left(\frac{\partial y}{\partial x} \right)_{t=0}$$

Using the binomial theorem to expand the first term in the square bracket on the RHS of the above expression and retaining only the first two terms of the expansion, we get

$$d(P.E.) = \frac{1}{2} T \left(\frac{\partial y}{\partial x} \right)^2 \Delta x$$

Again, we use Eq. (1.9d) to calculate $\partial y / \partial x$:

$$\frac{\partial y}{\partial x} = -ak \cos(\omega t - kx)$$

Thus, the elastic potential energy of the element Δx of the string is

$$d(P.E.) = \frac{1}{2} (T \Delta x) [a^2 k^2 \cos^2(\omega t - kx)] \quad (1.20)$$

So, the total mechanical energy of the element of length Δx of the string is

$$\begin{aligned} dE &= d(K.E.) + d(P.E.) \\ &= \frac{1}{2} \rho \Delta x a^2 \omega^2 \cos^2(\omega t - kx) + \frac{1}{2} T \Delta x a^2 k^2 \cos^2(\omega t - kx) \\ &= \frac{1}{2} \Delta x a^2 \cos^2(\omega t - kx) [\rho \omega^2 + T k^2] \end{aligned} \quad (1.21)$$

Using Eq. (1.21), we can write the **linear energy density**, $D(x, t)$ of the wave as (read the margin remark):

$$\begin{aligned} D(x, t) &= \frac{dE}{\Delta x} \\ &= \frac{1}{2} a^2 [(\rho \omega^2 + T k^2) \cos^2(\omega t - kx)] \\ &= \rho a^2 \omega^2 \cos^2(\omega t - kx) \end{aligned} \quad (1.22)$$

Eq. (1.22) shows that the space and time variation of energy density of a wave on a string is a function of $(\omega t - kx)$, which is same as the space and time variation of displacement, $y(x, t)$. What does this signify? It signifies that

- The energy density varies faster because of the square of cosine term.
- The energy of a wave propagates along its direction of propagation with velocity $v (= \omega / k)$.

Now, to calculate the total mechanical energy E contained in one wavelength of the wave, we integrate Eq. (1.22) from 0 to λ :

$$E = \int_0^\lambda D(x, t) dx = \rho a^2 \omega^2 \int_0^\lambda \cos^2(\omega t - kx) dx$$

For mathematical convenience, we set $t = 0$ for this integration. So, we get (study the margin remark):

$$E = \rho a^2 \omega^2 \int_0^\lambda \cos^2(kx) dx = \frac{1}{2} \rho a^2 \omega^2 \lambda \quad (1.23)$$

Eq. (1.23) gives the energy contained in one wavelength of the wave.

Now, power, P of the wave is the energy averaged over one time period, T and is given by

$$P = \frac{E}{T} = \frac{E}{(\lambda / v)} = \frac{1}{2} \rho a^2 \omega^2 v = 2\pi^2 \rho a^2 f^2 v \quad (1.24)$$

Binomial theorem:

$$\begin{aligned} (1+x)^n &= 1 + nx \\ &+ \frac{n(n-1)x^2}{2!} \\ &+ \frac{n(n-1)(n-2)x^3}{3!} + \dots \end{aligned}$$

We can write

$$\begin{aligned} (\rho \omega^2 + T k^2) \\ = \rho \omega^2 \left[1 + \frac{T}{\rho} \frac{k^2}{\omega^2} \right] \end{aligned}$$

We now use the relation for the velocity of the wave in terms of tension T in the string and mass per unit length ρ of the string given as:

$$v = \sqrt{T / \rho}$$

$$\text{and } \frac{k}{\omega} = \frac{1}{v}.$$

Then, we get

$$(\rho \omega^2 + T k^2) = 2\rho \omega^2$$

$$\begin{aligned} E &= \rho a^2 \omega^2 \int_0^\lambda \left[\frac{\cos 2(kx) + 1}{2} \right] dx \\ &= \rho a^2 \omega^2 \left[\int_0^\lambda \frac{dx}{2} + \frac{1}{2} \int_0^\lambda \cos(2kx) dx \right] \\ &= \rho a^2 \omega^2 \left[\frac{x}{2} + \frac{\lambda}{2\pi} \sin 2kx \right]_0^\lambda \\ &= \rho a^2 \omega^2 \left[\frac{\lambda}{2} \right] \end{aligned}$$

where $\omega = 2\pi f$ and T is the time period.

Eq. (1.24) shows that the rate of energy (i.e., power) transported by a wave varies linearly with wave velocity and the squares of its amplitude and frequency.

It is important to mention here that the expressions for average energy and power transported by a wave given, respectively, by Eqs. (1.23) and (1.24) are valid for all kinds of harmonic sinusoidal mechanical waves. We shall make use of this fact in Unit 3 where we discuss sound waves.

We now take a concrete example to give you some idea of the power transported by a mechanical wave. You can then solve a problem.

EXAMPLE 1.2: POWER TRANSPORTED BY A TRANSVERSE WAVE

A progressive transverse wave is travelling on a string of length 3.0 m and mass 300 g. The tension in the string is 40 N. If the frequency and the amplitude of the wave are 40 s^{-1} and 6 mm, respectively, calculate the average power of the wave over one time period.

SOLUTION ■ The average power of a progressive sinusoidal mechanical wave propagating in a medium of linear mass density ρ is given by Eq. (1.24):

$$P = \frac{1}{2} \rho a^2 \omega^2 v = 2\pi^2 \rho f^2 a^2 v \quad (\text{i})$$

We know that the amplitude of the wave $a = 6 \text{ mm} = 6 \times 10^{-3} \text{ m}$, frequency of the wave $f = 40 \text{ s}^{-1}$. The linear mass density of the string is its mass per unit length. It is:

$$\rho = \frac{m}{L} = \frac{300 \times 10^{-3} \text{ kg}}{3.0 \text{ m}} = 0.1 \text{ kg m}^{-1}$$

$$\text{Velocity of the wave, } v = \sqrt{\frac{T}{\rho}} = \sqrt{\frac{40 \text{ N}}{0.1 \text{ kg m}^{-1}}} = 20 \text{ m s}^{-1}$$

Substituting these values in Eq. (i), we get

$$P = 2 \times (3.14)^2 \times (6 \times 10^{-3} \text{ m})^2 \times (40 \text{ s}^{-1})^2 \times (0.1 \text{ kg m}^{-1})^2 \times (20 \text{ m s}^{-1}) \\ = 2.27 \text{ W}$$

This result shows that under the given conditions, the wave transports the average power 2.27 W or the average energy 2.27 J per second through the string.

We would now like you to work out an SAQ.

SAQ 5 – Power of a mechanical wave

A transverse wave is propagating on a string of length 1.0 m and mass 500 g. The tension in the string is 50 N, and the frequency and the amplitude of the wave are 50 s^{-1} and 8 mm, respectively. Calculate the average power of the wave.

The amount of energy transported by waves can be huge; these can cause immense destruction but can also be used constructively for generating electricity.

Energy Carried by Waves

A vivid demonstration of the energy carried by water waves is in the damage caused in coastal areas by **tidal waves** in stormy weather. You may know that the tidal waves generated in the super-cyclone in the Bay of Bengal in October, 1990 caused immense loss of life and property in coastal Orissa. More than ten thousand people lost their lives and millions were rendered homeless. Similar devastation by tidal waves was seen when typhoon Katrina struck US east coast in the year 2011. Tidal waves generated (due to an earthquake in Chile) carried huge amount of energy across 15,000 km of the Pacific Ocean and caused untold damage in Japan in the year 2010. *Do you know that a three metre high oceanic wave can lift 30 bags of wheat by about 10 ft?*

Seismic waves can also cause untold damage. The earthquake in South Gujarat on 26 Jan., 2001 reduced the area to rubbles killing an estimated one hundred thousand people. High-rise buildings, houses and hospitals collapsed and roads developed huge cracks. Similar devastation was experienced by the people in Jammu and Kashmir in October, 2005.

An earthquake under sea bed near Indonesia in the Indian Ocean on Dec. 26, 2004 caused a **tsunami** with waves of a height up to 30 feet and brought unimaginable misery in Indonesia, Thailand, Sri Lanka, Maldives, and India (Tamil Nadu).

You may like to read how the energy of tidal waves is being harnessed the world over to meet the increasing electricity requirements.

Electromagnetic (e.m.) waves

which include radio and microwaves, infrared, visible and ultraviolet light, X-rays and gamma rays, travel with speed $3 \times 10^8 \text{ ms}^{-1}$. Energy from the Sun reaches our planet in the form of electromagnetic radiations and sustains life.

In the context of waves, intensity of waves is another useful parameter. From our common experience, we know that the chirping of birds, the shout of a person, vehicular noise, sound of crackers or light from a lamp fade beyond a certain distance. If it were not true, noise pollution would have made our life miserable. Similarly, the surface temperatures of planets in the solar system depend on their respective distances from the Sun. **The energy and power of the waves do not give us any information about the variation in the strength of waves with distance.** We describe the strength of a wave at a given point in space by specifying the **intensity of a wave**. You will now learn about it.

1.5 INTENSITY OF WAVES

Intensity of a wave is defined as **the rate of energy transfer by the wave per unit area normal to the direction of wave propagation**. Mathematically, we can write the intensity (I) of the wave as

$$I = \frac{P}{A} \quad (1.25)$$

where P is the power and A is area of the surface intercepting the wave perpendicular to the direction in which the wave travels.

You must have experienced that loudness of sound decreases as we move away from its source. You may therefore like to know: **How does the intensity of sound vary with distance from the source?** Refer to Fig. 1.9 to

Note that we have considered a point source of waves such as a point source of sound or light. Unlike one-dimensional waves on a stretched string, a point source emits three-dimensional waves.

know answer to this question. It depicts a point source, S , emitting sound waves. Note that sound waves emanating from a point source spread out in 3-dimensional space over a spherical surface. To depict these spreading sound waves in 2-D, we have drawn concentric circles in Fig. 1.9 keeping S at the centre.

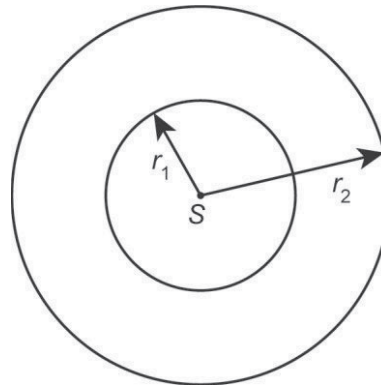


Fig. 1.9: Sound waves emitted by a point source S spread in the surrounding space.

Note from Fig. 1.9 that with increasing radius from centre S , energy transported by sound waves is distributed over a greater surface area. For example, at a distance r_1 , energy is distributed over a spherical surface of area $4\pi r_1^2$ whereas, at a distance of $r_2 (> r_1)$ the same amount of energy will be distributed over a larger spherical surface (having area $4\pi r_2^2$). So, if we assume that there is no dissipation of energy of the sound waves, the energy per unit area will be **less** at r_2 compared to the energy per unit area at r_1 . That is, energy per unit area decreases as the distance from the source increases or as we move away from the source. **In other words, the intensity of sound waves decreases as the distance from the source increases.**

Thus, for an arbitrary point located at distance r from the source, the energy of sound waves is distributed over a sphere of surface area, $A = 4\pi r^2$. Thus, from Eq. (1.25), we can write the intensity of sound at a distance r from the source as:

$$I = \frac{P}{4\pi r^2}$$

$$\text{That is } I \propto \frac{1}{r^2} \quad (1.26)$$

Eq. (1.26) is known as the **inverse square law**. It shows that the **intensity of sound waves varies inversely with square of the distance from the source**. This explains why we can be heard only up to a certain distance. Note that intensity is an extremely useful parameter because it gives us an idea about the actual amount of energy received at a given point in space.

Further, from Eq. (1.24), we recall that

$$P \propto a^2$$

where a is the amplitude of the wave. So, from Eq. (1.25), we can write

$$I \propto a^2 \quad (1.27)$$

Comparing Eqs. (1.26) and (1.27), we find that

$$a \propto \frac{1}{r} \quad (1.28)$$

Eq. (1.28) implies that **amplitude of a wave is inversely proportional to the distance of the observation point from the source**. That is, as we move away from the source, the amplitude of the wave diminishes.

You may now like to answer an SAQ.

SAQ 6 – Calculating the intensity of a wave

Calculate the intensity of sound at a distance of 2.0 m from a point source which delivers 50.0 W power.

Let us now summarize what you have learnt in this unit.

1.6 SUMMARY

Concept	Description
Wave	Wave is a disturbance which propagates in a medium progressively. Waves do not transport matter; these only transport energy and momentum.
Mechanical waves	Mechanical waves can exist only in material medium.
Types of waves	In a transverse wave, such as a wave on a stretched string, the particles of the medium oscillate in a direction perpendicular to the direction of propagation of the wave. On the other hand, in a longitudinal wave, such as sound wave in air, the particles of the medium oscillate along the direction of propagation of the wave.
Graphical representation	The graphical representation of wave motion can be done by two types of graphs: vibration graph and waveform graph. A vibration graph shows the wave behaviour at a single location along the wave path as time passes. A waveform graph is snap shot of a wave in space. A waveform graph depicts the wave behaviour at different locations along the wave path at a particular time.
Velocity of wave	The velocity, v of a wave is given as $v = f\lambda$, where f is the frequency of the wave and λ, is its wavelength. The frequency of a wave is a property of the source responsible for the generation of wave. In terms of angular frequency ω and wave number k, the wave velocity is given as $v = \omega/k$.
Mathematical representation of waves	A one-dimensional progressive wave propagating along the positive x-direction is described mathematically $y(x, t) = a \sin(\omega t - kx)$ On using the relations $\omega = 2\pi/T$ and $k = 2\pi/\lambda$, we can rewrite the above expression in other equivalent forms as $\begin{aligned} y(x, t) &= a \sin\left[\frac{2\pi}{\lambda}(\nu t - x)\right] \\ &= a \sin\left[\frac{2\pi}{T}\left(t - \frac{x}{\nu}\right)\right] \\ &= a \sin\left[2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right)\right] \end{aligned}$ A one-dimensional progressive wave travelling along the negative x-direction is given as $y(x, t) = a \sin(\omega t + kx)$

Phase of a wave

- The argument of the periodic function representing a periodic travelling wave is called the **phase angle** or simply the **phase**, $\phi(x, t)$ of the wave:

$$\phi(x, t) = \omega t - kx$$

Two points on the vibration graph of a wave separated by T or its integer multiples are in the same phase. Similarly, two points on the waveform graph of a wave separated by λ or its integer multiples are in the same phase.

Two waves are said to be **in phase** when the corresponding points of each wave reach their respective maximum or minimum displacements at the same time.

Energy of wave

- The **average mechanical energy** associated with a transverse wave on a string is given by

$$E = \frac{1}{2} \rho a^2 \omega^2 \lambda$$

where ρ is the linear mass density of the string.

Power of wave

- The average **power** (energy transported per unit time) transported by the wave is given by

$$P = \frac{1}{2} \rho a^2 \omega^2 v = 2\pi^2 \rho a^2 f^2 v$$

where v is the wave velocity.

Intensity of a wave

- The strength of a wave at a given point in space is specified by **intensity** of the wave:

$$I = \frac{P}{A}$$

where P is the power the wave and A the area of the surface at the point intercepting the wave perpendicular to the direction in which the wave travels.

Intensity of a wave (such as sound wave) at a point in space is inversely proportional to the square of the distance from the source:

$$I \propto \frac{1}{r^2}$$

where r is the distance between the source and the point under consideration. This law is known as **inverse square law**.

1.7 TERMINAL QUESTIONS

1. A transverse wave on a string has amplitude 12 mm, wavelength 2.2 m and speed 3.0 ms^{-1} . Determine the time period, frequency, angular frequency and wave number of the wave.
2. A harmonic wave on a string is given as

$$y(x, t) = (6.5 \text{ mm}) \sin[(1.42 \text{ m}^{-1})x - (3.86 \text{ s}^{-1}t)]$$

Determine the wavelength, frequency, speed, time period and direction of propagation of the wave.

3. The tension in a string is 100 N. If length of the string is 0.72 m and its mass is 5 g, determine the speed of wave along the string.
4. A string with linear mass density 0.15 kg m^{-1} is taut under a tension of 60 N. A transverse wave of amplitude 4.5 cm and angular frequency 120 s^{-1} propagates on it. Calculate the average power of the wave.
5. Hydrogen atoms produce energy due to fusion in the core of the Sun, some of this energy is radiated in the form of light, which is an electromagnetic radiation. The average intensity of sunlight at the surface of Earth is measured to be 1.33 kW m^{-2} . Take the total power produced in the Sun as equal to the power received per unit area of the Earth's surface multiplied by the surface area of a sphere of radius equal to the distance of the Earth from the Sun. Hence, calculate the mass of hydrogen that is converted per second into radiant energy in the Sun. Take the distance between the Sun and the Earth as $1.5 \times 10^8 \text{ km}$.

1.8 SOLUTIONS AND ANSWERS

Self-Assessment Questions

1. a) pulse; b) transverse; c) in east-west or west-east direction.
2. The frequency of the wave on the string is same as the frequency of the periodic motion of the end of the string. So, $f = 3.0 \text{ Hz}$. The distance between two consecutive crests of a wave is equal to the wavelength, λ of the wave.

The wave velocity is given as

$$v = f \lambda$$

So,

$$\lambda = \frac{v}{f} = \frac{6.0 \text{ ms}^{-1}}{3.0 \text{ s}^{-1}} = 2.0 \text{ m}$$

3. a) To show that the function

$$y(x, t) = (x - vt)^2$$

We need to calculate $\frac{\partial^2 y}{\partial t^2}$ and $\frac{\partial^2 y}{\partial x^2}$ and substitute the values in

Eq. (1.12)

$$\therefore \frac{\partial y}{\partial x} = 2(x - vt) \Rightarrow \frac{\partial^2 y}{\partial x^2} = 2 \quad \text{(i)}$$

$$\text{and } \frac{\partial y}{\partial t} = -2v(x - vt) \Rightarrow \frac{\partial^2 y}{\partial t^2} = 2v^2 \quad \text{(ii)}$$

Substituting Eqs. (i) and (ii) in Eq. (7.11), we find that the wave equation is satisfied.

- b) $y(x, t) = a \sin(kx - \omega t)$

$$\text{so, } \frac{\partial y}{\partial x} = k a \cos(kx - \omega t) \Rightarrow \frac{\partial^2 y}{\partial x^2} = -k^2 a \sin(kx - \omega t) \quad \text{(iii)}$$

$$\text{and } \frac{\partial y}{\partial t} = -\omega a \cos(kx - \omega t) \Rightarrow \frac{\partial^2 y}{\partial t^2} = -\omega^2 a \sin(kx - \omega t) \quad \text{(iv)}$$

On substituting Eqs. (iii) and (iv) in Eq. (1.12), we note that the wave equation is satisfied.

$$c) \quad y(x, t) = a \sin(kx + \omega t)$$

$$\text{so, } \frac{\partial y}{\partial x} = ka \cos(kx + \omega t) \Rightarrow \frac{\partial^2 y}{\partial x^2} = -k^2 a \sin(kx + \omega t) \quad (v)$$

$$\text{and } \frac{\partial y}{\partial t} = \omega a \cos(kx + \omega t) \Rightarrow \frac{\partial^2 y}{\partial t^2} = -\omega^2 a \sin(kx + \omega t) \quad (vi)$$

On substituting Eqs. (v) and (vi) in Eq. (1.12), we note that the wave equation is satisfied.

$$d) \quad y(x, t) = a e^{i(kx - \omega t)}$$

$$\frac{\partial y}{\partial x} = (ik) a e^{i(kx - \omega t)}$$

$$\frac{\partial^2 y}{\partial x^2} = -k^2 a e^{i(kx - \omega t)} \quad (vii)$$

$$\frac{\partial y}{\partial x} = (-ik) a e^{i(kx - \omega t)}$$

$$\frac{\partial^2 y}{\partial t^2} = -\omega^2 a e^{i(kx - \omega t)} \quad (viii)$$

4. Since $x_1 = 0$ and $x_2 = 2.0$ cm, we can write the given expressions for displacement as

$$y(0, t) = (0.02 \text{ cm}) \sin[(3\pi \text{ s}^{-1})t] \quad (i)$$

$$\text{and } y(2.0 \text{ cm}, t) = (0.02 \text{ cm}) \sin\left[(3\pi \text{ s}^{-1})t + \frac{\pi}{2}\right] \quad (ii)$$

On comparing Eq. (i) with Eq. (1.10d) with $x = 0$, we get amplitude, $a = 0.02$ cm. Angular frequency,

$$\omega = 3\pi \Rightarrow 2\pi f = 3\pi \Rightarrow f = 1.5 \text{ Hz}$$

Again, on comparing Eq. (ii) with Eq. (1.10d) we can write,

$$k \times (2.0 \text{ cm}) = \frac{\pi}{2}$$

$$\text{or } \frac{2\pi}{\lambda} \times (2.0 \text{ cm}) = \frac{\pi}{2} \Rightarrow \lambda = 8.0 \text{ cm}$$

Since the sign between ωt and kx is positive, the wave is travelling along negative x -direction.

The wave speed is given by

$$v = \lambda f = (8.0 \text{ cm}) \times (1.5 \text{ s}^{-1}) \Rightarrow v = 12 \text{ cm s}^{-1}$$

5. The power of a progressive wave is given by Eq. (1.24):

$$P = 2\pi^2 \rho f^2 a^2 v$$

As per the problem, $f = 50 \text{ s}^{-1}$ and $a = 8 \times 10^{-3} \text{ m}$. The linear mass density of the string is

$$\rho = \frac{0.5 \text{ kg}}{1.0 \text{ m}} = 0.5 \text{ kg m}^{-1}$$

∴ The velocity, v of the wave

$$v = \sqrt{\frac{T}{\rho}} = \sqrt{\frac{50 \text{ N}}{0.5 \text{ kg m}^{-1}}} = 10 \text{ ms}^{-1}$$

So,

$$\begin{aligned} P &= 2 \times (3.14)^2 \times (0.5 \text{ kg m}^{-1}) \times (50 \text{ s}^{-1})^2 \times (8 \times 10^{-3} \text{ m})^2 \times (10 \text{ ms}^{-1}) \\ &= 15.8 \text{ W} \end{aligned}$$

6. We note from Sec. 1.5 that the intensity of wave is

$$I = \frac{P}{4\pi r^2}$$

As per the problem, $P = 50.0 \text{ W}$ and $r = 2.0 \text{ m}$.

So,

$$\begin{aligned} I &= \frac{50.0 \text{ W}}{4 \times 3.14 \times (2.0 \text{ m})^2} \\ &= 1.0 \text{ W m}^{-2} \end{aligned}$$

Terminal Questions

1. Speed of wave, $v = f\lambda$. So,

$$f = \frac{v}{\lambda} = \frac{3.0 \text{ ms}^{-1}}{2.2 \text{ m}} = 1.36 \text{ s}^{-1}$$

$$\text{Time period, } T = \frac{1}{f} = \frac{1}{1.36 \text{ s}^{-1}} = 0.74 \text{ s}$$

$$\text{Angular frequency, } \omega = 2\pi f = 2 \times 3.14 \times 1.36 \text{ s}^{-1} = 8.54 \text{ s}^{-1}$$

$$\text{Wave number } k = \frac{2\pi}{\lambda} = \frac{2 \times 3.14}{2.2 \text{ m}} = 2.85 \text{ m}^{-1}$$

2. The given expression for the wave is

$$[y(x, t) = (6.5 \text{ mm}) \sin[(1.42 \text{ m}^{-1})x - (3.86 \text{ s}^{-1})t]$$

Comparing with standard equation for wave travelling in positive x -direction

$$y(x, t) = a \sin(kx - \omega t)$$

We have, $k = 1.42 \text{ m}^{-1}$ and $\omega = 3.86 \text{ s}^{-1}$

So,

$$\lambda = \frac{2\pi}{k} = \frac{2 \times 3.14}{1.42 \text{ m}^{-1}} = 4.42 \text{ m}$$

$$f = \frac{\omega}{2\pi} = \frac{3.86 \text{ s}^{-1}}{2 \times 3.14} = 0.61 \text{ s}^{-1}$$

$$\text{Speed, } v = f\lambda = (0.61 \text{ s}^{-1}) \times (4.42 \text{ m}) = 2.70 \text{ ms}^{-1}$$

Direction of propagation is positive x -direction.

3. We know that the speed of transverse wave on a string is given as

$$v = \sqrt{T / \rho}$$

$$T = 100 \text{ N} \quad \text{and} \quad \rho = \frac{5 \times 10^{-3} \text{ kg}}{0.72 \text{ m}} = 6.9 \times 10^{-3} \text{ kg m}^{-1}$$

So,

$$v = \sqrt{\frac{100 \text{ N}}{6.9 \times 10^{-3} \text{ kg m}^{-1}}} = 120.39 \text{ ms}^{-1}$$

4. We know that the average power of a wave on a string is given as

$$P = \frac{1}{2} \rho \omega^2 a^2 v \quad (\text{i})$$

The velocity, v of the wave is given by

$$v = \sqrt{\frac{T}{\rho}} = \sqrt{\frac{60 \text{ N}}{0.15 \text{ kg m}^{-1}}} = 20 \text{ ms}^{-1}$$

Angular frequency, $\omega = 120 \text{ s}^{-1}$; Amplitude, $a = 4.5 \times 10^{-2} \text{ m}$

On substituting the values of these parameters in Eq. (i), we get the average power,

$$\begin{aligned} P &= \frac{1}{2} \times (0.15 \text{ kg m}^{-1}) \times (120 \text{ s}^{-1})^2 \times (4.5 \times 10^{-2} \text{ m})^2 \times (20 \text{ ms}^{-1}) \\ &= 43.74 \text{ W} \end{aligned}$$

5. We are given that the intensity (I) of sunlight on the Earth's surface is

$$\text{Intensity } I = 1.33 \text{ kW m}^{-2} = 1330 \text{ W m}^{-2}$$

According to the question, the total power P produced in the Sun is

$$P = \text{Intensity} \times (4\pi R^2)$$

where R is the distance between the Sun and the Earth. Substituting the values of I and R , we get

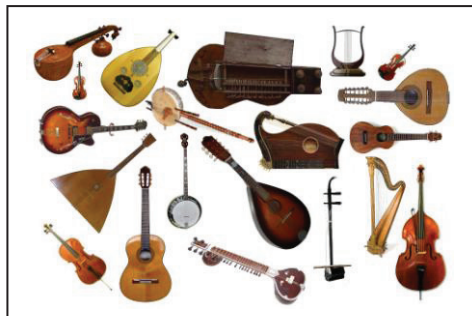
$$P = (1330 \text{ W m}^{-2}) \times (4 \times 3.14 \times (1.5 \times 10^{11} \text{ m})^2) = 4.356 \times 10^{26} \text{ W}$$

Since power is the energy radiated per second, we can say that the Sun radiates $4.356 \times 10^{26} \text{ J}$ per second. This signifies the amount of energy produced per second in the core of the Sun due to fusion. To determine how much mass is converted per second into energy due to fusion, we use Einstein's mass-energy equivalence relation:

$$E = mc^2$$

$$\therefore m = \frac{E}{c^2} = \frac{4.356 \times 10^{26} \text{ J}}{(3 \times 10^8 \text{ ms}^{-1})^2} = 4.08 \times 10^9 \text{ kg}$$

Thus, $4.08 \times 10^9 \text{ kg}$ hydrogen burns every second for the intensity of solar radiation we receive on the Earth. Obviously, this is a huge quantity of hydrogen. Yet, the Sun has been shining for billions of years and will continue as such for long. But a time will come when entire hydrogen will be burnt up. Can you imagine what would happen to life on our planet thereafter?



UNIT 2

SUPERPOSITION OF WAVES

The superposition of waves giving rise to standing waves is basic phenomena associated with the working of musical instruments.

(Source of picture: wikimedia.org)

Structure

- | | | | |
|-----|-------------------------------------|-----|--------------------------------|
| 2.1 | Introduction | 2.4 | Beats |
| | Expected Learning Outcomes | 2.5 | Wave Groups and Group Velocity |
| 2.2 | Principle of Superposition of Waves | 2.6 | Summary |
| 2.3 | Standing Waves | 2.7 | Terminal Questions |
| | Standing Waves on a String | 2.8 | Solutions and Answers |

STUDY GUIDE

In the previous unit, you have revised basic concepts related to wave motion and its mathematical representation. In this unit, you will learn how to apply the superposition principle to waves. You will learn how a variety of phenomena associated with waves such as standing waves, beats and wave groups can be explained by applying the superposition principle.

Superposition of waves is used to understand interference and diffraction of light. Thus, you should internalise the concepts discussed in this unit. Further, while studying the unit, the mathematical treatment of the formation of standing waves, beats and wave groups may seem similar. It is, therefore, necessary to focus on the apparently minor but significant differences in the mathematical analysis of each phenomenon. You will be using simple trigonometric relations, which you have learnt in your +2 mathematics. We, therefore, advise you to refresh your knowledge of basic trigonometry. Also, you are advised to focus on the physical significance of the mathematical results obtained due to superposition of waves under different conditions. Try to solve SAQs and TQs yourself to check your understanding of the concepts discussed in the unit.

“Ask the right questions, and nature will open the doors to her secrets.”

C.V. Raman

2.1 INTRODUCTION

In the previous unit, you have learnt that a progressive wave transports energy. When such a progressive wave meets a rigid boundary, a part of the wave is reflected and a part is transmitted. In case of one-dimensional progressive waves, such as transverse waves on a string, the reflected waves travel in the same medium but in a direction opposite to the direction of propagation of the incident wave. This is a situation where we have two waves in the same region of space at the same time.

In fact, there are many other situations where two or more waves exist in space simultaneously. For example, raindrops strike the surface of a pond and generate circular waves that meet each other. Similarly, you may find yourself in a situation when your ears are receiving sound waves generated by two persons speaking at the same time. So you may ask: **How do we analyse a phenomenon which arises due to the simultaneous existence of two, or more than two waves at a given point in space and time?** To analyse the behaviour of waves in such situations, we use the **principle of superposition**.

As such, the principle of superposition is not a new concept for you. In Unit 17 of the first semester course entitled Mechanics (BPHCT-131), you have learnt how to apply this principle to analyse the motion of a particle on which two simple harmonic oscillations acted simultaneously. In the case of oscillations, the superposition principle led us to conclude that the resultant motion of the particle depended on the amplitudes, frequencies, and phase difference of the superposing oscillations. The resultant was also determined by the directions of superposing oscillations – whether collinear or orthogonal. In the present unit, you will learn how to use the principle of superposition for waves and understand many related concepts such as standing waves, beats, and wave groups.

We begin this unit by revisiting the principle of superposition in Sec. 2.2 and discuss its implications for waves. Further, as you know, a wave is characterised by its amplitude, frequency (or wavelength) and initial phase. In Sec. 2.2, you will also learn that superposition of waves leads to a variety of outcomes depending upon the relation between the characteristic parameters of the superposing waves.

In Sec. 2.3, you will learn about formation of **standing waves**, also known as **stationary waves**, due to superposition of two identical waves (same frequency and same amplitude) travelling in opposite directions. In standing waves, energy is confined to a limited region of space. This phenomenon finds useful manifestation in the form of music – one of the most charming aspects of life – produced by wind pipes (such as flute, trumpet, *shehnai*, etc.) and string instruments (such as *sitar*, guitar, violin, etc.).

Yet another important phenomenon resulting due to superposition of waves of nearly equal frequencies is known as **beats**. The most vivid demonstration of

beats is observed in the case of sound waves, as periodic changes in the intensity of sound. You might have observed that in an orchestra, a violinist tightens up or loosens the pegs of her instrument while *tuning* with the instruments of other musicians. By adjusting the pegs, the violinist changes the frequency of the sound produced by her instrument so that different *notes* may be generated. *Tuning* two instruments is actually a way to eliminate *beats*. You will learn about beats in detail in Sec. 2.4.

In Sec. 2.5, we generalise the phenomenon of beats and introduce the concept of **wave groups** or **wave packets**, which arise due to superposition of two or more waves of slightly different frequencies. The concept of wave packet is of great importance in quantum mechanics where microscopic particles such as electrons and protons are represented by localised waves or wave packets.

In the next unit, you will study about acoustic waves.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ explain the principle of superposition for waves;
- ❖ apply the superposition principle to explain the formation of standing waves;
- ❖ derive the mathematical expression for a standing wave on a string;
- ❖ explain the concept of node and antinode in a standing wave;
- ❖ obtain expressions for the frequency of normal modes;
- ❖ explain the phenomenon of beats;
- ❖ explain the formation of wave groups or wave packets; and
- ❖ derive expressions for group velocity and phase velocity.

2.2 PRINCIPLE OF SUPERPOSITION OF WAVES

From Block 4 of the first semester course BHPCT-131, you will recall that harmonic oscillations obey the principle of superposition. Recall what happens when two or more simple harmonic oscillations act on a particle simultaneously? You have learnt that the resultant displacement of the particle at any instant of time is given by the algebraic sum of displacements due to individual oscillations acting separately.

You may now like to know: **Do waves also obey the principle of superposition?** The answer is, yes. When two or more waves travel in the same region in space, independent of one another, the resultant displacement of a particle of the medium at any given time is equal to the algebraic sum of the displacements due to individual waves. This is known as the **principle of superposition of waves**. That is, each wave influences the particles of the medium independent of the influence of other waves.

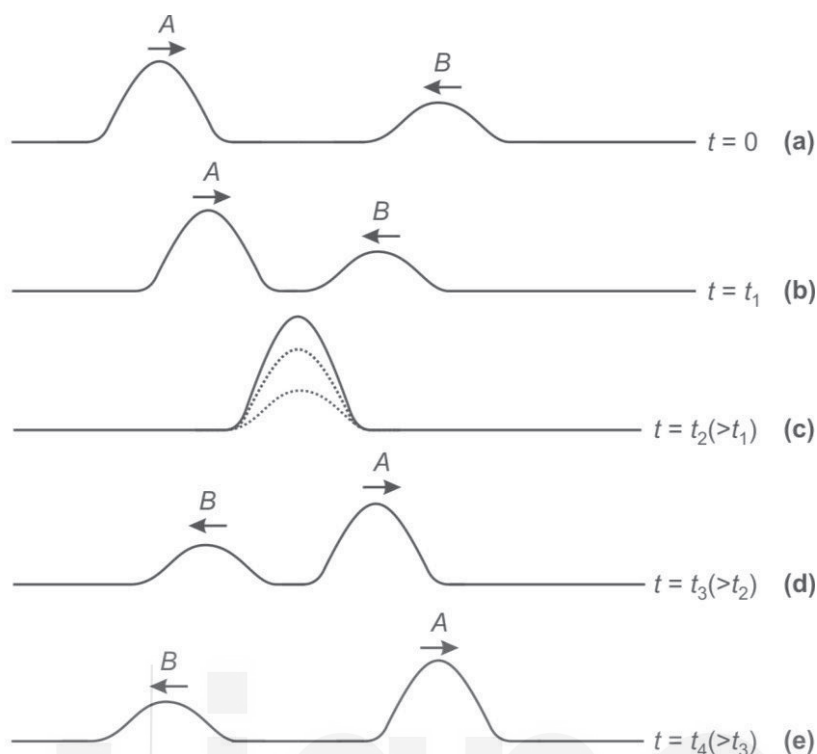


Fig. 2.1: Snapshots of two pulses A and B, having positive amplitudes but propagating in opposite directions, at different times: a) Two pulses approach each other; b) the two pulses meet each other; c) the superposition region where resultant amplitude increases; d) the pulses moving just beyond the superposition region; e) pulses much beyond the superposition or cross-over region.

To understand the principle of superposition of waves, it is convenient to consider two pulses A and B travelling on a string in opposite directions, as shown in Fig. 2.1a. Note that both the pulses have different but positive amplitudes. As time passes, these meet one another, as shown in Fig. 2.1b. Fig. 2.1c shows the superposition region, the region where the two pulses exist simultaneously. **Note that in this region, the amplitude (or displacement of the particles in the cross-over region) of the resultant pulse increases and it is equal to the sum of the amplitudes of the two superposing pulses.** Further, beyond the cross-over region, these two pulses continue to move in their respective original directions, as shown in Figs. 2.1d and 2.1e. It means that each pulse behaves as if it has not been influenced by the other in any way, i.e., they act completely independently.

Now, refer to Fig. 2.2a which shows two pulses (A and B) approaching each other with their amplitudes oppositely directed. When they cross each other, the resultant amplitude is the algebraic sum of the amplitudes of the two pulses (Fig. 2.2c). Note that the resultant amplitude (depicted by solid line in Fig. 2.2c in the cross-over region,) is positive but less than the amplitude of the pulse travelling from left to right. (The resultant amplitude is diminished because the pulse moving from right to left has negative amplitude.) Beyond the superposition region, the two pulses propagate along their respective directions as unaffected individual pulses (Fig. 2.2d and 2.2e).

Note that we have taken pulse(s) for illustrating and explaining superposition of waves. It is easier to depict as well as explain. Since a pulse is a part of the waveform representing a wave, the explanations and results will be valid for waves as well.

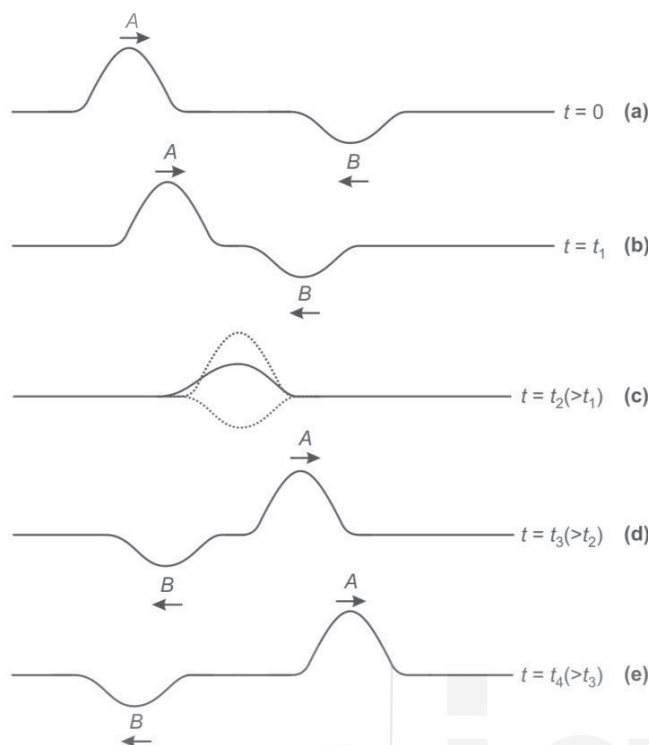


Fig. 2.2: Snapshots of two pulses A and B – one with positive amplitude and the other with negative amplitude – propagating in opposite directions at different times: a) Two pulses approach each other; b) the two pulses meet each other; c) in the superposition, the resultant amplitude is positive but smaller in magnitude compared to pulse A; d) the pulses moving just beyond the superposition region; e) much beyond the superposition region, the pulses move in their respective original directions.

Our discussion of superposition of waves has so far been qualitative.

Moreover, we have considered a rather simple case of two superposing pulses travelling on the same string. We considered this for its ease of depiction and understanding. **As such, the principle of superposition is valid for all kinds of waves wherever two or more waves exist simultaneously in a given region of space.** Let us now apply the superposition principle for waves and express it mathematically.

If $y_1(x,t)$ and $y_2(x,t)$ represent the displacement of two waves separately, then $y_1(x,t) + y_2(x,t)$ is the resultant wave when the two waves superpose. That is, the expression for the displacement of the resultant wave, $y(x,t)$ is given by

$$y(x,t) = y_1(x,t) + y_2(x,t) \quad (2.1)$$

In Unit 1 of this block, you have learnt that a wave is characterised by its amplitude, angular frequency, wave number and phase. So, you may like to know: **How do we determine these parameters of the resultant wave?** When two waves superpose, their resultant is determined by the relation between the corresponding parameters of the waves and their direction of propagation. Some possible cases are that waves have:

1. Same angular frequencies and phases but different amplitudes

$$(\omega_1 = \omega_2, \phi_1 = \phi_2 = 0, a_1 \neq a_2)$$

2. Same amplitudes and frequencies but finite phase difference

$$(a_1 = a_2, \omega_1 = \omega_2)$$

3. Identical waves travelling in opposite directions.
4. Same amplitudes and phases but different frequencies.

Let us apply the superposition principle to each case.

1. **Superposition of two waves having the same angular frequency, same phase but different amplitudes and both propagating along the positive x -direction.**

Mathematically, we represent these waves as

$$y_1(x, t) = a_1 \sin(kx - \omega t)$$

and $y_2(x, t) = a_2 \sin(kx - \omega t)$

The waveforms for these two waves at a given instant is shown by dashed and dotted curves in Fig. 2.3. Now, using the superposition principle, we can write the expression for the resultant wave as

$$\begin{aligned} y(x, t) &= a_1 \sin(kx - \omega t) + a_2 \sin(kx - \omega t) \\ &= (a_1 + a_2) \sin(kx - \omega t) \end{aligned} \quad (2.2)$$

From Eq. (2.2), we note that the amplitude of the resultant wave is equal to the sum of the amplitudes of individual waves. **However, the frequency and phase of the resultant wave remain unchanged.** The resultant waveform is also shown by a solid curve in Fig. 2.3.

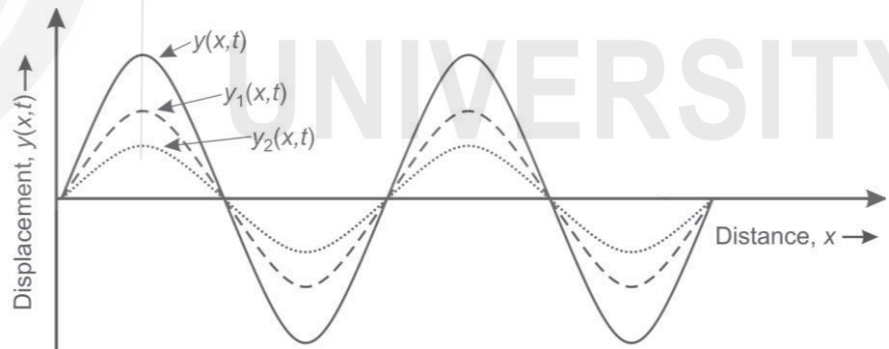


Fig. 2.3: Superposition of two in-phase waves of the same frequency but unequal amplitudes moving along the same direction.

2. **Superposition of two waves having a finite phase difference but same amplitude and frequency and propagating along the positive x -direction.**

We represent these waves as

$$y_1(x, t) = a \sin(kx - \omega t)$$

and $y_2(x, t) = a \sin(kx - \omega t + \phi)$

When such waves superpose, we observe the phenomenon of **interference** provided the phase difference, ϕ between the waves remains

constant. The phenomenon of interference is more vividly seen with light waves. You will learn about it in detail in Units 6 and 7 of this course.

3. Superposition of identical waves propagating in opposite directions.

The waves can be represented mathematically as

$$y_1(x, t) = a \sin(kx - \omega t) \quad (2.3)$$

and $y_2(x, t) = a \sin(kx + \omega t) \quad (2.4)$

Note that Eq. (2.3) represents a wave propagating along the positive x -direction while Eq. (2.4) represents a wave propagating in the negative x -direction. The resultant of the superposition of these two waves in a bound region gives rise to **standing waves**, also known as **stationary waves**. You will learn about standing waves and their characteristics in detail in Sec. 2.3.

4. Superposition of two waves of slightly different frequencies but same amplitude and phase and propagating along the same direction.

The waves can be represented mathematically as

$$y_1(x, t) = a \sin(kx - \omega_1 t) \quad (2.5)$$

and $y_2(x, t) = a \sin(kx - \omega_2 t) \quad (2.6)$

When such waves superpose, **beats** are produced. Although the phenomenon of beats is observed for all kinds of waves, we can hear them when sound waves superpose. **The beats are heard as periodic variation in the loudness of sound.** (We hear louder sound followed by a lower sound, with some time period.) In addition to the phenomenon of beats, superposition of waves of slightly different frequencies gives rise to the formation of **wave groups** or **wave packets**. The concept of wave packet has great significance in quantum mechanics. You will learn about the phenomenon of beats and the concept of wave packets in Sec. 2.4.

From the above qualitative discussion, you can note that the superposition of two or more waves can give rise to a variety of phenomena. In other words, we can say that a variety of natural phenomena can be understood on the basis of superposition principle.

It is, however, important to mention here that superposition principle can be applied only when the amplitudes of the superposing waves are small. It is so because a small-amplitude wave does not influence the properties of the medium significantly. In fact, the changes in the medium due to propagation of a small-amplitude wave are so small that the other waves are not affected by it. However, when the amplitude of the waves is large, such as that of shock waves generated by supersonic aircrafts, the superposition principle is not adequate to know the resultant.

In order to appreciate the principle of superposition, you should go through the following example carefully.

EXAMPLE 2.1: SUPERPOSITION OF TWO PULSES MOVING IN OPPOSITE DIRECTIONS

Consider two pulses on a string which are moving towards each other as shown in Fig. 2.4 below at the instant $t = 0$. Each pulse moves with a speed of 2 cm s^{-1} . Sketch the pulses at $t = 1.0\text{ s}$ and 2.0 s .

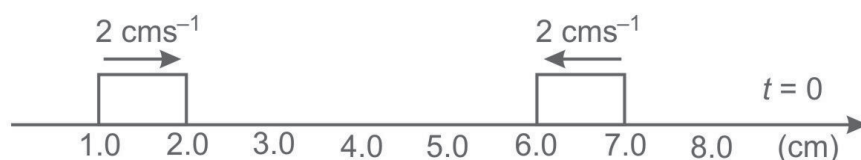


Fig. 2.4

SOLUTION ■ From the initial location of pulses, we note that at $t = 0$, the two pulses are situated between 1.0 cm and 2.0 cm and between 6.0 cm and 7.0 cm, respectively. Now these pulses are moving towards each other with a speed of 2 cm s^{-1} . So, the pulse moving to the right will be located between 3.0 cm and 4.0 cm after 1 s and the pulse moving towards the left will be located between 4.0 cm and 5.0 cm. On superposition, they will merge and appear as one big pulse between 3.0 cm and 5.0 cm as shown in Fig. 2.5.

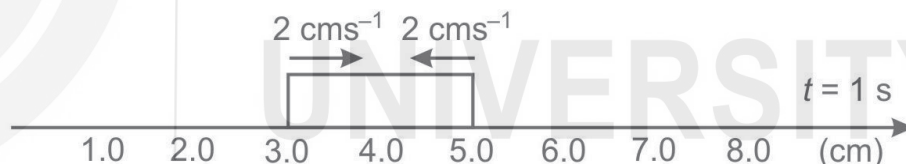


Fig. 2.5

After 2.0 s, these pulses will cross each other. The pulse moving to the right will be located between 5.0 cm and 6.0 cm, and the pulse moving towards the left will be located between 2.0 cm and 3.0 cm. These are depicted in Fig. 2.6.

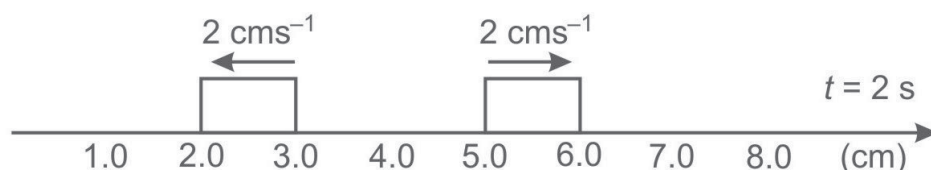


Fig. 2.6

Now you may like to solve the following SAQ.

SAQ 1 – Superposition of pulses

At time $t = 0$, the two pulses on a string approaching each other are shown in Fig. 2.7 below. The pulses are moving with a speed 1.0 cm s^{-1} . Sketch the pulses at $t = 1.0 \text{ s}$, 2.0 s and 4.0 s .



Fig. 2.7

Before proceeding further, let us recapitulate what you have learnt in this section.

Recap

- According to the superposition principle, when two or more waves travel in the same region in space, the resultant displacement of a particle in that region is equal to the algebraic sum of its displacement due to individual waves.
- The superposition principle is valid only for small amplitude waves.
- The phenomena of interference, beats and standing waves arise due to superposition of waves.
- When two waves having the same amplitude and phase but slightly different frequencies superpose, beats are produced. Beats can easily be heard for sound waves.

Let us now discuss in detail the superposition of waves travelling in opposite directions and leading to the formation of standing waves.

2.3 STANDING WAVES

You now know that superposition of two identical sinusoidal waves (of the same frequencies and amplitudes) travelling in opposite directions gives rise to **standing waves**. Let us try to visualise such waves and their superposition in an actual physical situation. Consider a stretched string and think: How can we obtain two identical waves with same amplitudes and frequencies but moving in opposite directions? In practice, we obtain such waves using the phenomenon of reflection; it is quite difficult otherwise.

You may now ask: **What type of boundary gives rise to a reflected wave which is identical to the incident wave?** An identical wave is generated when a wave is reflected from a **fixed** (or **rigid**) **boundary** such as a wall and the close-end of an organ pipe or a **free boundary** such as free end of a string and the open end of an organ pipe.

Refer to Fig. 2.8a which shows an incident wave on a stretched string completely reflected by a fixed boundary (wall). Since the incident and

reflected waves travel in the same medium (string), their speeds and frequencies remain the same. So, we find that if a sinusoidal wave is completely reflected by a boundary, the reflected wave (travelling in the opposite direction) will have the **same amplitude, frequency and speed** as that of the incident wave.

Note that a wave may be partially reflected and partially transmitted when it is incident on a boundary separating two different media.

Similarly, an identical reflected wave is obtained when reflection takes place at a **free boundary such as the free end of a string** (Fig. 2.8b).

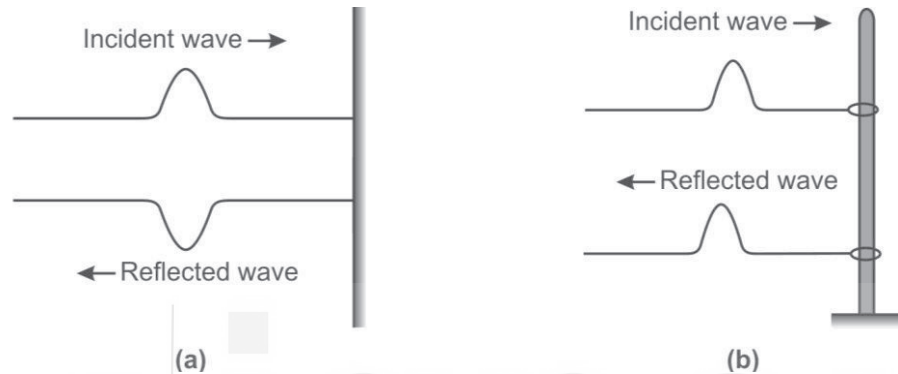


Fig. 2.8: Reflection of a wave propagating on a stretched string from a) fixed boundary; b) free boundary.

We hope you now know how to generate two identical waves travelling in opposite directions. When such identical waves travelling in opposite directions superpose, we obtain **standing waves**. You might have observed that in stringed musical instruments such as *sitar*, guitar and *veena*, both ends of a string are fixed. When the string is plucked, standing waves are generated due to the superposition of incident and reflected transverse waves between its fixed ends.

So far, we have discussed the conditions under which standing waves can be formed on a stretched string. It is, however, important to mention here that standing waves are also formed due to superposition of longitudinal waves in air columns, such as an organ pipe. Note that in the case of the string, transverse displacement of the particles of the string leads to formation of transverse waves. But in an organ pipe, longitudinal waves are generated due to longitudinal displacement of air molecules giving rise to compressions and rarefactions. However, the nature of mathematical functions representing either a transverse or a longitudinal wave is the same. Therefore, the mathematical analysis of standing waves due to superposition of transverse waves will be valid for longitudinal waves as well. We shall confine our discussion to the formation of standing waves on a string.

2.3.1 Standing Waves on a String

Consider a stretched string subject to uniform tension T and fixed at its ends, $x = 0$ and $x = L$, as shown in Fig. 2.9a. Suppose that a wave travels in the positive x -direction from the end $x = 0$ and is reflected at the fixed end at $x = L$. The incident and reflected waves are shown in Fig. 2.9b.

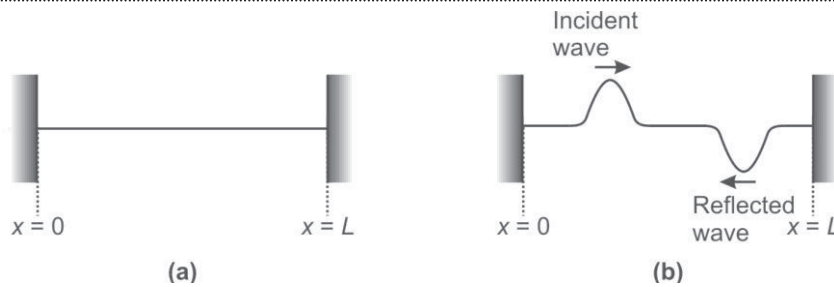


Fig. 2.9: a) A stretched string under tension T and fixed at $x = 0$ and $x = L$; b) the waves travelling in positive x -direction and negative x -direction. Note that for ease of visualisation, we have shown only a pulse of the waves travelling in opposite directions.

We can represent the wave travelling along the positive x -direction by

$$y_1(x, t) = a \sin(kx - \omega t) \quad (2.7)$$

And, we can represent the wave travelling in the negative x -direction as:

$$y_2(x, t) = a \sin(kx + \omega t) \quad (2.8)$$

When these two waves superpose, the resultant wave is given by

$$\begin{aligned} y(x, t) &= y_1(x, t) + y_2(x, t) \\ &= a \sin(kx - \omega t) + a \sin(kx + \omega t) \end{aligned}$$

or
$$y(x, t) = (2a \sin kx) \cos \omega t \quad (2.9)$$

Eq. (2.9) represents the resultant wave when two identical waves travelling in opposite directions superpose. Let us pause for a moment and ask: **What does Eq. (2.9) convey physically?** We can make the following inferences:

- The first and the most significant aspect of Eq. (2.9) is that it **does not represent a travelling wave. How do we say so?** It is because it does not contain a factor $(kx - \omega t)$ or $(kx + \omega t)$. Thus, we find that the superposition of two identical waves propagating in opposite directions gives rise to a wave which does not propagate; that is, **the resultant wave is stationary or standing**. You may like to know: How does the waveform of the standing wave look like? Fig. 2.10 shows a snapshot of the waveform representing a standing wave on a string.

$$\begin{aligned} y(x, t) &= y_i(x, t) + y_r(x, t) \\ &= a \sin(kx - \omega t) \\ &\quad + a \sin(kx + \omega t) \\ &= a[\sin kx \cos \omega t - \sin \omega t \cos kx + \sin kx \cos \omega t + \sin \omega t \cos kx] \\ &= (2a \sin kx) \cos \omega t \end{aligned}$$

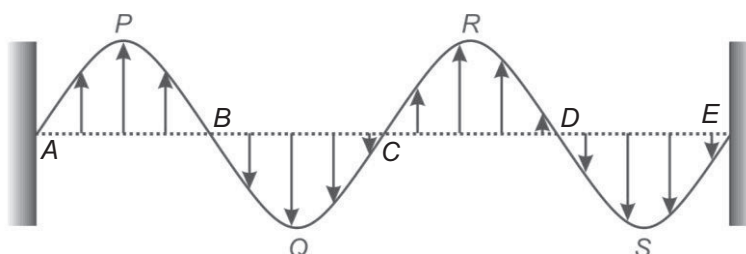


Fig. 2.10: A snapshot of a standing wave on a string fixed at both ends. The arrow-heads indicate the amplitudes with which various elements of the string vibrate.

It is important to mention here that Fig. 2.10 depicts the shape of the string (on which standing wave is formed) **at a given instant of time. This should not give you the impression that the string has such a shape**

at all times. To appreciate this point, refer to Fig. 2.11, which shows the snapshots of two waves moving in opposite directions and their resultant – the standing wave – at different times between $t = 0$ and $t = T$, the time period of the wave. Note that the shape of the string gradually changes with passage of time from $t = 0$ to $t = T$. (At $t = T/4$ and $t = 3T/4$, the string is straight.)

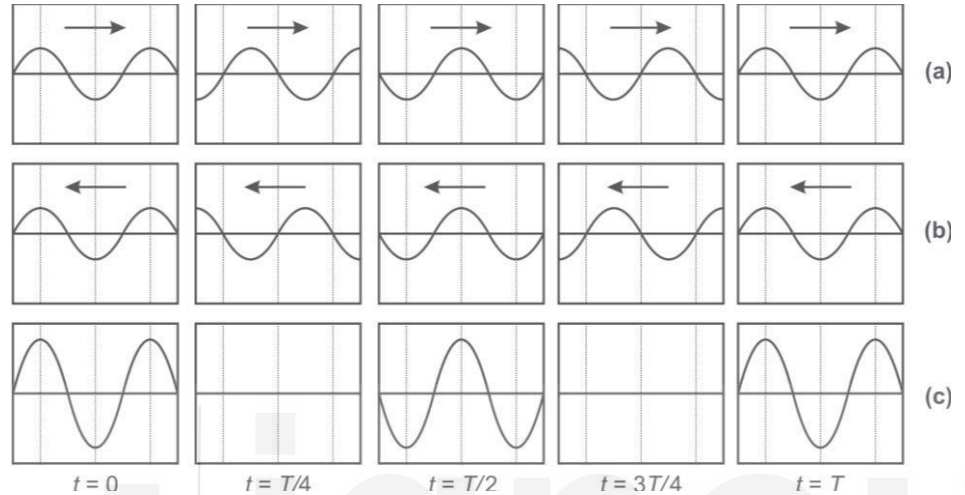


Fig. 2.11: Snapshots of standing waves at different times; a) wave moving in positive x-direction; b) wave moving in negative x-direction; c) resultant standing waves.

- ii) Further, from Eq. (2.9), we note that the resultant wave represents a simple harmonic motion whose amplitude given by $2a \sin kx$ varies from point to point. Note that the simple harmonic motion is unique. Every element of the string executes SHM with same frequency, ω . **However, the amplitude, $2a \sin kx$, of oscillatory motion of each element depends on the location x of the element of the string.** The position dependent amplitude of various elements of the string are depicted by arrow-heads of different heights in Fig. 2.10.
- iii) To check whether or not the expression for displacement of the resultant wave, given by Eq. (2.9), is consistent with the physical condition (that displacement is zero at $x = 0$), we apply the **boundary condition**

$$[y(x, t)]_{x=0} = 0$$

on Eq. (2.9). You can see that Eq. (2.9) satisfies this boundary condition.

- iv) Note that the amplitude, $2a \sin kx$ of an element of the medium (string) will be minimum (zero) for all such values of x which satisfy the condition

$$\sin kx = \sin n\pi, \quad n = 0, 1, 2, \dots$$

This means that for such values,

$$kx = n\pi \quad \Rightarrow \quad \frac{2\pi}{\lambda} x = n\pi$$

$$\text{or} \quad x = \frac{n\lambda}{2}, \quad n = 0, 1, 2, 3, \dots \quad (2.10)$$

The points on the string for which x is given by Eq. (2.10) represent the points where the amplitude of the standing wave is zero. Such points on a

standing wave are called **nodes** and are shown by points A, B, C, D and E in Fig. 2.10.

- v) We also note that the amplitude, $2a \sin kx$ of the standing wave will be maximum when

$$\sin kx = \pm 1 = \sin(2n+1)\frac{\pi}{2} \quad n = 0, 1, 2, \dots$$

or $kx = (2n+1)\frac{\pi}{2}$

$$\Rightarrow x = \frac{(2n+1)\lambda}{4} \quad n = 0, 1, 2, \dots \quad (2.11)$$

The points on the string corresponding to the values of x given by Eq. (2.11) are called **antinodes**. These are shown by points P, Q, R and S in Fig. 2.10.

- vi) Since the string is of finite length L and is fixed at both the ends, the displacement at the fixed ends of the string must be zero at all times. That is, the expression for the resultant standing wave given by Eq. (2.9) should also satisfy the following **boundary condition**:

$$[y(x, t)]_{x=L} = 0$$

that is, the displacement at the other end ($x = L$) of the string, which is fixed, must be zero at all times. When we apply this boundary condition to Eq. (2.9), we get

$$\sin kL = 0 = \sin n\pi \quad n = 0, 1, 2, \dots$$

$$\Rightarrow k_n L = n\pi$$

$$\text{or } L = \frac{n\lambda_n}{2} \quad n = 1, 2, 3, \dots \quad (2.12)$$

Eq. (2.12) expresses a very significant result. But, before we discuss its significance, let us understand why have we added a subscript n with k and λ and why have we dropped $n = 0$ as one of the possible values of n .

The subscript n signifies that the wave number k and hence the wavelength λ can take only discrete set of values associated with integer n . Further, $n = 0$ is a permissible value technically. But, for $n = 0$, the resultant displacement, $y(x, t)$ is identically zero and will correspond to a trivial situation. Thus, the value of n effectively starts with 1. The physical significance of Eq. (2.12) is that it enables us to understand the concept of normal modes.

Normal Modes

Since the string is fixed at both ends (Fig. 2.9), the boundary condition $[y(x, t)]_{x=L} = 0$ applied to the resultant standing wave [Eq. (2.9)] gives us Eq. (2.12). Note that Eq. (2.12) implies that the **length L of the string must be an integral multiple of $\lambda/2$** . In other words, on a string of a given length, L , fixed at both ends, standing waves of only certain wavelengths, determined by Eq. (2.12), can be sustained or allowed. **That is, stationary or standing waves of only certain values of wavelength are possible.** The allowed values of wavelength (or frequency) of the standing waves determine the **normal modes** of vibration of the string fixed at both ends.

The discrete allowed values of wavelength, λ of the standing waves are given by [Eq. (2.12)]:

Note that here we are dealing with **three types of amplitudes**. First is the **amplitude, a , of the superposing waves**. Second is the amplitude, $2a \sin kx$ of simple harmonic motion of **each element of the medium** (string) which depends on its location. This implies that the elements of the string on which standing waves are formed have different amplitudes of oscillation. This is quite unlike travelling waves in which each element of the medium oscillates with the same amplitude, a and the amplitude of the wave is the same as the amplitude of the oscillation of elements. And, thirdly, the **amplitude of the standing wave** which is $2a$.

$$\lambda_n = \frac{2L}{n} \quad n = 1, 2, 3, \dots \quad (2.13)$$

And, the corresponding values of the frequencies can be obtained using the relation $v = f\lambda$. Thus, the **normal mode frequency** (allowed values of the frequency) of the standing waves on a string fixed at both ends are given by

$$f_n = \frac{nv}{2L}; \quad n = 1, 2, 3, \dots \quad (2.14)$$

The expression for normal mode frequencies can also be written in terms of tension in the string and mass per unit length. To do so, we write the velocity of a transverse wave on a stretched string as

$$v = \sqrt{\frac{T}{\rho}}$$

where T is the tension in the string and ρ is the mass per unit length of the string. Substituting the above value of v in Eq. (2.14), we get

$$f_n = \frac{n}{2L} \sqrt{\frac{T}{\rho}} \quad n = 1, 2, 3, \dots \quad (2.15)$$

The frequency corresponding to $n = 1$ is called **fundamental frequency** of the standing wave and is given by [Eqs. (2.14), (2.15)]:

$$f_1 = \frac{v}{2L} = \frac{1}{2L} \sqrt{\frac{T}{\rho}} \quad (2.16)$$

A mode of vibration refers to the 'way of vibration' of the string characterised by the frequency of vibration or the wavelength of the standing wave.

Note that fundamental frequency is the minimum frequency with which a string fixed at both ends can vibrate. For this reason, the fundamental frequency is also called the **first harmonic**.

You may now ask: **What are the other normal modes of vibration of the string?** In general, the string can vibrate in n ($= 1, 2, 3, \dots$) number of segments and the corresponding frequency of vibration, f_n , is given by Eq. (2.14) or Eq. (2.15). The frequencies f_n (for $n \geq 2$) are called **higher harmonics** or **overtones**. **Note that frequencies of higher harmonics are integral multiples of the fundamental frequency.** Some of the normal modes of vibration of a string fixed at both ends are depicted in Fig. 2.12. Note that the string gets divided into larger number of loops as higher harmonics are excited.

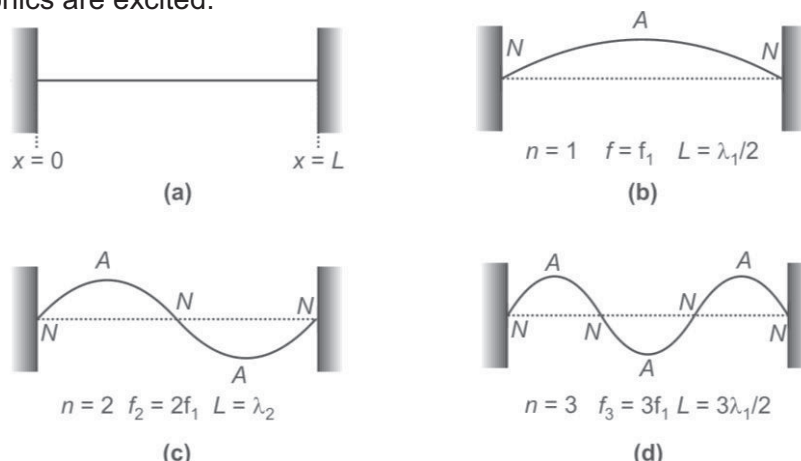


Fig. 2.12: a) A stretched string of length L fixed at both ends; waveforms for b) the fundamental mode or the first harmonic ($n = 1$); c) the second harmonic ($n = 2$); d) the third harmonic ($n = 3$).

From the above discussion, you may conclude that waves of only certain discrete values of wavelength (or frequency) can be sustained or generated on a string fixed at both ends. This discreteness is referred to as **'quantisation' of wavelength (or frequency)**. Note that it is a direct consequence of the boundary conditions at the fixed ends of the string.

To appreciate the concept of standing waves, you should now go through the following examples.

EXAMPLE 2.2: COMPONENT WAVES OF A STANDING WAVE

Two waves of the same frequency, velocity and amplitude travelling in opposite directions on a string fixed at both ends are represented as:

$$y_1(x, t) = a \sin(kx - \omega t)$$

and $y_2(x, t) = a \sin(kx + \omega t)$

On superposition, these give rise to a standing wave given by

$$y(x, t) = 2 \sin \frac{\pi x}{4} \cos 40 \pi t \text{ cm}$$

- Determine the equations representing the component waves $y_1(x, t)$ and $y_2(x, t)$;
- Calculate the distance between two successive nodes.

SOLUTION ■ The resultant wave due to superposition of two waves of same amplitude and frequency and travelling in opposite directions is given by [Eq. (2.9)]:

$$y(x, t) = (2a \sin kx) \cos \omega t \quad (i)$$

Comparing Eq. (i) with the expression for the resultant wave given in the problem:

$$y(x, t) = 2 \sin \frac{\pi x}{4} \cos 40 \pi t$$

We get, $k = (\pi/4)$, $\omega = 40\pi$ and $a = 1 \text{ cm}$

So, the equations representing the actual component waves will have these values of parameters a , k and ω . Thus, we get the expressions for component waves as:

$$y_1(x, t) = (1.0 \text{ cm}) \sin \left(\frac{\pi x}{4} - 40 \pi t \right)$$

and $y_2(x, t) = (1.0 \text{ cm}) \sin \left(\frac{\pi x}{4} + 40 \pi t \right)$

b) Since $k = (\pi/4) \Rightarrow \frac{2\pi}{\lambda} = \frac{\pi}{4} \Rightarrow \lambda = 8 \text{ cm}$

And, as you know, the distance between two successive nodes is equal to

$$\frac{\lambda}{2} = \frac{8 \text{ cm}}{2} = 4 \text{ cm}$$

EXAMPLE 2.3: FREQUENCIES OF HIGHER HARMONICS

The fundamental frequency of a guitar string is 540 Hz. a) Calculate the frequency of the first and second harmonics excited on it. b) If the tension in the guitar string is doubled, calculate the new fundamental frequency.

SOLUTION ■ a) From Sec. 2.3.1, we recall that when standing waves are formed on a string fixed at both ends, the fundamental frequency is also known as the first harmonic and the higher harmonics are integral multiples of the fundamental frequency. In a guitar, strings are fixed at both ends. So, as per the problem, the frequency of the first harmonic is 540 Hz and the frequency of the second harmonic is $2f_1 = 2 \times 540 \text{ Hz} = 1080 \text{ Hz}$.

b) From Eq. (2.15), we recall that frequencies of standing waves formed on a string fixed at both ends are given by

$$f_n = \frac{n}{2L} \sqrt{\frac{T}{\rho}}, \quad n = 1, 2, 3, \dots \quad (\text{i})$$

Therefore, the fundamental frequency ($n = 1$) is given by

$$f_1 = \frac{1}{2L} \sqrt{\frac{T}{\rho}}$$

When the tension T in the guitar string is increased to $2T$, the new fundamental frequency f'_1 is given [Eq. (2.16)] as

$$f'_1 = \frac{1}{2L} \sqrt{\frac{2T}{\rho}} = \sqrt{2} f_1 = 763.7 \text{ Hz}$$

Before studying further, you should solve the following SAQs.

SAQ 2 – Standing waves on a string and fundamental frequency

A string of length 2 m is clamped at both ends. Suppose that a wave propagates on this string with speed 500 ms^{-1} . (a) Calculate the minimum frequency of the disturbance which must be generated on this string to obtain a standing wave. (b) What will be the next higher frequency of the standing wave?

While studying the characteristics of standing waves, you might have noted that these differ from progressive waves in the following ways:

1. Standing waves do not transport energy like progressive waves.
2. Each element or particle of the medium oscillates with the same amplitude in progressive waves. But, in standing waves, each element or particle of the medium has different amplitude.

You may now like to know: **What gives rise to these differences?** To address this question, we need to examine a little more closely the displacement and velocity of each element of the medium, in which a standing wave is formed.

Displacement and Velocity of a Particle in a Standing Wave

From elementary mechanics, you know that the velocity of a particle is defined as the time rate of change of displacement. So, to obtain an expression for the

velocity of a particle in a standing wave, we differentiate the resultant displacement, $y(x, t)$ given by Eq. (2.9) with respect to time. Doing so, we get

$$v = \frac{dy(x, t)}{dt} = \frac{d}{dt}(2a \sin kx \cos \omega t) = -2a\omega \sin kx \sin \omega t \quad (2.17)$$

Eq. (2.17) shows that the magnitude of the velocity of a particle in the medium will be maximum when $\sin kx = \pm 1$ and it will be zero when $\sin kx = 0$. If you examine the expression for displacement [Eq. (2.9)] of the particle, you will note that the displacement of the particle also attains maximum values for $\sin kx = \pm 1$ and becomes zero for $\sin kx = 0$. Further you know that $\sin kx = \pm 1$ and $\sin kx = 0$, respectively, define **antinodes** and **nodes** on a standing wave. **It means that the displacement as well as velocity of the particle attain maximum value at the antinodes and both are zero at the nodes.** At any intermediate position, the displacement and velocity of the particles gradually change between maximum and zero.

To appreciate the above conclusions, refer to Fig. 2.13, which depicts the velocity of the particles of the string supporting a standing wave at a given instant. The lengths of arrow heads perpendicular to the x-axis represent the magnitude of velocities of the particles at that instant. From the figure, you may note that, in standing waves, particles of the medium (string) get divided into segments like *M*, *N*, and *O* in such a manner that the **particles in any one segment, say *M*, always move along the same direction**. Also, note that the particles in two adjacent segments, such as *M* and *N*, always move in opposite directions. That is, when particles in segment *M* move up, particles in segment *N* move down and *vice-versa*. Now compare the variation of velocity of particles depicted in Fig. 2.13 with the variation of displacement of the particles depicted in Fig. 2.10. You will note that both are identical.

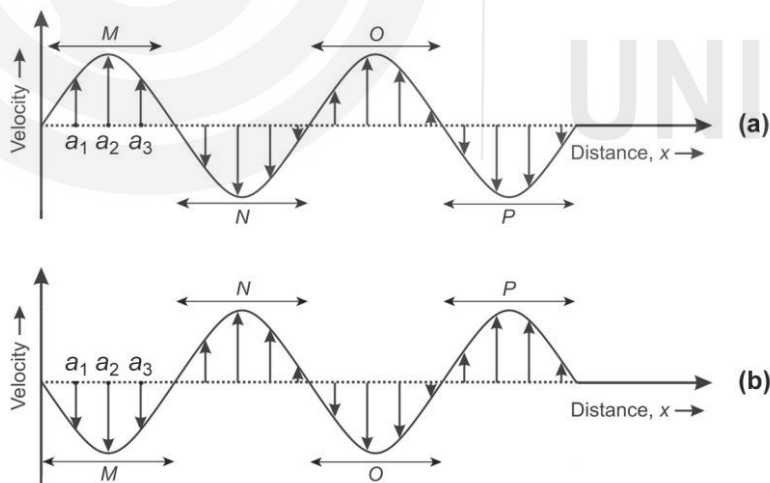


Fig. 2.13: a) Velocity-position graph for the particles in a standing wave on a string fixed at both ends. The lengths of the arrow-heads denote the magnitude of the velocity of the particles at a given instant of time; b) snapshot of the standing wave at $t = t_2$.

Now, study Fig. 2.13 *a* and *b* again. While Fig. 2.13*a* depicts the snapshot of the standing wave at time, say t_1 , Fig. 2.13*b* depicts it at a later time, $t_2 (= t_1 + (T/2))$. Note that the magnitude of the velocity (shown by arrow-heads) of the particles in any one segment, say particles a_1, a_2 and a_3 in

segment M , are different. Particle which has to cover larger distance possess greater velocity and *vice-versa*. As a result, all particles in a segment – crest or trough – of the standing wave reach their respective extreme positions at the same time. **This happens because all particles of the string have the same time period but different velocities.**

Now, to concretise your understanding about the behaviour of the particle of the string when a standing wave is formed on it, you should study the following example.

EXAMPLE 2.4: THE POSITION OF NODES IN A STANDING WAVE

Two waves propagating on a stretched string fixed at both ends are given by

$$y_1(x, t) = (0.016) \sin\left(\frac{x}{2} - 20t\right)$$

$$y_2(x, t) = (0.016) \sin\left(\frac{x}{2} + 20t\right)$$

where y_1 , y_2 and x are in metres and t is in seconds. a) Determine the position of the nodes of the resulting standing wave. b) Calculate the amplitude of a particle of the string located at $x = 0.40$ m.

SOLUTION ■ a) From the given expressions for the waves, we note that the two waves are identical and are travelling in opposite directions. Therefore, they will superpose and give rise to a standing wave. From Eq. (2.9), we recall that the instantaneous displacement of the particles in the resultant standing wave is given by

$$y(x, t) = 2a \sin kx \cos \omega t \quad (i)$$

From the given expressions for $y_1(x, t)$ and $y_2(x, t)$, we note that

$$a = 0.016 \text{ m}; k = \frac{2\pi}{\lambda} = \frac{1}{2} \Rightarrow \lambda = 4\pi \text{ m}; \omega = 2\pi f = 20 \text{ s}^{-1} \Rightarrow f = \left(\frac{10}{\pi}\right) \text{ s}^{-1} \quad (ii)$$

On substituting these values of a , k and ω in Eq. (i), we get

$$y(x, t) = (0.032 \text{ m}) \sin(0.5x) \cos(20t)$$

From Eq. (2.10), we recall that the position of nodes is given by

$$x = \frac{n\lambda}{2}; \quad n = 0, 1, 2, \dots$$

On substituting the value of λ from Eq. (ii), we get

$$x = \frac{n4\pi}{2}; \quad n = 0, 1, 2, \dots$$

So, the nodes will be located at

$$x = 0, 6.28 \text{ m}, 12.56 \text{ m}, \dots$$

b) We know that amplitude of a standing wave is $2a \sin kx$. So, the amplitude of the particle located at $x = 0.40$ m will be

$$(a)_{x=0.40 \text{ m}} = 2 \times (0.016 \text{ m}) \sin\left(\frac{0.40}{2}\right) = 0.011 \text{ m}$$

You may now like to answer an SAQ.

SAQ 3 – Determination of the position of nodes and antinodes

The superposition of two sinusoidal waves travelling in opposite directions on a string fixed at both ends gives rise to a standing wave represented as:

$$y(x,t) = (0.12 \text{ m}) \sin(4.0x) \cos(20t)$$

where x is in metres and t is in seconds. Determine the a) wavelength, and frequency of the standing wave, and b) position of nodes and antinodes.

Now, before proceeding further, let us recapitulate what you have learnt in this section.

Recap

- A **standing wave** is formed when two identical progressive waves travelling in opposite directions superpose. If standing waves are formed on a string with both ends fixed, the string may be divided into one or more segments. Each segment has a point of zero displacement, called **node**, and a point of maximum displacement, called **antinode**.
- The **normal mode** frequencies of the standing wave on a string of length L are given by

$$f_n = \frac{nv}{2L} = \frac{n}{2L} \sqrt{\frac{T}{\rho}}; \quad n = 1, 2, 3, \dots$$

- The lowest frequency, f_1 (for $n = 1$) is called the **fundamental frequency** or the first harmonic of the standing wave. The higher frequencies $f_2, f_3, \dots$ etc. are, respectively, called second, third and higher harmonics.
- The particles in any given segment of a standing wave always move along the same direction. However, the particles in two adjacent segments move in opposite directions.

So far, we have considered superposition of two identical waves travelling in opposite directions. As pointed out in Sec. 2.2, when two or more waves of slightly different frequencies travelling in the same direction superpose we observe the phenomena of beats, formation of wave groups, etc. Let us learn about these now.

2.4 BEATS

When sound waves of different frequencies superpose, the intensity of the resultant sound increases and decreases alternately with time. This phenomenon is called **beats**. Though the name beats comes for the superposition of sound waves, the phenomenon is true for all types of waves if the superposing waves have slightly different frequencies.

To understand the phenomenon of beats qualitatively, recall from Sec. 2.2 that the resultant of the superposition of two harmonic waves of same frequency travelling in the same direction depends on the phase difference between the two waves. If the two waves are in phase, crest for crest, the amplitude of the resultant wave will be maximum. When they are out of phase, crest for trough, the amplitude of the resultant wave will be minimum. **So, if**

the superposing waves are sound waves, the intensity of the resultant wave will be maximum when the superposing waves are in phase and *vice-versa*. This is because the intensity or loudness of the sound is proportional to the square of its amplitude.

Note that, at any given point in space, if the two superposing waves of the same frequency are in-phase or out-of phase, they will remain so at all subsequent times. This means that if we fix our position in space, we can either observe / hear enhanced intensity (due to in-phase superposition) or diminished intensity (due to out-of phase superposition). **But, what we observe / hear in case of beats is quite contrary: even if we are located at a fixed position in space, we perceive periodic increase and decrease in intensity of sound.** This means that the superposing waves are in-phase at one instant and are out-of phase at the other instant at the same position in space! The question is: How does it happen?

The variation in the phase difference between the superposing waves giving rise to beats is caused due to the different values of their frequencies. To visualise this, refer to Figs. 2.14. Fig. 2.14a and 2.14b depict two sinusoidal waves of the same amplitude but slightly different frequencies ω_1 and ω_2 , respectively, where $\omega_1 > \omega_2$. Note that these two waves are being depicted on vibration graphs; that is, these are displacement-time graphs for the two waves at a fixed value of position, x .

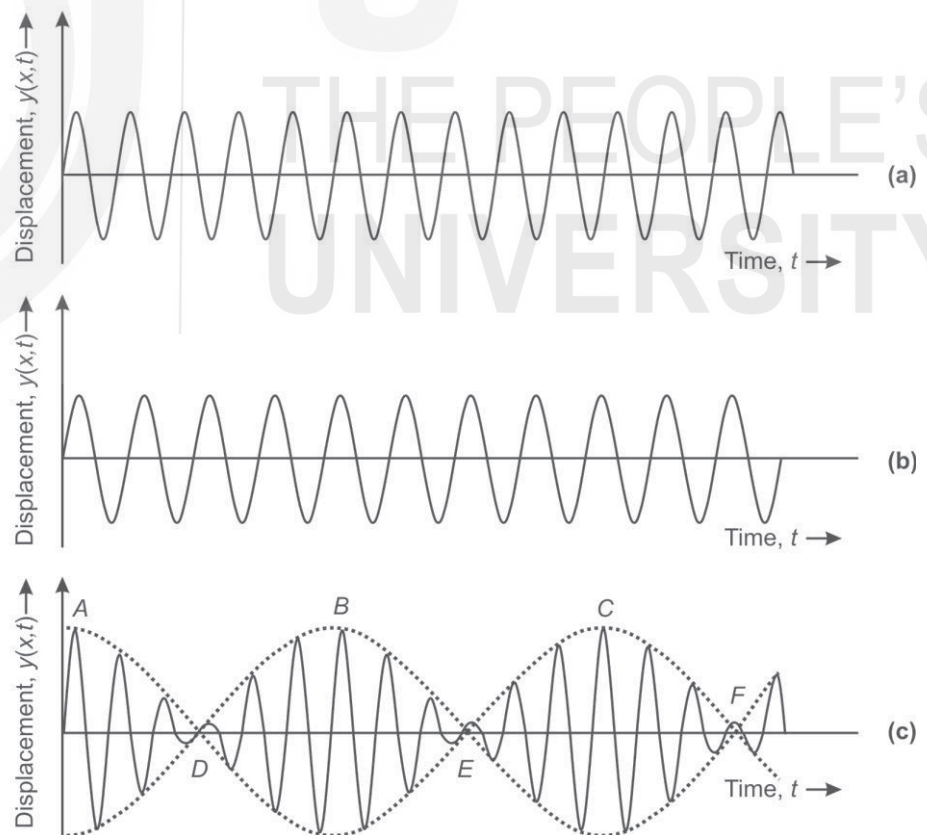


Fig. 2.14: Superposition of two waves of the same amplitude but slightly different frequencies: a) vibration graphs of wave of frequency ω_1 ; b) vibration graph of wave of frequency ω_2 ; c) vibration graph of the resultant wave due to superposition of waves depicted in a) and b).

Fig. 2.14c depicts the vibration graph for the resultant wave due to superposition of the waves depicted in Figs. 2.14a and 2.14b. You may note the following from Fig. 2.14a to c:

- In the regions where the two superposing waves are in-phase, such as the regions A, B and C, the amplitude of the resultant wave is large.
- In the regions where the superposing waves are out of phase, such as the regions D, E and F, the amplitude of the resultant wave is very small.
- **Also, the amplitude of the resultant wave varies periodically. This gives rise to alternations of intensity called beats.**

Now, for obtaining a relation between the beat frequency and the frequencies of the superposing wave, let us consider two waves of same amplitude but slightly different angular frequencies ω_1 and ω_2 given as:

$$y_1(x, t) = a \sin(\omega_1 t - k_1 x)$$

$$y_2(x, t) = a \sin(\omega_2 t - k_2 x)$$

Since we are interested in a physical situation where the position of the observer/listener is fixed, we can take $x = 0$ in the above expressions for waves and consider only the time dependence of the waves (the vibration graph). So, the expression for the superposing waves reduces to:
 $y_1(t) = a \cos \omega_1 t$ and $y_2(t) = a \sin \omega_2 t$. When these two waves superpose, the resultant wave is given by the superposition principle as:

$$\begin{aligned} y(t) &= a \sin \omega_1 t + a \sin \omega_2 t \\ &= 2a \sin \left[\frac{\omega_1 + \omega_2}{2} t \right] \cos \left[\frac{\omega_1 - \omega_2}{2} t \right] \\ &= 2a \cos \frac{\Delta \omega}{2} t \sin \omega_{av} t \end{aligned} \quad (2.18)$$

where $\omega_{av} = (\omega_1 + \omega_2)/2$ defines the **average angular frequency** of the resultant wave and $\Delta \omega = \omega_1 - \omega_2$ represents the difference in the angular frequencies of the superposing waves.

Eq. (2.18) is the expression for the resultant wave which contains a sinusoidal term, $\sin \omega_{av} t$ which is modulated by an amplitude factor $2a \cos(\Delta \omega/2)t$.

Eq. (2.18) suggests that

- at a given point in space, the amplitude of the resultant wave, $2a \cos(\Delta \omega/2)t$, varies with time between $2a$ and zero;
- the resultant waveform follows sinusoidal variation **within an envelope** represented by the amplitude term, as shown in Fig. 2.14c; and
- the frequency of the resultant wave is ω_{av} ; if ω_1 and ω_2 are nearly equal, the amplitude of the resultant wave will vary very slowly.

So, the time interval, Δt between two successive beats will correspond to the cosine term in Eq. (2.18). Since intensity of a wave is proportional to the square of amplitude, the time interval between two successive beats will be the time in which amplitude squared, that is, $(\cosine)^2$ factor ranges from $+1$ to zero to $+1$.

Thus, we can write

$$\frac{\Delta\omega}{2} \Delta t = \pi$$

$$\text{or} \quad \Delta t = \frac{2\pi}{\Delta\omega} \quad (2.19)$$

This gives the **time period for beats**. Since frequency is reciprocal of the time period, the **frequency of beats** can be written as

$$\begin{aligned} f_b &= \frac{1}{\Delta t} \\ &= \frac{\Delta\omega}{2\pi} = |f_1 - f_2| \end{aligned}$$

Thus the frequency of beats is equal to the difference of the frequencies of two superposing waves.

You may now like to answer an SAQ.

SAQ 4 – Calculating beat frequency

A note of a piano is sounded with a tuning fork of frequency 512 Hz. If six beats are heard every second, calculate the frequency of the note.

You now know that beats arise due to superposition of waves of slightly different frequencies. **On the basis of the above discussion, we can say that beats are the manifestation of the collective behaviour of two or more waves of different frequencies propagating simultaneously.** You may recall that while analysing the phenomenon of beats, we assumed that $x = 0$. This condition basically means that the value of the position or location of observation or listening has been fixed. The situation corresponds to superposition of vibration graphs (displacement-time graph representing a wave at a fixed position) generated by waves of different frequencies. For this reason, beats are also known as ‘superposition in time’. You may ask: **What will happen if we relax the condition that the value of position, x is fixed?** In that situation, the analysis of the superposition leads us to the concept of wave groups or wave packets. You will learn it now.

2.5 WAVE GROUPS AND GROUP VELOCITY

Consider two waves of equal amplitude but slightly different frequencies ω_1 and ω_2 propagating in the positive x -direction. We can express such waves mathematically as

$$y_1(x, t) = a \sin(\omega_1 t - k_1 x)$$

$$\text{and} \quad y_2(x, t) = a \sin(\omega_2 t - k_2 x)$$

The resultant wave due to the superposition of these two waves is given by

$$\begin{aligned} y(x, t) &= y_1(x, t) + y_2(x, t) \\ &= a \sin(\omega_1 t - k_1 x) + a \sin(\omega_2 t - k_2 x) \end{aligned} \quad (2.20)$$

To proceed further, we use the trigonometric identity

$$\sin A + \sin B = 2 \sin\left(\frac{A+B}{2}\right) \cos\left(\frac{A-B}{2}\right)$$

and take $A = \omega_1 t - k_1 x$ and $B = \omega_2 t - k_2 x$. Then, we can write

$$\begin{aligned}
 y(x, t) &= 2a \sin \left[\frac{\omega_1 t - k_1 x + \omega_2 t - k_2 x}{2} \right] \cos \left[\frac{\omega_1 t - k_1 x - \omega_2 t + k_2 x}{2} \right] \\
 &= 2a \sin \left[\frac{(\omega_1 + \omega_2)t}{2} - \frac{(k_1 + k_2)x}{2} \right] \cos \left[\frac{(\omega_1 - \omega_2)t}{2} - \frac{(k_1 - k_2)x}{2} \right] \\
 &= 2a \cos \left[\left(\frac{\Delta\omega}{2} \right) t - \left(\frac{\Delta k}{2} \right) x \right] \sin(\omega_{av} t - k_{av} x) \quad (2.21)
 \end{aligned}$$

where $\omega_1 - \omega_2 = \Delta\omega$, $k_1 - k_2 = \Delta k$, and we have defined **average angular**

frequency and **average wave number** as $\omega_{av} = \frac{\omega_1 + \omega_2}{2}$ and

$k_{av} = \frac{k_1 + k_2}{2}$, respectively.

We can rewrite Eq. (2.21) as

$$y(x, t) = a_m \sin(\omega_{av} t - k_{av} x) \quad (2.22)$$

where

$$a_m = 2a \cos \left(\frac{\Delta\omega}{2} t - \frac{\Delta k}{2} x \right) \quad (2.23)$$

Let us pause for a moment and try to understand what different terms in Eq. (2.22) signify physically. Refer to Fig. 2.15 which depicts the waveform graphs corresponding to Eq. (2.21). Fig. 2.15a and b depict the sine and cosine terms of Eq. (2.21) respectively. And, Fig. 2.15c depicts the resultant of the graphs in Fig. 2.15a and b.

Firstly, the function is not a harmonic function because its amplitude, a_m varies both in space (x) and time (t). **The varying or modulated amplitude gives rise to a wave envelop** as shown in Fig. 2.15b.

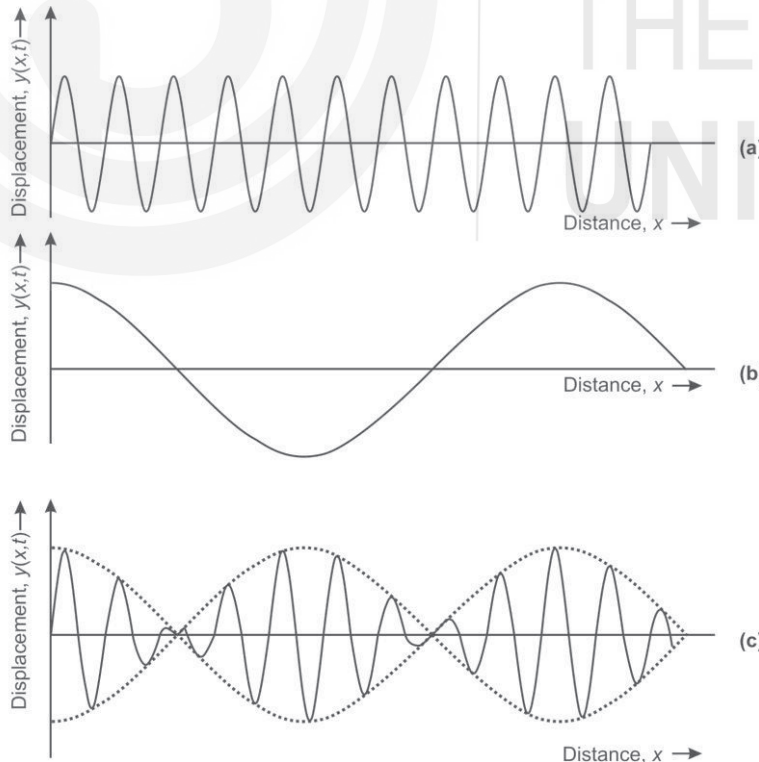


Fig. 2.15: The waveform arising due to superposition of two waves of slightly different frequencies moving along the same direction; a) waveform corresponding to the sine term in Eq. (2.21); b) waveform corresponding to the cosine term in Eq. (2.21); c) waveform obtained by plotting the graphs in a) and b) together.

Secondly, we note that the function representing the resultant wave comprises two parts:

- i) the sine part which varies (rapidly) with a frequency that is average of the frequencies of the component waves (this part is depicted graphically in Fig. 2.15a); and
- ii) the cosine part which varies (very slowly) with a frequency that is half of the difference of the frequencies of the superposing waves (this part is depicted graphically in Fig. 2.15b). The product of the above two parts of Eq. (2.21) is plotted in Fig. 2.15c.

We further note from Fig. 2.15c that the superposition of waves of slightly different frequencies moving along the same direction results in the formation of groups (or packets) of waves. These are called **wave groups** or **wave packets**.

You may ask: **How do these wave groups behave? What is their velocity of propagation?** In general, a **wave group propagates with a velocity different from that of the individual superposing waves**. The velocity of the wave group is called **group velocity**. We denote it by the symbol v_g . Let us now obtain an expression for v_g and check the validity of the statement regarding the velocity of wave groups.

Since wave groups are formed due to the modulated part (represented by the cosine term) of Eq. (2.21), we can take the angular frequency, $\omega = (\Delta\omega/2)$ and wave number, $k = (\Delta k/2)$ for the wave group. So, using the expression for wave velocity, $v = (\omega/k)$, we can write the group velocity, v_g for the wave group as

$$v_g = \frac{\Delta\omega/2}{\Delta k/2} = \frac{\Delta\omega}{\Delta k} \quad (2.24)$$

If a wave group is formed due to superposition of not only two but a large number of waves with angular frequencies lying between ω_1 and ω_2 (with $\omega_1 \approx \omega_2$), and wave numbers lying between k_1 and k_2 , the expression for group velocity can be written as

$$v_g = \frac{d\omega}{dk} \quad (2.25)$$

Here $d\omega$ and dk , respectively, denote the spreads (or difference between the maximum and the minimum) of angular frequencies and wave numbers of the superposing waves that gives rise to the wave group.

You may ask: **What is the velocity of the resultant wave represented by the sine function** in Eq. (2.21)? The velocity of the resultant wave is called the **phase velocity**. We denote it by the symbol v_p . From Eq. (2.21), we can write the expression for phase velocity, v_p as

$$v_p = \frac{\omega_{av}}{k_{av}} \quad (2.26)$$

Now let us consider a situation in which wave velocity of the superposing waves is the same:

$$\frac{\omega_1}{k_1} = \frac{\omega_2}{k_2} = v(\text{say})$$

Then, in that condition, we can write the phase velocity as

$$v_p = \frac{\omega_1 + \omega_2}{k_1 + k_2} = \frac{k_1 v + k_2 v}{k_1 + k_2} = v \quad (2.27)$$

And, for this particular case, we can write group velocity [Eq. (2.24)] as

$$v_g = \frac{\Delta\omega}{\Delta k} = \frac{\omega_1 - \omega_2}{k_1 - k_2} = \frac{k_1 v - k_2 v}{k_1 - k_2} = v \quad (2.28)$$

From Eqs. (2.27) and (2.28) we note that, **if the individual wave velocities are equal, the group velocity is equal to the phase velocity.**

We have mentioned earlier that the resultant wave travels with phase velocity. **However, the energy transfer by such waves is governed by group velocity.** That is why group velocity is a more fundamental physical quantity and of greater interest in physics.

Now, you may like to ask: **How are phase velocity and group velocity related?** We can establish a relation between the *phase velocity* and the *group velocity* if we replace $d\omega$ by $d(kv_p)$ in the expression for v_g [Eq. (2.25)]:

$$v_g = \frac{d\omega}{dk} = \frac{d(kv_p)}{dk}$$

On simplification, we obtain

$$v_g = v_p + k \frac{dv_p}{dk} \quad (2.29)$$

Since $k = 2\pi/\lambda$, we can write Eq. (2.29) in terms of λ as (read the margin remark):

$$v_g = v_p - \lambda \frac{dv_p}{d\lambda} \quad (2.30)$$

Yet another important aspect of wave groups is its spatial extent. Refer to Fig. 2.16 which depicts wave groups formed due to superposition of waves of different frequencies. You may note that the **length of a wave group is half of the wavelength of the enveloping wave.** To substantiate this statement, we obtain expressions for the wavelength of the resultant wave as well as the enveloping (or the modulating) wave. The wavelength λ_r of the resultant wave can be written using the sine function of Eq. (2.21):

$$\lambda_r = \frac{2\pi}{k_{av}} \approx \frac{2\pi}{k} \quad (\text{because } k_1 \approx k_2)$$

And the wavelength λ_e of the enveloping wave can be written using the cosine function of Eq. (2.21):

$$\lambda_e = \frac{2\pi}{\Delta k/2} = \frac{4\pi}{\Delta k}$$

Since Δk is very small compared to k we have: $\lambda_e \gg \lambda_r$.

If λ_1 and λ_2 denote the wavelengths of the component waves, (using $\Delta k = k_1 - k_2$) you can verify that

$$\frac{\lambda_e}{2} = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} \quad (2.31)$$

To express v_g in terms of λ , we note that

$$\frac{dv}{dk} = \frac{dv}{d\lambda} \cdot \frac{d\lambda}{dk}$$

Since $k = \frac{2\pi}{\lambda}$, it readily

follows that

$$\frac{dk}{d\lambda} = -\frac{2\pi}{\lambda^2}$$

Inserting this result in above equation, we get

$$\frac{dv}{dk} = -\frac{\lambda^2}{2\pi} \frac{dv}{d\lambda}$$

Hence

$$\begin{aligned} v_g &= v_p + \frac{2\pi}{\lambda} \left(-\frac{\lambda^2}{2\pi} \right) \frac{dv}{d\lambda} \\ &= v_p - \lambda \frac{dv_p}{d\lambda} \end{aligned}$$

Eq. (2.30) gives the general relation between group velocity and phase velocity. If you pursue higher studies in physics, you will learn that the phase velocity is independent of wavelength if the wave is propagating in a non-dispersive medium. Thus, in a non-dispersive medium, group velocity is equal to phase velocity.

Since the spatial extent of the wave group is half of the wavelength of the enveloping wave, Eq. (2.31) gives the length (or the extent) of the wave group. You can see from Fig. 2.16 that the length of the wave group is half of the wavelength of the enveloping wave, i.e., it is equal to $\lambda_e/2$.

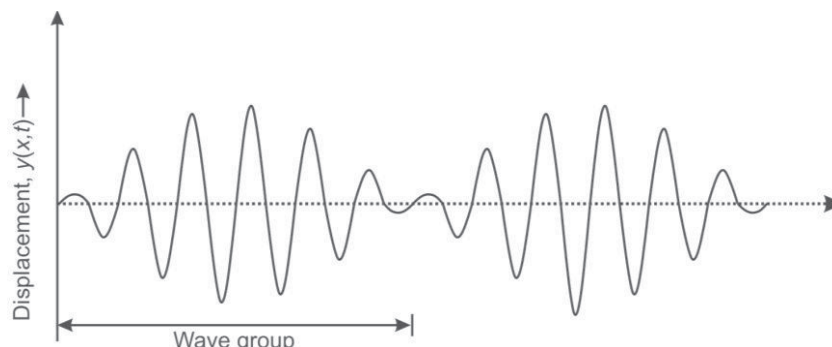


Fig. 2.16: Wave group and its extent.

You may now like to solve a SAQ.

SAQ 5 – Phase and group velocities

The phase velocity of a wave in a medium is given by

$$v_p = C_1 + C_2\lambda$$

where C_1 and C_2 are constants. Obtain an expression for its group velocity.

Let us now summarise what you have learnt in this unit.

2.6 SUMMARY

Concept	Description
Superposition principle	When two or more waves superpose, the displacement of the particles of the medium in the resultant wave is given by the algebraic sum of the individual displacements due to superposing waves. This is known as principle of superposition of waves.
Standing waves	Standing waves arise because of the superposition of two identical waves travelling in opposite directions and confined between two points. - The expression for the standing waves due to superposition of two waves $y_1(x,t) = a \sin(kx - \omega t)$ and $y_2(x,t) = a \sin(kx + \omega t)$ is given as $y(x,t) = y_1(x,t) + y_2(x,t) = 2a \sin kx \cos \omega t$
Nodes and antinodes	In a standing wave, the region of space is divided into well defined segments. The points of zero and maximum displacement are referred to as nodes and antinodes, respectively. The distance between any two consecutive nodes or antinodes is equal to one-half of the wavelength of the standing wave.
Normal modes of a string	The normal mode frequencies of the standing wave on a string of length L and fixed at both the end is given as

$$f_n = \frac{nv}{2L} = \frac{n}{2L} \sqrt{\frac{T}{\rho}} \quad n = 1, 2, 3, \dots$$

The value of the frequency corresponding to $n = 1$ is called **fundamental frequency**. The fundamental frequency is the minimum frequency with which the string can vibrate. The values of normal mode frequencies for $n = 2$ and above are called **higher harmonics**.

Beats

- When two waves of slightly different frequencies travelling along the same direction superpose, we obtain periodic variations in the amplitude of the resultant wave at a fixed point in space. These are referred to as **beats**. In case of sound waves, the beats are heard in the form of waxing and waning of sound. The number of beats produced per second; that is, the **beat** frequency, is given as

$$f_b = |f_1 - f_2|$$

The expression for the resultant wave giving rise to beats is given as

$$y(t) = 2a \cos \frac{\Delta\omega}{2} t \sin \omega_{av} t$$

Note that beats are also known as superposition in time; that is, the position of observation/listening is fixed.

Wave groups

- The superposition of many waves of slightly different frequencies travelling in the same direction results in the production of a **wave packet** or **wave group**.

Group velocity and phase velocity

- The velocity with which a wave group or wave packet travels is called the **group velocity**. The group velocity is given as

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

The velocity of the resultant wave is called **phase velocity** and it is given as

$$v_p = \frac{\omega_{av}}{k_{av}}$$

The group velocity and phase velocity are related as

$$v_g = v_p + k \frac{dv_p}{dk}$$

2.7 TERMINAL QUESTIONS

1. The length of a piano string fixed at both ends is 1.0 m and its mass per unit length is 0.015 kg m^{-1} . If the frequency of the fundamental note is 220 Hz, calculate the tension in the string.
2. A string of length 3.5 m and mass 0.70 kg is fixed at both ends and the tension in the string is 60 N. a) Calculate the fundamental frequency of the standing waves on the string. b) If the standing wave generated on the string has two antinodes, what is its frequency?
3. Standing waves having a frequency of 8.0 Hz is established on a rope of length 3.5 m fixed at both ends. The wavelength is twice the length of the string. Calculate the speed of the wave.
4. Sixty five tuning forks are arranged in order of increasing frequency and any two successive forks produce 4 beats per second when sounded together. If the frequency of the last tuning fork is twice the frequency of the first tuning fork, calculate the frequency of the first tuning fork.
5. Wave groups are formed due to the superposition of two waves of equal amplitude but slightly different frequencies propagating along the positive x-direction. If the wavelengths λ_1 and λ_2 of the two superposing waves are 0.50 m and 0.75 m, respectively, calculate the spatial extent of the wave group.

2.8 SOLUTIONS AND ANSWERS

Self-Assessment Questions

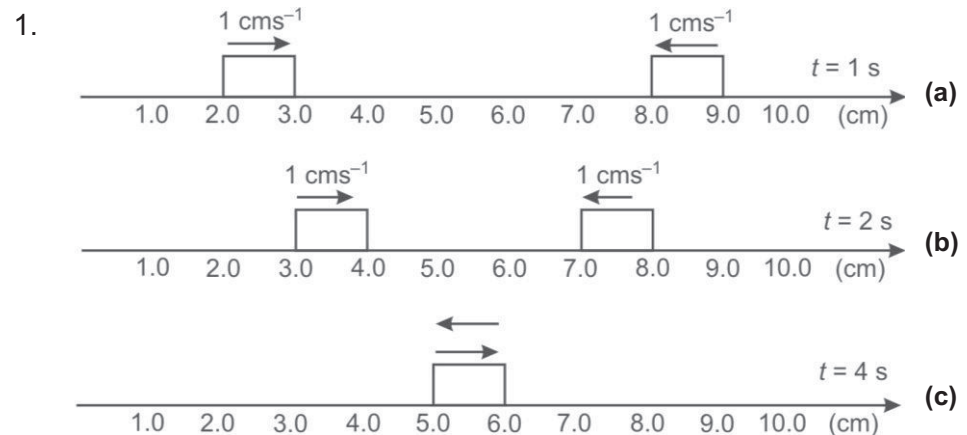


Fig. 2.17

2. a) We know that the frequency of the standing wave on a string clamped at both ends is given by

$$f_n = \frac{n}{2L} v \quad n = 1, 2, 3, \dots$$

The minimum frequency of the standing wave is obtained by taking $n = 1$ in the above relation. So,

$$f_1 = \frac{1}{2 \times (2\text{ m})} \times 500 \text{ ms}^{-1} = 125 \text{ s}^{-1}$$

- b) Next frequency of the standing wave which can be sustained on the string is

$$f_2 = 2f_1 = 2 \times 125 \text{ s}^{-1} = 250 \text{ s}^{-1}$$

3. a) The given expression for the standing wave is:

$$y(x, t) = (1.2 \text{ m}) \sin(0.40 x) \cos(200 t)$$

On comparing this expression with the standard expression for standing wave on a string fixed at both ends:

$$y(x, t) = 2a \sin kx \cos \omega t$$

we get:

$$k = 4.0 \text{ m}^{-1} \Rightarrow \lambda = \frac{2\pi}{4.0 \text{ m}^{-1}} = 1.57 \text{ m}$$

$$\omega = 20 \text{ s}^{-1} \Rightarrow f = \frac{20 \text{ s}^{-1}}{2\pi} = 3.2 \text{ s}^{-1}$$

- b) The position of the nodes is given by

$$x = \frac{n\lambda}{2}$$

$$\text{for } n = 1; \quad x_1 = \frac{1.57 \text{ m}}{2} = 0.79 \text{ m}$$

$$\text{for } n = 2; \quad x_2 = \frac{2 \times (1.57 \text{ m})}{2} = 1.57 \text{ m} \quad \text{etc.}$$

The position of antinodes is given by

$$x = \frac{(2n+1)\lambda}{4}$$

$$\text{for } n = 1; \quad x_1 = \frac{3 \times (1.57 \text{ m})}{4} = 1.18 \text{ m}$$

$$\text{for } n = 2; \quad x_2 = \frac{5 \times (1.57 \text{ m})}{4} = 1.96 \text{ m} \quad \text{etc.}$$

4. Let the frequency of the note be f_2 . We know that the beat frequency, f_b is given by

$$f_b = |f_1 - f_2|$$

$$\text{or} \quad 6 = |512 - f_2|$$

$$\therefore f_2 = 512 \pm 6$$

In this case, one frequency of the note cannot be identified; it will be either 518 Hz or 506 Hz.

5. We know that the group velocity and phase velocity are related to each other by the relation [Eq. (2.30)]:

$$v_g = v_p - \lambda \frac{dv_p}{d\lambda}$$

From the given expression for the phase velocity of the wave, we have:

$$\frac{dv_p}{d\lambda} = C_2.$$

Inserting this result in the expression for v_g , we obtain

$$v_g = C_1 + C_2\lambda - \lambda C_2 = C_1$$

Terminal Questions

1. Frequency of the fundamental note of a piano string fixed at both ends can be obtained by using Eq. (2.15):

$$f_n = \frac{n}{2L} \sqrt{\frac{T}{\rho}}$$

On taking $n = 1$ (fundamental note) and substituting the value of other parameters in the above expression, we get:

$$220 \text{ s}^{-1} = \frac{1}{2 \times 1 \text{ m}} \sqrt{\frac{T}{0.015 \text{ kg m}^{-1}}}$$

$$\text{or} \quad T = (440 \text{ ms}^{-1})^2 \times (0.015 \text{ kg m}^{-1}) = 2904 \text{ N}$$

2. a) The frequency of normal modes of the standing waves on a string fixed at both ends is given as

$$f_n = \frac{nV}{2L} = \frac{n}{2L} \sqrt{\frac{T}{\rho}} \quad n = 1, 2, 3, \dots$$

For the fundamental frequency, $n = 1$. So, the fundamental frequency, f_1 is

$$f_1 = \frac{1}{2L} \sqrt{\frac{T}{\rho}}$$

The mass per unit length, $\rho = \frac{m}{L} = 0.25 \text{ kg m}^{-1}$

And, as per the problem, tension, $T = 60 \text{ N}$.

$$f_1 = \frac{1}{2 \times (3.5 \text{ m})} \times \sqrt{\frac{60 \text{ N}}{0.25 \text{ kg m}^{-1}}} = 2.21 \text{ s}^{-1}$$

- b) When the standing wave on a string fixed at both ends has two antinodes, it is called second harmonic (Fig. 2.12c). And, frequency, f_2 of the second harmonic is

$$f_2 = 2f_1 = 2 \times 2.21 \text{ s}^{-1} = 4.42 \text{ s}^{-1}$$

3. The frequency of the normal modes of a string fixed at both ends is given as

$$f_n = \frac{nv}{2L}$$

We know that when the length of the string is twice the wavelength, the standing wave corresponds to the fundamental mode ($n = 1$). So, the given frequency is the fundamental frequency. So,

$$f_1 = \frac{v}{2L} \Rightarrow v = f_1 \times (2L) = (8 \text{ s}^{-1}) \times (2 \times 3.5 \text{ m}) = 56 \text{ ms}^{-1}.$$

4. Let the frequency of the first turning fork be f_1 . Then the frequency of the second fork will be $f_2 = f_1 + 4$. We can rewrite it as

$$f_2 = f_1 + (2 - 1)4$$

Similarly, the frequency of the third fork

$$f_3 = f_2 + 4 = f_1 + (3 - 1)4$$

and the frequency of the 65th fork

$$f = f_1 + (65 - 1) \times 4 = f_1 + 256$$

Since the frequency of 65th fork is $2f_1$, we can write

$$f_1 + 256 = 2f_1 \Rightarrow f_1 = 256 \text{ Hz}$$

5. From Fig. 2.15c, you may note that the spatial extent (that is, the length of a wave group) is half the wavelength λ_e of the enveloping wave. And, from Eq. (2.31), we have

$$\frac{\lambda_e}{2} = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1}$$

So, substituting the values of λ_1 and λ_2 , we get

$$\frac{\lambda_e}{2} = \frac{(0.50 \text{ m}) \times (0.75 \text{ m})}{(0.75 \text{ m}) - (0.50 \text{ m})} = 1.5 \text{ m}$$



UNIT 3

ACOUSTIC WAVE |

A concert hall has to be specially designed to ensure that sound energy is uniformly distributed through the entire hall. This area of study is called **acoustics of buildings**.

(Source of picture: wikimedia.org)

Structure

- | | |
|--|---|
| <p>3.1 Introduction
Expected Learning Outcomes</p> <p>3.2 Production of Sound Waves
Graphical and Mathematical
Representation of Sound Waves
Sound Waves in Different Media
Forced Vibrations and Resonance
Musical Instruments</p> <p>3.3 Intensity of Sound
Intensity Level: Decibel</p> | <p>3.4 Musical Notes and Scale
Musical Scales</p> <p>3.5 Acoustics of Buildings</p> <p>3.6 Summary</p> <p>3.7 Terminal Questions</p> <p>3.8 Solutions and Answers</p> |
|--|---|

STUDY GUIDE

You will learn about acoustic waves in this unit. The word ‘acoustics’ means “the science of sound” and acoustic waves refers to sound waves. Since sound is a wave phenomenon we shall use the concepts related to wave motion (Unit 1) in this unit. You should, therefore, revise Unit 1. The focus of our discussion in this unit is on production, propagation, and detection of sound waves. The phenomena of forced oscillation and resonance play a crucial role in production of sound. You will be able to understand it better if you revise the basics of oscillatory motion from Units 16 and 18 of the first semester course entitled Mechanics (BPHCT-131).

Yet another phenomenon which plays a crucial role in production of sound is the formation of standing waves in strings and air columns. The basic physics involved in formation of standing waves in air columns such as organ pipes is the same as for the formation of standing waves in strings discussed in Unit 2. You should, therefore, read that unit once again before studying this unit. Also, you should keep comparing the salient features of transverse waves on a string discussed in Units 1 and 2 with the salient features of the sound waves discussed in this unit. This will help you appreciate the differences and similarities between the two types of waves.

“If I were not a physicist, I would probably be a musician. I often think in music. I live my daydreams in music. I see my life in terms of music...”

Albert Einstein

3.1 INTRODUCTION

In the previous two units, you have learnt the general characteristics of wave motion and the superposition of waves by considering transverse waves on a string. You now know that waves are characterised by parameters like amplitude, wavelength, frequency and phase. You also know that the study of waves is important because these transport energy from one place to another. In the present unit, you will study about acoustic waves, which is another name for sound waves.

Acoustics means the science related to production, propagation and detection of sound. In fact, the term acoustics is not limited to audio; it includes the design of buildings, musical instruments and much more. As you know, sound is an important part of our every day experience. Just as light is responsible for our visual perception, sound is responsible for our perception of hearing.

Sound is a mechanical wave produced by vibrating objects and it requires a material medium for propagation. For example, when we speak, our vocal cords vibrate and produce sound waves which propagate through air. Human beings, however, cannot 'hear' mechanical waves of all frequencies. The human ear is capable of detecting/hearing mechanical waves having frequency in the range of 20 Hz to 20,000 Hz. **Therefore, mechanical waves in this frequency range are called sound waves.**

Moreover, waves of higher frequencies, called ultrasound waves are used in medical diagnostics and industry. Also, some creatures like bats and dolphins use these high frequency waves for navigating their path in the dark. Our understanding about sound and related phenomena is based on the physics of waves, which you studied in the previous two units. You will learn how to apply the concepts related to waves for the study of sound.

We begin this unit with a discussion, in Sec. 3.2, on the basic process involved in the production and propagation of sound waves. You will discover that sound is a longitudinal wave comprising compressions and rarefactions of the particles of the medium. We have also discussed the graphical and mathematical description of sound waves taking analogy with the transverse waves on a string discussed in Unit 1. For being audible at a given point, the intensity of sound waves must have a value above the **threshold value** ($\sim 10^{-12} \text{ W m}^{-2}$) at that point. Normally, the intensity of sound waves needs to be enhanced. In many musical instrument as well as sound producing organs of living beings, this enhancement is obtained by forced oscillations and resonance. You will learn about forced oscillations and resonance and its role in sound in this section. You will also learn the role of formation of standing waves on strings and organ pipes in production of sound and enhancement of its intensity.

The range of intensity of sound which human beings can hear is very large. Thus, a logarithmic scale, called **decibel**, is used for the intensity of sound about which you will study in Sec. 3.3. In Sec. 3.4, we have discussed why human beings find some sounds pleasing and what are the physical principles underlying musical notes and musical scales. You will also learn about the factors such as echo, reverberation, etc. which impact the hearing experience in a closed room or hall. In Sec. 3.5, we have discussed these factors under **acoustics of buildings** which play important role in design of buildings, particularly, auditoriums and concert halls.

In the next unit, we have discussed electromagnetic waves. As you know, light is an electromagnetic wave. So, the study of electromagnetic waves will help you understand the phenomena such as polarisation, interference and diffraction of light.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ describe how sound waves are produced by vibrating objects;
- ❖ understand the graphical and mathematical representation of sound waves;
- ❖ obtain expressions for speed of sound in different media;
- ❖ explain the role of forced vibrations and resonance in production of sound;
- ❖ explain the role of standing waves in production of sound by musical instruments;
- ❖ explain the decibel scale for intensity of sound and obtain an expression for intensity level of sound on this scale;
- ❖ discuss diatonic and tempered musical scales; and
- ❖ list and explain the factors which affect the acoustics of building.

3.2 PRODUCTION OF SOUND WAVES

Sound is an important sensory experience in our daily life. As light is the stimulus for our visual experiences, sound is the stimulus for hearing. Have you ever thought: How is sound produced by human beings or birds and animals or musical instruments like *sitar* and drum? **Sound is produced by vibrating objects.** For example, when we speak, our vocal cords vibrate and produce sound. In case of string musical instruments like *sitar* or guitar, sound is generated by vibrating strings. Likewise, you will always find a vibrating object as source of sound in any other situation. However, for producing the sensation of hearing, the sound produced by a vibrating object needs a material medium such as air to travel from the vibrating object to our ear. Most of the sound may also travel in liquid and solids. **The presence of material media is a necessary requirement for propagation of sound.**

Sound is a wave phenomenon. It comprises waves generated by a vibrating object and propagates through a medium. Generally, the sound we hear in our day to day life travels through air in the form of longitudinal waves.

To understand how sound waves are produced and how they propagate, let us consider a long tube containing air with a piston fitted at its one end and the other end open as shown in Fig. 3.1. When the piston is at rest, the air inside the tube has uniform density and pressure. Suppose the piston is made to move back and forth in a periodic manner. Now, think: **What is the impact of this periodic motion of the piston on the air in the tube?** As the piston moves to the right, it will push the air particles in its vicinity to the right. This will create a layer of air in which particles are compressed. This compressed layer of the air will, in turn, push the adjacent layer on its right. As a result, a moving region of **compression** is formed in the tube. The compression will

keep moving towards right even when the piston begins moving back to the left. In the compression region, the density and pressure of air is higher than the equilibrium density and pressure (when the piston was at rest).

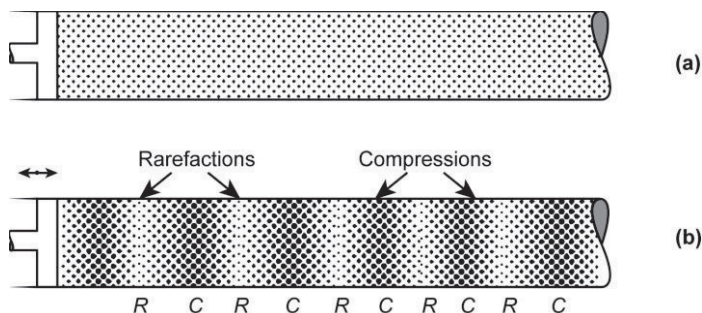


Fig. 3.1: a) A long tube containing air with a piston fitted at one end; b) Periodic motion of piston generates longitudinal sound waves comprising regions of compressions (C) and rarefactions (R).

From Unit 1, you know that mechanical waves are of two types: transverse waves and longitudinal waves. You also know that in transverse waves, such as the waves on a string, particles of the medium oscillate perpendicular to the direction of wave motion. In longitudinal waves, particles of the medium oscillate along the direction of wave motion.

When the piston moves backwards or to the left, a layer of low density is created. This low density (and low pressure) layer of air is called **rarefaction**. The rarefaction layer of air will pull the adjacent layer on its right. Thus, the region of rarefaction also keeps moving to the right in the tube. Again, when the piston moves to the right, a layer of compression is created, which propagates along the tube. Thus, the periodic motion of the piston creates successive regions of compressions and rarefactions. **These successive regions of compressions (C) and rarefactions (R) moves along the tube as longitudinal waves.** This is a longitudinal wave because the displacements of the particles of the medium is in the same direction along which the wave propagates.

The regions of compressions and rarefactions in a sound wave are the regions of high pressure and low pressure, respectively. The ear drum of our ear detects these pressure changes and causes the sensation of hearing.

You may ask: **As the sound waves propagate in tube, is there a translational motion of the particles of the medium?** No; the particles of the medium do not move along the wave. The particles of the medium execute oscillatory motion about their equilibrium position.

In Unit 1, you have learnt that waves are characterised by parameters like amplitude, wavelength, frequency and speed. The amplitude of the sound wave is the maximum displacement of the particles of the medium from their respective mean positions. The distance between two consecutive compressions (or rarefactions) is wavelength of the sound wave (Fig. 3.1b). The time taken by a particle of the medium to complete one complete oscillation is called the time period, T of the wave and the frequency, f of the wave is equal to $1/T$. The frequency of the sound wave is same as the frequency of the vibrating object producing the sound. The speed v of the sound wave is the speed with which each compression or rarefaction moves through the medium. You have learnt in Unit 1 that the relation between wavelength λ , frequency f and speed v of wave is given as

$$v = \lambda f \quad (3.1)$$

After having understood how longitudinal waves are generated and how they propagate in a medium (air), you may now ask: **Are all longitudinal waves**

sound waves? In other words, do we perceive all longitudinal waves as sound? No, it is not so. For being perceived/heard as sound by human beings, longitudinal wave must satisfy the following two conditions:

- Its frequency must be in the range 20 Hz and 20 KHz, and
- the value of its intensity, at the position of the listener, must be above (more) than “the threshold of hearing”.

You will learn about ‘threshold of hearing’ later in this unit.

When a longitudinal wave satisfy these conditions, it is heard as sound by human beings. We shall discuss these aspects of sound in detail later in the unit.

3.2.1 Graphical and Mathematical Representation of Sound Waves

You have learnt in Unit 1 that mechanical waves such as transverse waves on a string can be represented graphically as well as mathematically. Can we do the same for sound wave which is a longitudinal wave? Yes, we can do that taking analogy of transverse waves on a string discussed in Unit 1. Let us learn it now.

Graphical Representation

In Unit 1, you have learnt that transverse wave on a string can be represented graphically by two types of graphs namely vibration graph and waveform graph. Similarly, longitudinal waves can also be represented graphically by these two types of graphs. Recall from Unit 1 that in the **vibration graph**, we consider a particle of the medium at a fixed position x and plot its displacement with respect to its mean position as time passes. On the other hand, in the **waveform graph**, we keep the time fixed and plot the displacement of the particles located at different positions, x . That is, the waveform graph is the snapshot of the displacement of the particles of the medium with respect to their mean position at a given instant of time. The vibration graph and the waveform graph for a sound wave are shown in Fig. 3.2a and b. The amplitude, time period and wavelength of the sound wave are also indicated in these graphs.

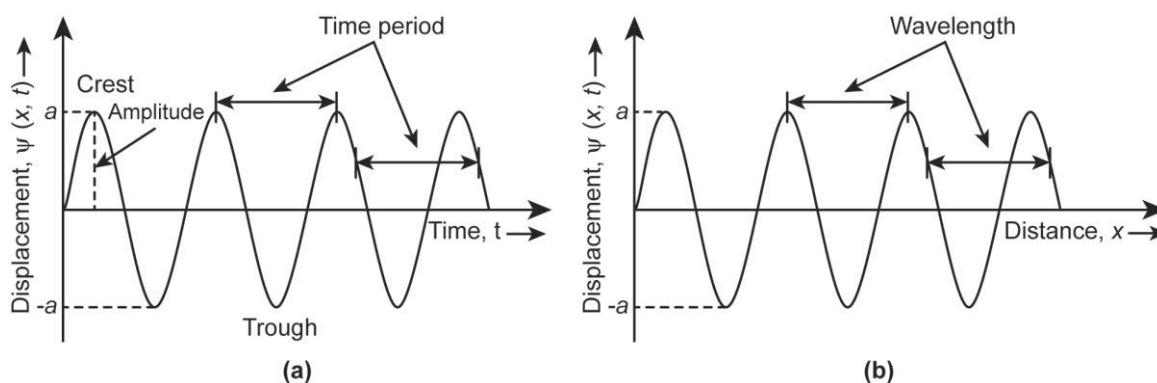


Fig. 3.2: a) The vibration graph of a sound wave showing displacement $\psi(x, t)$ of the particles of the medium at a *given position* as a function of time; b) The waveform graph of a sound wave showing displacement of the particles of the medium as a function of position at a *given time*.

You may note that the graphical representation of sound waves is identical to that of transverse wave on a string except for an important difference: While the displacement, $y(x,t)$ in Fig. 1.3 of Unit 1 denotes **transverse displacement** of the particles of the medium (string) on which transverse waves propagate, the displacement, $\psi(x,t)$ in Fig. 3.2 represents the **longitudinal displacement** of the particles of the medium (air) in which longitudinal waves propagate.

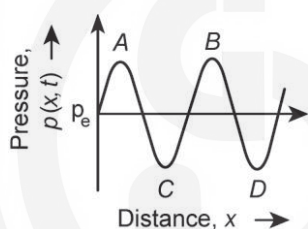


Fig. 3.3: Representation of sound wave as pressure wave.

It was shown by Fourier that complex periodic waveforms such as those generated by violin and piano can be considered as a sum of harmonic waves. The process of determining the component harmonic waves constituting a complex periodic wave is called Fourier analysis. You will learn about it if you pursue higher studies in physics.

Sound as Pressure Waves

The sound waves are also described as pressure waves. As you have learnt above, the regions of compressions and rarefactions in the sound wave in air are the regions of high pressure (or density) and low pressure (or density), respectively. To understand what we mean by pressure wave, refer again to Fig. 3.1a. Assume that, when the piston is stationary, the air in the tube is at equilibrium pressure, say p_e . As the piston begins to oscillate periodically, it generates longitudinal sound waves comprising compressions and rarefactions. The pressure in the compression regions is higher than equilibrium pressure, p_e and that in rarefaction regions is lower than p_e . So, the longitudinal wave is a successive regions of high and low pressures compared to the equilibrium pressure. Thus, the variation of pressure in a sound wave at a given time can be represented graphically as shown in Fig. 3.3. Note that points like A, B, ... represent compressions having positive value of pressure compared to the equilibrium value and points like C, D, ... represent rarefactions having negative values of pressure compared to the equilibrium value. On this pressure graph, the wavelength is the distance between two consecutive value of maximum (or minimum) pressure. From this graph, we can also determine the amplitude of the pressure wave which is the maximum positive or negative value of the excess pressure. The human ear drums can detect these pressure changes in a sound wave and cause the sensation of hearing.

Further, note that sound waves, produced due to *continuous and periodic forced oscillations* in an open medium (medium without boundaries) are **travelling periodic waves**. **Travelling waves** propagate through the medium which supports them. For example, the periodic motion of the piston (Fig. 3.1a) forces the particles of the surrounding medium (air) to oscillate and a longitudinal wave is generated. If the oscillations of the piston are simple harmonic, the travelling waves generated by it are called **harmonic waves**. Refer to Fig. 3.4 which depicts three waveforms for the sound waves: a harmonic (sinusoidal) wave, wave generated due to vibrations of the strings in a violin and a piano.

Note from Fig. 3.4 that the waveforms of the sound produced by violin and piano are periodic but not harmonic (sinusoidal). Such waves are called **anharmonic waves**.

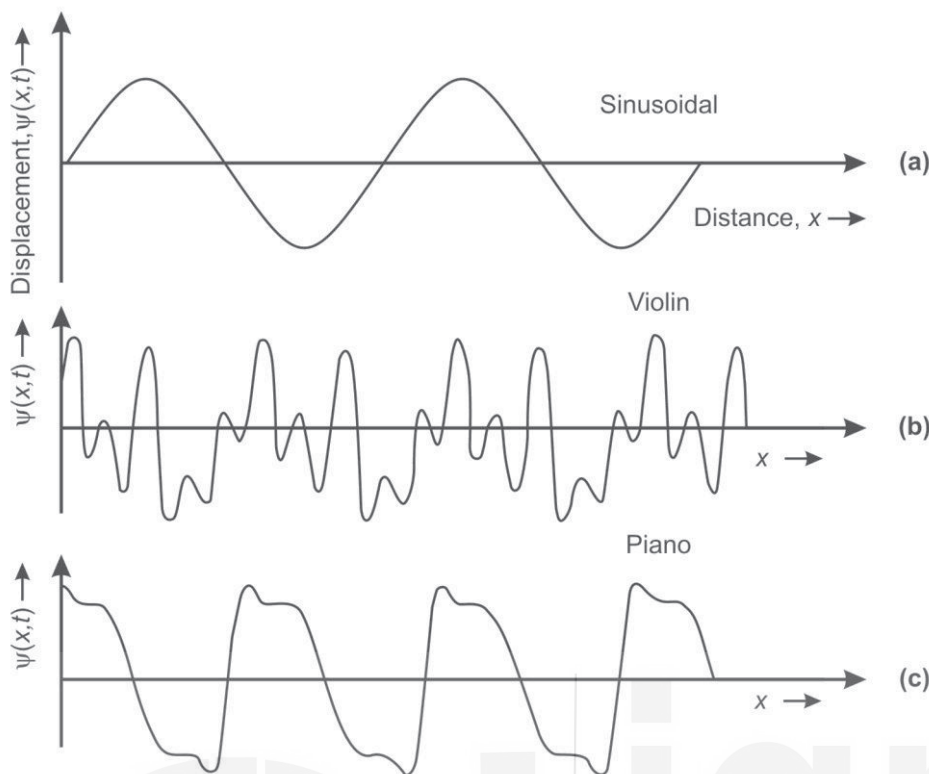


Fig. 3.4: Waveforms for a) harmonic (sinusoidal) wave; b) sound produced by a violin; c) sound produced by a piano.

In general, harmonic waves are good approximation for sound waves and we shall consider only harmonic waves for our discussion on sound.

Before proceeding further, you should answer a simple SAQ to check your understanding of what you have learnt so far in this unit.

SAQ 1 – Production and propagation of sound waves

Mark the following statements as True or False:

- Sound waves are produced by vibrating objects.
- The frequency of sound waves is not the same as the frequency of the vibrating object which produce them.
- As the sound waves propagate in a medium, there is a translational motion of the particles of the medium.
- The regions of compressions and rarefactions in a sound wave are the regions of high pressure and low pressure, respectively.

Mathematical Representation

Since sound waves are travelling harmonic waves, we can represent it mathematically taking analogy with the mathematical representation of a one-dimensional (1-D) travelling transverse waves on a string discussed in Unit 1. Thus, we can write the expression for 1-D longitudinal wave travelling in positive x -direction as

$$\psi(x, t) = a \sin(kx - \omega t) \quad (3.2)$$

where $\psi(x, t)$ is the instantaneous displacement of the particle of the medium at x with respect to its mean position at time t along the direction of

In Unit 1, you have learnt that both

$$y(x, t) = a \sin(kx - \omega t)$$

and

$$y(x, t) = a \sin(\omega t - kx)$$

are valid mathematical representation of one-dimensional harmonic wave travelling in the positive x -direction because both of them satisfy the 1-D wave equation.

propagation of the longitudinal wave, the amplitude a represents the maximum magnitude of displacement, $k \left(= \frac{2\pi}{\lambda} \right)$ is the wave number, λ is the wavelength and $\omega (= 2\pi f)$ is the angular frequency.

The wave parameters such as a , λ and f characterising the longitudinal wave have the same definition as given for transverse waves in Unit 1. Also recall from Unit 1 that the velocity of wave in terms of angular frequency ω and wave number k is given as: $v = \lambda f = (2\pi/k)(\omega/2\pi) = \frac{\omega}{k}$.

If you compare Eq. (3.1) with Eq. (1.15a), you will note that we have replaced $y(x,t)$ by $\psi(x,t)$. While $y(x,t)$ represents displacement of the particles of the medium (string) **perpendicular to the direction of propagation** of the transverse wave, $\psi(x,t)$ represents the displacement of the particles of the medium **along the direction of propagation** of longitudinal wave.

Now, before proceeding further, you should go through an example to get an idea about the numerical values of the wave parameters associated with sound wave.

EXAMPLE 3.1: PARAMETERS OF SOUND WAVES

In a fluid medium, the speed of sound waves is 275 ms^{-1} and its frequency is 500 Hz . a) Calculate the distance between two points in the medium which have phase difference of $(\pi/3)$. b) Calculate the phase difference between two displacements of a particle of the medium at a point at time 10^{-3} s apart.

SOLUTION ■ a) We know that the speed of sound wave of wavelength λ and frequency f is given by $v = \lambda f$. Thus, the wavelength is

$$\lambda = \frac{v}{f} = \frac{275 \text{ ms}^{-1}}{500 \text{ s}^{-1}} = 0.55 \text{ m}$$

We also know that two points separated by a distance equal to one wavelength, λ have phase difference equal to 2π . Thus, we can write the distance, x between two points having phase difference $(\pi/3)$ as

$$x = \frac{\lambda}{2\pi} \times \frac{\pi}{3} = 0.10 \text{ m}$$

b) The time period of the wave is given as

$$T = (1/f) = (1/500 \text{ s}^{-1}) = 2 \times 10^{-3} \text{ s}.$$

We know that the phase difference between two displacements of the particle at a point at time T apart is 2π . Therefore, the phase difference ϕ between two displacements separated in time by 10^{-3} s is

$$\phi = \frac{2\pi}{2 \times 10^{-3} \text{ s}} \times 10^{-3} \text{ s} = \pi$$

In Unit 1, you have learnt that the behaviour of a wave in space and time is described by the wave equation. So, you would like to know: **What is the**

wave equation for sound wave? Again, we take analogy of wave equation for 1-D transverse wave on a string and write the **wave equation** for the 1-D sound waves in a gaseous medium as:

$$\frac{\partial^2 \psi}{\partial t^2} = v^2 \frac{\partial^2 \psi}{\partial x^2} \quad (3.3)$$

where v is the speed of the sound wave. The speed of sound waves in a gaseous medium is given in terms of the properties of the medium as

$$v = \sqrt{B/\rho} \quad (3.4)$$

where B is the bulk modulus of the medium and ρ is the density of the medium. So, in terms of the properties of the medium, Eq. (3.3) can be written as

$$\frac{\partial^2 \psi}{\partial t^2} = \frac{B}{\rho} \frac{\partial^2 \psi}{\partial x^2} \quad (3.5)$$

Eq. (3.3) is the standard form of the **1-D wave equation for sound waves in a gaseous medium** with v given by Eq. (3.4). Note that Eq. (3.3) is similar to Eq. (1.12) – the 1-D wave equation for transverse waves on a stretched string. You must, however, recognise that the physical properties of the medium which determine the wave velocity in the two cases are different. For transverse waves on a stretched string, wave velocity is determined by the tension in the string and its linear mass density, whereas for longitudinal waves in a gaseous medium, it is determined by the bulk modulus and mass density of the medium.

Before proceeding further, let us revise the main points discussed so far.

- Sound waves are longitudinal waves produced by vibrating objects.
- Sound waves require a material medium to propagate from one point to other.
- Longitudinal waves having frequency in the range 20 Hz to 20kHz are perceived by human beings as sound.
- The intensity of longitudinal wave at the place of listener must have a value more than the threshold value called “threshold of hearing” for being heard as sound.
- The 1-D sound wave travelling in the positive x -direction is represented as

$$\psi(x, t) = a \sin(kx - \omega t); \quad \psi(x, t) = a \cos(kx - \omega t)$$

- The wave equation for one-dimensional sound wave is

$$\frac{\partial^2 \psi}{\partial t^2} = v^2 \frac{\partial^2 \psi}{\partial x^2} \quad (3.6)$$

Note that Eq. (3.3) gives the wave equation for 1-D longitudinal wave where displacement ψ is a function of x and t . But, sound waves in air are three-dimensional (3-D) longitudinal wave. Thus, the displacement ψ of the particles of the medium will be a function of x , y , z and t . So, the wave equation for three dimensional longitudinal sound wave is

$$\frac{\partial^2 \psi}{\partial t^2} = v^2 \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) \quad (3.6)$$

Recall from Unit 1 that the speed of transverse wave on a string is given in terms of properties of medium namely the tension in the string T and linear mass density ρ of the string as

$$v = \sqrt{T/\rho}$$

Recap

Now, solve an SAQ before proceeding further.

SAQ 2 – Equation of sound wave

Write down the equation for 1-D sound waves travelling along negative x-direction and having wave parameters given as: amplitude = 0.5 cm, wavelength = 32 cm and time period = 0.4 s.

From your school physics, you may recall that in an isothermal change, the temperature of the system remains constant, whereas in an adiabatic change, the temperature of the system changes.

Note that in the previous units, we have used the symbol ρ for linear mass density. We shall use this symbol for bulk (volume) mass density as well as linear mass density of a material. Thus, you should be mindful of the context of its use.

For an isothermal process, Boyle's law tells us that

$$pV = \text{constant}$$

For infinitesimal changes in p and V , we can write

$$p dV + V dp = 0$$

Hence

$$\left(\frac{\partial p}{\partial V}\right)_T = -\frac{p}{V}$$

$$\therefore p = -V \left(\frac{\partial p}{\partial V}\right)_T \equiv B_T$$

That is, pressure equals isothermal bulk modulus.

The wave equation [Eq. (3.5)] gives the speed of sound waves in terms of the physical properties of the medium. This enables us to compare the theoretical value of the speed of sound in different media with the experimentally determined values. In the following section, we discuss the speed of sound waves in different media.

3.2.2 Sound Waves in Different Media

Sound waves can propagate in solid, liquid and gaseous medium. The speed of sound varies significantly in these media due to their physical properties. Let us learn about it now.

Sound Waves in Air

The speed of sound in air can be calculated by using the value of the bulk modulus (B) and density (ρ) of the medium in Eq. (3.4). But, for a gaseous medium such as air, the bulk modulus depends on the thermodynamic changes in the medium. When sound waves propagate in the medium, it causes thermodynamic changes in it. The regions of compression have higher temperature and the regions of rarefaction have lower temperature. These changes can be **isothermal** or **adiabatic**.

Newton calculated the speed of sound in air, assuming that changes in the medium were isothermal. That is, the compression and rarefaction of the medium are not accompanied by a change in temperature. For an isothermal change, the bulk modulus of elasticity is equal to the atmospheric pressure p (see margin remark):

$$B = B_T = p$$

where B_T is the isothermal bulk modulus. Thus, we can write Eq. (3.4) as

$$v = \sqrt{p/\rho} \quad (3.7)$$

Eq. (3.7) is known as **Newton's formula** for speed of sound. For air at STP, $\rho = 1.29 \text{ kg m}^{-3}$ and $p = 1.01 \times 10^5 \text{ Nm}^{-2}$. Hence, the speed of sound in air, using Newton's formula, is

$$v = \sqrt{\frac{1.01 \times 10^5 \text{ Nm}^{-2}}{1.29 \text{ kg m}^{-3}}} = 280 \text{ ms}^{-1}$$

But, experiments show that speed of sound in air at STP is 332 ms^{-1} , which is about 15% higher than the value predicted by Newton's formula. You may now like to know: **How was this discrepancy resolved?** The discrepancy was resolved by Laplace who argued that it was wrong to assume that the thermodynamic changes in the medium caused due to propagation of sound are isothermal. It means that we cannot use Boyle's law to study wave propagation in air because it holds only for isothermal conditions. **Laplace further suggested that sound waves produce adiabatic changes: the regions of compression were hotter while the regions of rarefaction were**

cooler. That is, local changes in temperature occur when sound propagates in air. This gives rise to a larger value for the speed of sound.

For an adiabatic change, the bulk modulus, B_s is γ times the pressure (see margin remark); $\gamma \left(= \frac{C_p}{C_v} \right)$ is the ratio of molar heat capacities of a gas at constant pressure (C_p) and at constant volume (C_v). Thus, we have $B_s = \gamma p$. Then Eq. (3.4) modifies to:

$$v = \sqrt{\frac{\gamma p}{\rho}} \quad (3.8)$$

For air, $\gamma = 1.4$ and the value of the speed of sound in air at STP, calculated using Eq. (3.8) comes out to 331 ms^{-1} . This value of the speed of sound is in excellent agreement with the experimentally determined value. This shows that Laplace's argument was indeed correct.

Let us now understand how temperature affects the speed of sound in air. To do so, we express v_{sound} as a function of temperature. From your school physics and the third semester course BPHCT-135, you may recall that the equation of state of an ideal gas is

$$pV = nRT = Nk_B T \Rightarrow p = \frac{Nk_B T}{V}$$

where p is pressure, V , the volume, n , the number of moles of gas, T , the temperature, k_B , Boltzmann constant and N , the number of gas particles.

Substituting $p = \frac{Nk_B T}{V}$ in Eq. (3.8), we get:

$$v_{\text{sound}} = \sqrt{\frac{\gamma k_B T}{\rho(V/N)}} \quad \text{or} \quad v_{\text{sound}} = \sqrt{\frac{\gamma k_B T}{m}} \quad (3.9)$$

where m is the mass of an air molecule. **Eq. (3.9) tells us that the speed of sound in air (or any other gaseous medium) is a function of $\sqrt{T}$.**

We note from Eq. (3.8) that the speed of sound in a gaseous medium, such as air, is inversely proportional to the square root of gas density. Is it true even for the speed of sound in liquids and solids? If so, will sound travel slower in media denser than air? Let us find out.

Sound Waves in Liquids

The mechanism of propagation of sound waves in a liquid medium is the same as that of longitudinal waves in a gaseous medium. Thus, the wave equation given by Eq. (3.3) is also valid for sound waves in liquids. We can, therefore, write the speed of sound in liquids as [Eq. (3.4)]:

$$v = \sqrt{\frac{B}{\rho}}$$

where B is bulk modulus of elasticity for the liquid. Unlike gases, liquids are, in general, almost incompressible. This means that for producing the same change in a given volume of a liquid as that of a gas, we need to apply more pressure in the case of former. As a result, the value of the bulk modulus for liquids is very high compared to gases. Thus, the above expression indicates that the speed of sound in liquids will be significantly higher than that in gases.

For an adiabatic change, the equation of state of a gas (air) is

$$pV^\gamma = \text{constant.}$$

Therefore, infinitesimal changes in p and V are connected through the relation

$$V^\gamma dp + p \gamma V^{\gamma-1} dV = 0$$

or

$$-V \left(\frac{\partial p}{\partial V} \right)_s \equiv B_s = \gamma p$$

The ability to measure the speed of sound has been put to many uses in military operations. During World War I, a technique called **sound navigation and ranging (SONAR)** was developed to locate the position of enemy guns by using the sound of cannons in action. The technique is also being used for exploration of under sea minerals and oil. The off-shore explorations are now a routine

For water, $\rho = 10^3 \text{ kg m}^{-3}$ and $B = 2.22 \times 10^9 \text{ Nm}^{-2}$. This gives the speed of sound to be about 1500 ms^{-1} . Compare it with the speed of sound in air at STP (331 ms^{-1}). **Note that though water is about 10^3 times denser than air, yet sound propagates about five times faster in water.** This means that we can send audio messages from one ship to another faster via water than via air. This has led to the development of *sonar* (Sound Navigation and Ranging). High frequency sound waves are used in a **SONAR** to measure the depth of sea bed, detect submarines, enemy torpedoes, and for off-shore oil exploration.

Sound Waves in Solids

In a solid medium, both transverse and longitudinal waves can propagate. The mechanism of propagation of longitudinal waves in a solid is same as in case of the propagation of longitudinal waves in a gaseous medium. Thus, the wave equation for the propagation of sound waves in a solid given by Eq. (3.3).

You may now like to ask: **Which properties of the medium determine the speed of longitudinal waves in a solid?** When sound waves propagate in a solid rod, changes take place in its length; the volume remains almost constant. Thus for solids, bulk modulus (which relates the volume changes with pressure changes) of elasticity is replaced by Young's modulus of elasticity (Y) defined as

$$Y = \frac{\text{Stress}}{\text{Longitudinal strain}} = \frac{\Delta p}{\left(\frac{\Delta l}{l}\right)}$$

So, by replacing B by Y in Eq. (3.4), we can write the speed of sound in a solid as

$$v = \sqrt{Y/\rho} \quad (3.10)$$

You should now go through the following example to get a feel for the numerical values of the speed of sound in different media.

EXAMPLE 3.2: SPEED OF SOUND IN DIFFERENT MEDIA

Calculate the speed of sound in a) air at STP; b) sea water; and c) steel. Take for air at STP, $\gamma = 1.39$ and $\rho = 1.29 \text{ kg m}^{-3}$, for sea water, $\rho = 1023 \text{ kg m}^{-3}$; $B = 2.34 \times 10^9 \text{ Nm}^{-2}$; for steel, $Y = 20 \times 10^{10} \text{ Nm}^{-2}$, $\rho = 7800 \text{ kg m}^{-3}$.

SOLUTION ■ a) From Eq. (3.7), we know that the velocity of sound in air is given as

$$v_{\text{air}} = \sqrt{\gamma p / \rho} \quad (i)$$

For air at STP, we have $p = 0.76 \times 13.6 \times 10^3 \times 9.8 \text{ Nm}^{-2}$; $\gamma = 1.39$ and $\rho = 1.29 \text{ kg m}^{-3}$.

On substituting these values in Eq. (i), we get

$$v_{\text{air}} = \sqrt{\frac{(1.39) \times (0.76 \times 13.6 \times 10^3 \times 9.8 \text{ Nm}^{-2})}{1.29 \text{ kg m}^{-3}}} = 330.4 \text{ ms}^{-1}$$

The values of the Young's modulus (Y) and bulk modulus (B) of some materials are:

Material	Y (Nm^{-2})	B (Nm^{-2})
Steel	200×10^9	160×10^9
Aluminium	70×10^9	70×10^9
Copper	110×10^9	140×10^9
Glass	55×10^9	31×10^9
Diamond	1120×10^9	540×10^9
Water	0	2.2×10^9
Mercury	0	27×10^9
Most gases at STP	0	1.1×10^5

b) The expression for velocity of sound in liquid is given by Eq. (3.3):

$$v_{\text{sea-water}} = \sqrt{B/\rho}$$

As per the problem, $B = 0.23 \times 10^{10} \text{ Nm}^{-2}$ and $\rho = 1023 \text{ kg m}^{-3}$

$$\text{So, } v_{\text{sea-water}} = \sqrt{\frac{0.23 \times 10^{10} \text{ Nm}^{-2}}{1023 \text{ kg m}^{-3}}} = 1.4 \times 10^3 \text{ ms}^{-1}$$

(c) The expression for velocity of sound in solids is given by Eq. (3.10):

$$v_{\text{steel}} = \sqrt{Y/\rho}$$

As per the problem, $Y = 20 \times 10^{10} \text{ Nm}^{-2}$ and $\rho = 7800 \text{ kg m}^{-3}$

$$\text{So, } v_{\text{steel}} = \sqrt{\frac{20 \times 10^{10} \text{ Nm}^{-2}}{7800 \text{ kg m}^{-3}}} = 5.1 \times 10^3 \text{ ms}^{-1}$$

So, we find that the velocity of sound increases as the density of the medium increases. That is,

$$v_{\text{steel}} > v_{\text{water}} > v_{\text{air}}$$

Now before proceeding further, you may like to solve an SAQ.

SAQ 3 – Speed of sound wave in solids

Seismic waves travel with speed $5 \times 10^3 \text{ ms}^{-1}$. Assuming the average Earth density to be 1500 kg m^{-3} , calculate Young's modulus of the material of the Earth.

You now know that sound waves are produced and propagate in a medium when a vibrating object forces the particles of the medium to execute vibratory / oscillatory motion. The forced vibration and a related phenomenon called resonance plays another important role with respect to sound: they enhance the intensity of sound so as to make it audible at a distance. You will now learn about it.

3.2.3 Forced Vibrations and Resonance

Forced vibrations refers to the phenomenon in which an external periodic force induces vibrations in another system. To understand this process qualitatively, consider two identical tuning forks *A* and *B* of equal frequency, say 384 Hz. Let tuning fork *A* be set into vibration and brought close to tuning fork *B*. It is observed that *B* also begins to vibrate. **The tuning fork *B* starts to vibrate despite the fact that it was not struck and it did not come in contact with *A*.** The vibration of *B* is called **forced vibration** and it is caused by the compressions and rarefactions produced in air by the vibrating tuning fork *A*. That is, tuning fork *B* is **forced by tuning fork *A*** to vibrate through sound waves.

Another manifestation of the forced vibration is observed when we place a vibrating tuning fork in contact with a table or some other hard surface of larger area. The vibrating tuning fork forces the table top to vibrate. The

In the context of forced vibrations, the external periodic force is called **driving force** or **driver** and the system on which this force is applied is called **driven system**.

vibrating table top generates compressions and rarefactions in a larger volume of air because its surface area is larger, than that of the vibrating tuning fork. As a result, sound produced by the tuning fork is **enhanced** due to the forced vibrations.

In most musical instruments, forced vibrations and resonance are responsible for enhancing the intensity of sound. When an external periodic force (called **driving force**) is applied on a system (called **driven system**) capable of vibratory motion, it influences the motion of the system by altering its frequency and amplitude. To understand how this happens, let us consider a spring-mass system in which a block of mass m is attached to a spring (Fig. 3.5).

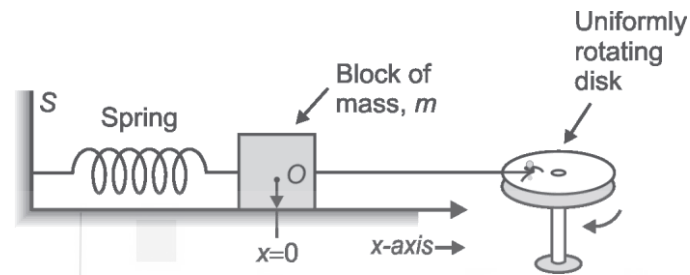


Fig. 3.5: A schematic diagram of a spring-mass system executing forced vibrations.

Let an external periodic force, given by $F(t) = F_0 \cos \omega t$, where F_0 is amplitude of the periodic force and ω is its frequency, be applied to the oscillator. Due to the periodic driving force, the spring-mass system is pushed back and forth periodically and it ultimately oscillates with the frequency ω of the driving force.

At any instant, the mass of the oscillator is subjected to three forces:

- i) restoring force, $-kx$, due to the spring,
- ii) damping force, $-\gamma \frac{dx}{dt}$ and
- iii) driving force, $F_0 \cos \omega t$.

So, the net force acting on the system is

$$F = -kx - \gamma \frac{dx}{dt} + F_0 \cos \omega t$$

According to Newton's second law of motion, this force must be equal to

$m \frac{d^2x}{dt^2}$. Thus, we have

$$m \frac{d^2x}{dt^2} = -kx - \gamma \frac{dx}{dt} + F_0 \cos \omega t \quad (3.11)$$

On dividing both sides of Eq. (3.11) by m and rearranging terms, we can write the above differential equation governing the motion of a damped forced oscillator as:

$$\frac{d^2x}{dt^2} + 2b \frac{dx}{dt} + \omega_0^2 x = f_0 \cos \omega t \quad (3.12)$$

where $2b = \gamma/m$, $\omega_0^2 = k/m$ and $f_0 = F_0/m$, and ω_0 is the natural frequency of the oscillator.

You may now ask: Does Eq. (3.11) apply only to the damped spring-mass system? No, it applies to any oscillatory or vibratory system whose natural frequency is ω_0 and which is subjected to **an external periodic driving force**.

Note that Eq. (3.12) is second order, linear inhomogeneous differential equation with constant coefficients. Its solution gives the expression for the instantaneous displacement of the forced oscillator. Here, we shall write the solution without solving Eq. (3.12) and discuss the physical significance of the expression in the context of sound waves.

The steady-state solution of Eq. (3.12) is given as

$$x(t) = a(\omega)\cos(\omega t - \theta) \quad (3.13)$$

where, amplitude $a(\omega)$ is given as

$$\begin{aligned} a(\omega) &= \frac{f_0}{[(\omega_0^2 - \omega^2)^2 + 4b^2\omega^2]^{1/2}} \\ &= \frac{F_0}{m[(\omega_0^2 - \omega^2)^2 + 4b^2\omega^2]^{1/2}} \end{aligned} \quad (3.14)$$

Eq. (3.13) represents the instantaneous steady-state displacement of a weakly damped oscillator, driven by a periodic force. From Eqs. (3.13) and (3.14), we note that:

1. **The frequency of oscillation of a forced oscillator is equal to the frequency, ω of the driving force, which is different from its natural frequency, ω_0 .** This finding tells us that the sound waves generated due to forced vibrations has the same frequency as the frequency of the sound producing vibrating object.
2. The amplitude, $a(\omega)$ of the forced oscillator depends on the amplitude, F_0 and frequency, ω of the driving force. The frequency dependence of the amplitude of the forced oscillator gives rise to the phenomena of resonance. Let us learn about it now.

Resonance

From Eq. (3.14), we note that the amplitude of vibration of the system increases manifold when the frequencies ω and ω_0 of the driving force and the driven system, respectively, are equal. This phenomenon is called **resonance**. To understand the phenomenon of resonance, let us re-examine Eq. (3.14) a little more closely. Note that for low ($\omega \ll \omega_0$) as well as for high ($\omega \gg \omega_0$) driving frequencies, the amplitude will be small because the term $(\omega_0^2 - \omega^2)^2$ in the denominator will be large. As the driving frequency, ω approaches ω_0 , the amplitude will increase gradually. The amplitude of oscillation acquires maximum value when ω is very close to ω_0 . This value of frequency, say ω_r , is called the **resonance frequency** and the condition when amplitude attains a very high value is known as **resonance**.

Now, to obtain an expression for the resonance frequency ω_r , we use a concept from differential calculus. In the language of differential calculus, the value of amplitude $a(\omega)$ will correspond to a maxima

The motion of a damped forced oscillator is governed by superposition of two oscillations: one of damped oscillations and the other of harmonic driving force. When the effect of damping is overcome by the energy supplied to the oscillator by the external periodic force, the system begins to oscillate with the frequency of the driving force. The oscillatory system is then said to be in **steady-state**.

- if the first derivative of $a(\omega)$ with respect to ω is equal to zero, and
- its second derivative is less than zero (negative) quantity,

The corresponding frequency is said to be the resonance frequency.

Let us first calculate the first derivative of $a(\omega)$ with respect to ω and equate it to zero. From Eq. (3.14) we can write

$$\begin{aligned}\frac{da(\omega)}{d\omega} &= \frac{d}{d\omega} \left[\frac{f_0}{[(\omega_0^2 - \omega^2)^2 + 4b^2\omega^2]^{1/2}} \right] \\ &= -\frac{f_0[-4\omega(\omega_0^2 - \omega^2) + 8b^2\omega]}{2[(\omega_0^2 - \omega^2)^2 + 4b^2\omega^2]^{3/2}}\end{aligned}$$

The roots of the quadratic equation

$$ax^2 + bx + c = 0$$

are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Here $b = 0$, $a = 1$ and $c = 2b^2 - \omega_0^2$. So we have

$$\begin{aligned}x &= \pm \frac{1}{2} \sqrt{-4(2b^2 - \omega_0^2)} \\ &= \sqrt{(\omega_0^2 - 2b^2)}\end{aligned}$$

At resonance frequency ($\omega = \omega_r$), we must have $da(\omega)/d\omega = 0$. Therefore, we get:

$$-4\omega_r(\omega_0^2 - \omega_r^2) + 8b^2\omega_r = 0$$

or

$$4\omega_r[\omega_r^2 - \omega_0^2 + 2b^2] = 0$$

This equation is satisfied in two situations. The first is when $\omega_r = 0$. It is ignored because it is trivial. The second situation is:

$$\omega_r^2 - \omega_0^2 + 2b^2 = 0$$

This equation is quadratic in ω_r and the acceptable root is (see margin remark):

$$\omega_r = (\omega_0^2 - 2b^2)^{1/2} \quad (3.15)$$

Note that the root corresponding to the negative sign has been ignored as it is physically meaningless.

Further, you can verify that, for the value of ω_r given by Eq. (3.15), $(d^2a/d\omega^2)_{\omega=\omega_r}$ is negative. Physically, it means that the amplitude of oscillation is maximum at resonance frequency. When such a situation is obtained, we say that **amplitude resonance** has occurred. The frequency, ω_r is referred to as the **resonance frequency**.

Eq. (3.15) shows that resonance frequency depends on damping in the system. However, if damping in the system is zero, we have $b = 0$ and we get from Eq. (3.15):

$$\omega_r = \omega_0 \quad (3.16)$$

That is, the resonance frequency is equal to the natural frequency of the driven system.

At this stage, pause for a moment and think what is happening when there is resonance between the driver (the external periodic force) and the driven system (the spring-mass system in the instant case) due to forced vibrations. Firstly, the amplitude of the driven system increases manifold. From Unit 16 of BPHCT-131, you know that the energy of an oscillator is proportional to its amplitude squared. Thus, we can say that the transfer of energy from the driver to the driven system is maximum at resonance.

Secondly, from Eq. (3.16) we note that the resonance takes place when the frequency $\omega (= \omega_r)$ of the driver is equal to the natural frequency, ω_0 of the driven system. This implies that the transfer of energy from driver to a driven system due to forced vibrations is most effective or maximum when the frequencies of the driver and the driven system are equal.

This is an important result in the context of production of sound by musical instruments like guitar and flute. You will learn about the role played by resonance in production of sound by these musical instruments later in this unit. Before proceeding further, let us recapitulate the important points of this section discussed so far.

Recap

- In forced vibration, the driven system vibrates with the frequency of the driving force. In the context of sound waves, it means that the frequency of the sound wave is the same as the frequency of the vibrating object generating the sound waves.
- If the driven system does not suffer any damping, the resonance frequency is equal to the natural frequency of the driven system.
- At resonance, the amplitude of oscillations / vibrations is enhanced manifold.

On the basis of the above discussion, you now know that forced vibrations can cause resonance under suitable conditions. Also, the amplitude of a driven system increases at resonance. These features of forced vibrations and resonance enable us to understand the production of rich variety of sound of sufficient intensity by string and air column (organ pipe) based musical instrument. Let us now briefly learn about them.

3.2.4 Musical Instruments

Musical instruments are broadly classified into two categories: string musical instruments and air column (organ pipe) based musical instruments. In the string musical instruments such as *sitar*, guitar and violin, transverse vibrations of the string is the source of sound. The strings in these musical instruments are in a stretched state and fixed at both ends. When the string is plucked (that is, disturbed from its equilibrium state), standing waves are formed. You have learnt in Unit 2 about the formation of standing waves on a string fixed at both ends. Recall that

- a) the frequency of the standing waves on a string fixed at both ends is given by [Eq. (2.14)]:

$$f_n = \frac{n}{2L} v = \frac{n}{2L} \sqrt{\frac{T}{\rho}} \quad n = 1, 2, 3, \dots \quad (3.17)$$

where T is the tension in the string, ρ , its linear mass density and L , the length of the string. Eq. (3.17) shows that the string can vibrate with different frequencies for different values of n . These frequencies are called the **normal frequencies** or **normal mode frequencies** of the string. The frequency, f_1 corresponding to $n = 1$ is called the **fundamental frequency** of the string and frequencies corresponding to higher values of n are called **harmonics** or **overtones**. Using Eq. (3.17), you can show that the

permitted frequencies of the standing waves on the string constitute a **harmonic series**:

$$f_1 : f_2 : f_3 : \dots = 1 : 2 : 3 : \dots \quad (3.18)$$

For this reason, the normal mode frequencies of the standing waves on a string are called **harmonics**: fundamental frequency is called first harmonic, the next higher frequency is called second harmonic and so on. These harmonics are also called **overtones**. However, there is a difference: **the numbering of overtones does not take into account the fundamental frequency as such**. That is, the second harmonic is called first overtone, third harmonic is called second overtone and so on. At this stage, you should refer to Fig. 2.12 which depicts the configuration of standing waves on a string fixed at both ends for the first few harmonics.

- b) From Eq. (3.17), we note that the wavelength of the standing wave on a string fixed at both ends is given as

$$\lambda_n = \frac{2L}{n}$$

So, in the fundamental mode ($n = 1$), the length L of the string accommodates half wavelength.

- c) The displacement of the string at the fixed ends is zero; such points on the standing wave are called **nodes**. And, between two nodes, there is an **antinode** which is the position of maximum displacement of the string.

From Eq. (3.17), it is obvious that sounds of different frequencies can be produced by string musical instruments. Also note from Eq. (3.17) that the frequency of the standing waves on a string is a function of tension in the string, linear mass density of the string and length of the string. Musicians, therefore, are able to change the frequency of the sound by adjusting one or the other of these physical parameters of the strings in the instrument.

The sound produced by an isolated string is usually difficult to hear from a distance of a few metres. It is so because the area of the string is small and it does not move much air in its surroundings. That is, the power delivered by the vibrating string to the longitudinal waves generated by it in the surrounding air is very low and hence it is not heard as sound from a distance. This is true even for the sound waves generated by the strings of a musical instrument.

To enhance the power (or intensity) of the sound waves generated by strings, string musical instruments invariably comprise a hollow box called **sounding box**. Sounding box is a hollow wooden box of suitable size. The enhancement in the power of the sound waves by sounding box is caused due to formation of standing waves in the air column in the sounding box. To understand this phenomenon, you need to learn the formation of standing waves in air columns called organ pipes. Let us now briefly discuss it.

Organ Pipes

An organ pipe is any tube, which has been bent into different shapes or has holes cut into it. The rich variety of sound produced by organ pipe (or air column) based musical instruments such as flute and saxophone is due to formation of standing waves. In the context of sound waves, formation of

standing waves in air columns such as organ pipes plays **two important roles**:

Firstly, a group of musical instruments called wind instruments, such as flute and saxophone produce sound due to vibrations of air columns in organ pipes. The vibrations are caused by the mouth piece of the instrument. The energy supplied by these vibrations maintains the vibrations in the air column of the organ pipe.

Secondly, air columns called **sounding box** are attached with string musical instruments to enhance the intensity of the sound produced by the vibrating strings of the instrument.

Both the above functions of organ pipes are based on the formation of standing waves in the air columns in them. The organ pipe or tubes which are normally used in the musical instruments are classified into three categories: closed-closed organ pipe, closed-open organ pipe and open-closed organ pipe.

We will not go into the detailed mathematical analysis of the formation of standing waves in these organ pipes as it is beyond the scope of this course. It is sufficient to know that the formation of standing waves in organ pipes is analogous to the formation of standing waves on a string; that is, the standing waves are formed due to the superposition of two waves travelling in opposite directions. In case of string, the waves travelling in opposite directions are transverse waves and in case of organ pipes, the two waves are longitudinal waves. Also, the position of nodes and antinodes in the standing waves depends on its construction; that is, whether it is a closed-closed, or closed-open, or open-open pipe.

Before proceeding further, solve an SAQ.

SAQ 4 – Calculating frequency of harmonics of a musical instrument

The wavelength of the standing wave on a guitar string vibrating in its fundamental mode is 2 m. If the speed of wave on the string is 400 ms^{-1} , what is the frequency of its second overtone?

With these brief description of organ pipes and the standing waves formed in its air column, we can understand the role of sounding box in sound produced by string musical instruments. As we mentioned above, the sound waves generated by the vibrating string of the instrument is barely audible at a distance of a few meters or so. But, if a sounding box is attached near the vibrating string, the air column inside the sounding box is set to forced vibrations. And, if the frequency of the waves generated on the string matches with one of the normal modes frequencies of the air column, resonance takes place and a loud sound is produced.

So far, we have discussed how sound is produced, its speed in different media and the role of forced vibrations and resonance in production of sound. When a sound wave reaches human ear, the ear converts the pressure changes into nerve impulses which is interpreted by brain as something

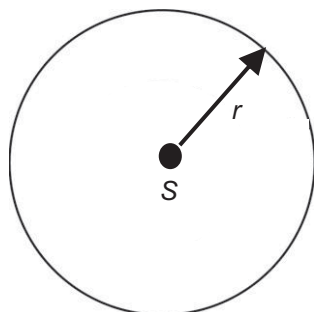


Fig. 3.6: Sound waves emitted by a point source S spread in the surrounding space.

heard. **However, the sound wave must have sufficient intensity to produce appreciable pressure changes at the ear drum.** Let us now learn about intensity of sound at a point away from the source and how it is measured.

3.3 INTENSITY OF SOUND

In Unit 1, you have learnt that the **intensity of a wave** is defined as **the rate of energy transfer by the wave per unit area normal to the direction of wave propagation**. Mathematically, we can write the intensity (I) of a wave as

$$I = \frac{P}{A} \quad (3.19)$$

where P is the power of the wave and A is area of the surface intercepting the wave perpendicular to the direction in which the wave travels (Fig. 3.6).

You must have experienced that loudness of sound decreases as we move away from its source. You may therefore like to know: **How does the intensity of sound vary with distance from the source?** You have learnt about this in Sec. 1.4 of Unit 1. Recall that for an arbitrary point located at distance r from the source, the intensity of sound is given as:

$$I = \frac{P}{4\pi r^2}$$

That is $I \propto \frac{1}{r^2}$ (3.20)

Eq. (3.20) is known as the **inverse square law**. You have also learnt in this section (Sec. 1.4, Unit 1) that the amplitude a of the wave is inversely proportion to the distance r :

$$a \propto \frac{1}{r} \quad (3.21)$$

Eq. (3.21) shows that **amplitude of a wave decreases with increasing distance of the listener from the source of sound**.

3.3.1 Intensity Level: Decibel

At this stage, you may ask: **How is our hearing related to the intensity of sound waves?** A normal person can hear sounds over a very wide range of intensities. When we speak in normal tone, we emit energy at the rate of about 10^{-5} W . Suppose that a person opens his/her mouth by about 10 cm^2 while speaking. Then, the intensity of sound is $(10^{-5} \text{ W})/10^{-3} \text{ m}^2$. Its value is 10^{-2} Wm^{-2} . Let us denote this by I_0 . A normal person can hear a sound wave of intensity as low as $10^{-10} I_0$ or 10^{-12} Wm^{-2} . The sound of this intensity is called the **threshold of hearing** (ToH) for human beings. On the other hand, we can hear sound of as high an intensity as 10 Wm^{-2} without any damage to our ears.

So, we find that the range of intensities which a human ear can detect is very large ($\sim 10^{13} \text{ Wm}^{-2}$). Physicists, therefore, use a *logarithmic scale* called the **decibel (dB) scale** to measure the intensity of sound. Decibel scale is a **relative scale** which expresses the intensity of a given sound wave with respect to the threshold of hearing.

The logarithmic scale used to express intensity of sound is Bel, named after **Alexander Graham Bell**. For finer gradation, it is customary to express intensity of sound by tenths of a bel or a decibel (dB).
(1 bel = 10 decibel).

The decibel scale differs from most measurement scales in two ways. Firstly, most of the measurement scales such as that for length, mass and temperature are absolute and linear. **The decibel scale is, however, relative rather than absolute and logarithmic rather than linear.** A linear scale such as temperature is characterised by the fact that the difference in temperature between 20°C and 21°C is same as the difference in temperature between 90°C and 91°C: the difference in temperature is 1°C in both the cases. So, in a linear scale, jumping from one value to another means **adding** a constant value. But, in non-linear scale like logarithmic scale, jumping from one scale value to another requires **multiplying** by a constant.

The decibel scale is a relative scale in the sense that it expresses intensity not in absolute terms but in relation to some fixed value called threshold of hearing ($\sim 1 \times 10^{-12} \text{ W m}^{-2}$).

The unit for measuring intensity level of sound, bel is defined as

$$\text{bel} = \log_{10} \left(\frac{I}{I_0} \right)$$

where I is the measured intensity and I_0 is the threshold of hearing. To have a wider range of values for intensity of sound, intensity is measured in **decibel** which is 10 times the **bel**. Thus, the **intensity level** (β) of sound in decibel scale corresponding to intensity I is given by

$$\beta = (10 \text{ dB}) \log_{10} \left(\frac{I}{I_0} \right) \quad (3.22)$$

where, I_0 is the intensity corresponding to the threshold of hearing ($1 \times 10^{-12} \text{ W m}^{-2}$).

Since the decibel scale is logarithmic, you can see that it is based on multiples of 10. The threshold of hearing is assigned a decibel level of zero dB. This corresponds to the intensity of $10^{-12} \text{ W m}^{-2}$. A sound of intensity $10^{-11} \text{ W m}^{-2}$, which is 10 times more than the threshold of hearing is assigned a decibel level of 10 dB. A 1000 times more intense (10^{-9} W m^{-2}) sound is assigned a level of 30 dB. You can verify these numbers using Eq. (3.22).

In Table 3.1, we have listed the intensity of waves generated by different sources of sound and also provided the corresponding intensity level in dB at a distance of 1 m from the source.

Table 3.1: Intensities of sound at a distance of 1m from typical sources

Source of Sound	Intensity (W m^{-2})	Intensity Level (dB) (with respect to ToH)
Threshold of hearing (ToH)	10^{-12}	0
Rustling of leaves	10^{-11}	10
Whisper	10^{-10}	20
Normal conversation	10^{-6}	60
Street traffic	10^{-5}	70
Bursting of crackers (at 1m)	8×10^{-5}	78
Heavy road traffic	10^{-2}	100
Threshold of pain	10	130
Jet take off	10^2	140

To understand how a logarithmic scale works, let us consider the relation

$$a = \log b$$

where a and b are variables. If we multiply b by 10, we get

$$\begin{aligned} a' &= \log (10 b) \\ &= \log 10 + \log b \\ &= 1 + a \end{aligned}$$

That is, when we multiply b by 10, a increases by 1.

Similarly, if we multiply b by 10^6 , we get

$$\begin{aligned} a' &= \log (10^6 b) \\ &= 6 \log 10 + \log b \\ &= 6 + a \end{aligned}$$

That is, when b is multiplied by 10^6 , a increases by 6, the value in the exponent.

Note that the zero point on the decibel scale (dB) does not imply 'no sound'. Instead, it signifies the threshold of hearing i.e., the lowest sound that can be heard. Also, the sound (~ 140 dB) of jet taking off can cause pain in our ears. To save workers in airports, buildings are made sound proof. However, people living in the vicinity of airports are likely to suffer noise pollution.

To check your understanding of decibel scale for intensity levels of sound you may like to answer an SAQ.

SAQ 5 – Intensity level of sound

In a factory, it is required that the sound intensity level should remain below 80 dB to protect workers from noise pollution. Calculate the maximum sound intensity permitted in the factory.

Sound, in general, is classified into two types: musical sound and noise. Although this classification is based on our perception, we can always try to understand the difference in terms of concepts of physics. You will now learn the physical characteristics of musical sound and also about musical scale used for producing musical sound.

3.4 MUSICAL NOTES AND SCALE

In the context of musical sound, 'note' and 'tone' both refer to its 'pitch'. Since, 'pitch' of a sound is related to its frequency, we essentially refer to the frequency of sound when we mention 'note' or 'tone' of a sound.

As far as perception or hearing is concerned, we can define musical sound as sound which produces a pleasant sensation to the ear. A musical sound contains regularly sustained **notes** produced by a vibrating object. On the other hand, noise is an abrupt sound which is not sustained for long. Due to this abruptness, noise produces unpleasant sensation to ear.

From the point of view of perception, a sound is characterised by three parameters, namely, pitch, loudness and quality. These three **attributes of sound** are broadly related to physical characteristics of the sound wave. The pitch of sound is related to the wave frequency, loudness to wave intensity (or amplitude) and quality to the waveform. Let us briefly discuss these relations.

Pitch and Frequency

As a matter of fact, the subjective perception of sound by human ear is a physiological process. But, this physiological process is influenced by the physical characteristics of sound waves. It is found that the frequency of sound influences the perceived pitch of sound: higher the frequency, higher is the perceived pitch. You know that pitch of sound enables us to **distinguish a grave sound (or note) from a sharp one. It is found that two frequencies that differ by a factor of 2 are perceived as pleasant sound when they are sounded together. Two such frequencies are said to differ by an octave and this plays an important role in developing musical scales** as you will learn later in this section. The subjective perception of equal steps in pitch level roughly corresponds to multiples in the frequency of harmonic sound waves. **Thus, two successive notes (pitch level) in music are not equally spaced in frequency.**

Loudness and Intensity

The loudness of perceived sound is correlated to its intensity (that is, energy carried by sound waves per unit area at the position of listener). As you have learnt in the previous section, the intensity of sound at the position of a listener must have a value more than the threshold of hearing. Further, the correlation between loudness and intensity is similar to the correlation between pitch and frequency; that is, equal steps in loudness correspond to equal multiples in intensity.

Quality

Even if two sound waves have the same pitch and loudness, they may be perceived as two distinct sounds. This difference is caused due to the **waveforms** of the two sounds if they are not harmonic waves. To understand what is meant by the 'waveform' of sound, refer to Fig. 3.4 where we have depicted the waveforms of the sound produced by violin and piano. You may note that these two waveforms are different from each other which influences the perceived quality of respective sounds. You may like to know: How such periodic but anharmonic waveforms are generated?

It was shown by Fourier that a complex periodic waveform (that is, a waveform which is periodic but not harmonic) can be regarded as a sum of harmonic waves. So, when a musical instrument generates sounds of different harmonics or overtones, they combine to produce a complex waveform. Further, the possible harmonics or overtones which can be generated depend up on the nature of the musical instrument. Therefore, the waveforms, corresponding to a given pitch and loudness, produced by a musical instrument is its unique characteristic. That is why we are able to distinguish between the same notes produced by flute and *sitar*.

A complex periodic waveform can be analysed using Fourier analysis and the relative amount of fundamental frequency and the higher harmonics constituting the waveform can be ascertained. Similarly, we can add various amounts of higher harmonics to the fundamental frequency of a musical instrument and produce a desired waveform.

With the above background, you are now ready to learn about musical scales.

3.4.1 Musical Scale

Musical scale used by musicians to produce musical sound is based on the concept of **octave**. **In a musical scale, an octave is defined as a set of two frequencies wherein the value of the higher frequency is two times the value of the lower frequency.** For example, if 220 Hz is the frequency of the lower note of the octave, the value of the higher note will have the frequency 440 Hz. The importance of octave in music is due to the fact that two frequencies that differ by a factor of 2 are particularly pleasing when sounded together and they are said to form an octave.

The human ear can distinguish a number of notes of definite frequencies between a note and its octave. The musical scale used in Indian (Hindustani as well as Carnatic) music as well as Western music is based on this series of

distinguishable notes in an octave. The lowest frequency note in this series is called **key note** or **tonic**. The pitch corresponding to frequency of 256 Hz is taken as key note for scientific analysis of musical sound. However, musicians use pitch corresponding to frequency of 264 Hz as key note or tonic for creating / producing music.

A musical scale is characterised by the relation among the notes in an octave. Let us now learn about a few musical scales.

1. Diatonic Scale

This is the most common musical scale. It consists of a series of **eight notes** with the interval between the last and the first note in the ratio 2:1. Thus, the range of notes is called an octave. This musical scale is common to Indian traditional music and Western music. In Indian traditional music, it is also called *saptak* comprising seven basic notes; the eighth (or the last) note of the octave is not counted because it is double the frequency of the first note. The pitch of a note is given a name such as *sa, re, ga, ...* etc. in Indian music tradition and *do, re, mi, ...* etc. in Western music tradition. A note is a sound wave of given frequency. The seven pure notes of the octave in the diatonic scale are assigned names and the frequency of each note is related to the frequency of the preceding note by a fixed ratio. The names of these pure note in the Indian and Western music traditions and the ratio specifying the relation between two consecutive notes of the diatonic scale is given in Table 3.2 below.

Table 3.2: Frequency relation between notes of the diatonic scale

	Name of pure notes							
	1 (Keynote or tonic)	2	3	4	5	6	7	8
Indian Traditional Music	<i>sa</i>	<i>re</i>	<i>ga</i>	<i>ma</i>	<i>pa</i>	<i>dha</i>	<i>ni</i>	<i>sa</i>
Western Music (C-Major Scale)	<i>do</i>	<i>re</i>	<i>mi</i>	<i>fa</i>	<i>sol</i>	<i>la</i>	<i>si</i>	<i>do</i>
Frequency Ratio	1	9/8	5/4	4/3	3/2	5/3	15/8	2
	1	1.125	1.250	1.333	1.500	1.667	1.875	2

Note from Table 3.2 that the frequencies of the notes in the octave are arranged in ascending order. Also, note that the frequency intervals of the musical scale are not uniform. However, the sensation of pitch produced by these notes is of equal interval. If we take 1 as the key note of an octave comprising *saptak* – the basic seven notes – then the frequencies of these seven basic notes are derived according to the ratio given in Table 3.2.

To understand this scale in terms of frequencies of the basic notes, study the following example.

EXAMPLE 3.3: FREQUENCY OF BASIC NOTES OF AN OCTAVE

Suppose the frequency of the key note (or tonic) of an octave is 256 Hz. Calculate the frequencies of the seven basic notes of the corresponding octave of the diatonic scale.

SOLUTION ■ Since the frequency of the key note (lower frequency) of the given octave of the diatonic scale is 256 Hz, the higher frequency of the octave is 512 Hz. And, the frequencies of other six basic notes are as given in the table below:

Note	<i>sa</i>	<i>re</i>	<i>ga</i>	<i>ma</i>
Frequency (in Hz)	256	256×1.125 = 288	256×1.250 = 320	256×1.333 = 341.3

<i>pa</i>	<i>dha</i>	<i>ni</i>	<i>sa</i>
256×1.500 = 284	256×1.667 = 426.7	256×1.875 = 480	256×2 = 512

You should note that the relation among the frequencies of diatonic scale as given in Table 3.2 is with respect to the key note, that is, the lower frequency of the octave. You may ask: **How are the frequencies of the basic notes of the scale related to each other?** The ratio between frequencies of two successive notes of a musical scale is called **musical interval**. The musical intervals of diatonic scale is given in the Table 3.3.

Table 3.3: Musical intervals of diatonic scale

Notes	<i>sa</i> (f_1)	<i>re</i> (f_2)	<i>ga</i> (f_3)	<i>ma</i> (f_4)	<i>pa</i> (f_5)	<i>dha</i> (f_6)	<i>ni</i> (f_7)	<i>sa</i> (f_8)
Musical Intervals	256	(f_2 / f_1) = 9/8	(f_3 / f_2) = 10/9	(f_4 / f_3) = 16/15	(f_5 / f_4) = 9/8	(f_6 / f_5) = 10/9	(f_7 / f_6) = 9/8	(f_8 / f_7) = 16/15

From Table 3.3, we note that the musical interval between key note and the next higher note in the diatonic scale is 9/8. Also, from Table 3.3 we note that the three main musical intervals in the diatonic scale are 9/8, 10/9 and 16/15 which are, respectively, known as a major tone, minor-tone and semi-tone. Thus, the sequence of intervals in the diatonic scale are major-tone, minor-tone and semi-tone. Since major-tone occurs more frequently, this scale is also called **major diatonic scale**. If we neglect the difference between major and minor tones, then diatonic scale will be tone, tone, semi-tone, tone, tone, tone, semi-tone. **In view of having mainly two basic ratios as musical intervals, this scale is called diatonic scale.**

As evident from Table 3.3, the musical interval between two successive notes (frequencies) are not equal in the diatonic scale. This puts severe limitation on a large number musical instruments. To overcome this shortcoming of the

diatonic scale, another musical scale, called tempered scale was developed. Let us learn about it now.

Tempered Scale

In the tempered scale, the octave is divided into 12 **equal intervals**. Thus, if x is the musical interval between two consecutive notes on tempered scale, then

$$(x)^{12} = 2 \Rightarrow x = (2)^{1/12} = 1.05946$$

Thus, the equal musical interval between two consecutive notes, (that is, the ratio between the frequencies of two consecutive frequencies) on a tempered scale is 1.059456 .

In this scale, all the other notes in the octave are defined in relation to tonic (sa). For notation purposes, each of the 12 notes in an octave has a unique identity/name/notation as given below:

Table 3.4: Notations of Indian and Western music for tempered scale

Note Notation		(Key note or tonic)	2	3	4	5	6	7	8	9	10	11	12	13
	Indian Music Note (Symbol)	S	r	R	g	G	M	m	P	d	D	n	N	S'
	Western Music Note (Symbol)	C	C [#]	D	D [#]	E	F	F [#]	G	G [#]	A	A [#]	B	C'

Because of equal musical interval between two consecutive notes on a tempered scale, it is also called **Equal Tempered Scale (ETS)**.

In the Indian music tradition, the 12-notes scale comprises **seven pure notes** and **five impure notes**. In the Table 3.4, pure notes are shown by upper case letters and the impure notes are shown by lower case letters. So, as per Indian music system, pure notes are S, R, G, M, P, D, and N. Also note that the seven pure notes correspond to C-Major scale of the Western music tradition. In Indian traditional music, an octave was divided into 22 intervals. But, the 12 tone based tempered scale is more prevalent.

Like light, sound waves are also reflected, refracted and diffracted. Due to these phenomena, hearing experience is affected, particularly in rooms, halls and auditoriums. The study of impact of these phenomena and the remedial measures to minimise their undesirable impact on hearing experience is called acoustics of buildings. You will now learn about it.

3.5 ACOUSTICS OF BUILDINGS

Sound produced in a room, auditorium or any closed space can lead to various undesirable effects such as generation of noise and lack of clarity. The

purpose of studying the acoustics of building is to minimise these undesirable effects.

The procedure to design a building which do not have any undesirable acoustic effects, is to first anticipate the possibilities of noise and make provisions to reduce the disturbances. For clear and comfortable listening experience, the following factors are taken into consideration at the stage of planning, design and construction of a building:

- a) Reverberation
- b) Echo
- c) Focussing
- d) Echelon Effect
- e) Loudness
- f) Resonance
- g) Noise

Let us now briefly discuss how the above factors affect sound in buildings and what are the remedial measures adopted to minimise their adverse impact on the quality of the sound heard by us.

a) Reverberation

The sound waves produced in a big room or auditorium are reflected from its walls, ceilings and all other objects. A listener will receive direct sound waves from the source and the sound waves reflected from all the surfaces of the room. So, even when the production of sound wave has stopped and the listener does not receive any direct sound, the same sound waves are present in the hall in the form of reflected waves. The listener, therefore, will continue to hear the same sound. **The persistence of sound even when the source has stopped producing it is called reverberation.**

Reverberation, as such, is not a completely undesirable effect.

Reverberation for some time is desirable because it enhances the richness/quality of sound. But, if it is for a longer time, it produces undesirable effect because it interferes with the next sound. So, the question is: For how long reverberation is desirable? The reverberation time is the time for which a sound persists in a hall. **It is measured as the time between when the source stopped producing sound and when the sound becomes inaudible.**

Thus, you may ask: What is the optimum reverberation time? For a room or hall, reverberation time is quantitatively defined as the time in which the intensity of sound reduces by 60 dB. For example, suppose the intensity of a sound of 140 dB reduces to 80 dB in 5 seconds. Then, the reverberation time will be 5 seconds and it is written as T_{60} time. Typically, the reverberation time less than 3 seconds is considered good.

Since the reflection of sound waves is the main cause of reverberation, the walls, ceilings and other reflecting surfaces in the hall are covered with sound absorbing materials such as cardboard sheets, perforated sheets, felt, heavy curtains, etc.

b) Echo

Echo is another undesirable phenomenon caused due to reflection of sound waves from various surfaces in a hall. When the time difference between the arrival of direct sound and the reflected sound to a listener is less than 60 ms, the two sounds are heard as one composite sound. However, if this time difference is more than 60 ms, we hear two distinct sounds; the first is the direct sound and the second is the reflected sound. The second sound is called **echo**. Echo is generally caused due to direct reflection of sound of short duration from a surface of large area.

The remedial measures for echo are to design the construction of internal structure of the auditorium in such a way as to avoid large surface areas and to cover the surfaces with porous materials which have high absorbing capacity for sound waves.

c) Focussing

If the interior walls and ceiling of an auditorium are such that reflected sound waves converge or focus at a point, there is an abnormally high intensity of sound at that point. This effect is called focussing. The undesirable effect of large intensity due to focussing can be remedied by proper design of the interior walls and ceiling of the hall and covering the reflecting surfaces with porous absorbing materials.

d) Echelon effect

This effect refers to the production of musical notes when in-phase reflected sound waves combine with each other. Such in-phase reflected waves are most likely to be generated if the hall has steps of roughly equal widths. The sound produced due to echelon effect interferes with the original sound and makes it confusing.

The echelon effect can be avoided by designing the steps of unequal widths and cover them with sound absorbing materials.

e) Loudness

For pleasant and satisfactory hearing experience, the loudness of the sound must be sufficient and uniform all across the hall. Loudness of the sound is adversely affected by the use of absorbing materials to minimise the effects of echo, reverberation, etc. Thus, absorbing materials to cover the interior walls and ceiling of the hall should be used judiciously. Also, the loudness of the sound is enhanced by designing hard and smooth reflecting surfaces near the source of sound.

f) Resonance

Sound waves in a building can induce vibratory motion in its window panes, walls, air enclosed between walls, etc., which in turn, produce sound waves when they vibrate. If the frequency of the sound waves produced by them matches with the frequency of the original sound produced in the hall, resonance may take place. Due to such resonances, intensity of some of the original musical notes (sound) get enhanced. This results in a distortion in the hearing experience.

The resonance effect can be remedied by proper maintenance of the objects in the hall so that their vibrations are suitably damped.

g) Noise

It refers to any unwanted sound in the hall which distorts the hearing experience. The noise can be due to sound coming from outside the hall such as the sound of traffic, market place, nearby railway station or airport. The effect of such external sources of noise can be minimised by proper noise insulation of the hall as well as by having green cover in the form of trees around the hall. Noise can also be due to internal sources such as air conditioners, movement of people, etc. The impact of such internal sources of noise can be minimised by proper maintenance and having carpeted floors in the hall.

Finally, we end this unit with a brief qualitative discussion on Doppler effect for sound.

Doppler Effect for Sound

Doppler effect refers to the apparent change in frequency of a wave due to relative motion between source and detector (an instrument or a human being). We use the word **apparent** because the difference in perceived and emitted frequencies lasts only as long as there is relative motion between the source and the listener.

This effect is valid for mechanical waves as well as electromagnetic waves. Since sound (mechanical waves) and light (electromagnetic waves) can be 'detected' (perceived) by human beings, it is common to use the term listener in place of detector in case of sound and observer in case of light.

According to Doppler effect, when the source approaches the observer/listener or the observer/ listener approaches the source, or both approach each other, the apparent frequency is higher than the actual frequency of the wave emitted by the source. On the other hand, when the source moves away from the observer/ listener, or when the observer/ listener moves away from the source, or when both move away from each other, the apparent frequency is lower than the actual frequency of the wave emitted by the source. The difference between the actual frequency and the apparent frequency is called **Doppler shift**. In the following, we confine our discussion to the Doppler effect for sound.

To analyse the Doppler effect for sound and obtain an expression for the apparent frequency, let us assume that: (a) the motion of the source and/or the listener is along a line joining the two; and (b) the speed of source is significantly less than the speed of sound. Under these simplifying assumptions, we consider the following cases:

- i) The source is stationary and the listener is in motion.
- ii) The source is in motion and the listener is stationary.

Source Stationary and Listener in Motion

Let us suppose that a stationary source S is producing sound of frequency f and wavelength $\lambda \left(= \frac{v}{f} \right)$; where v is the speed of sound. The sound waves

emitted by the source spread out from S as spherical wavefronts. Since the frequency of the emitted sound is f , the number of complete wavelengths (or f wavefronts) produced per second by the source will be f . The frequency of the sound heard by the listener equals the rate at which she receives the wavefronts. If the listener is stationary, the rate at which she receives wavefronts is equal to f , the rate at which wavefronts are produced by the source. So, if the source and the listener both are stationary, we should expect no change in the frequency of sound.

Now suppose that the listener moves with speed v_L towards the source.

Since the listener is moving towards the source, she will receive more wavefronts per unit time compared to the case when she was stationary. Therefore, the frequency, say f' , of sound heard by the listener will be greater than the frequency of the sound emitted by the source. The expression for the apparent frequency f' is given as

$$f' = f \left(\frac{v + v_L}{v} \right) \quad (3.23)$$

where f is the frequency of sound wave emitted by the source, v is the velocity of sound and v_L is the velocity of the listener moving towards the source.

Now, you may ask: **Will the apparent frequency f' be greater than or equal to the actual frequency f when the listener moves away from the source?** In this case, the apparent frequency is given as

$$f' = f \left(\frac{v - v_L}{v} \right) \quad (3.24)$$

Eq. (3.24) shows that, when the listener moves away from a stationary source, the value of apparent frequency is less than the actual frequency f . So, the combined expression for the apparent frequency for stationary source and listener in motion is obtained by combining Eqs. (3.23) and (3.24):

$$f' = f \left(\frac{v \pm v_L}{v} \right) \quad (3.25)$$

Source in Motion and Listener Stationary

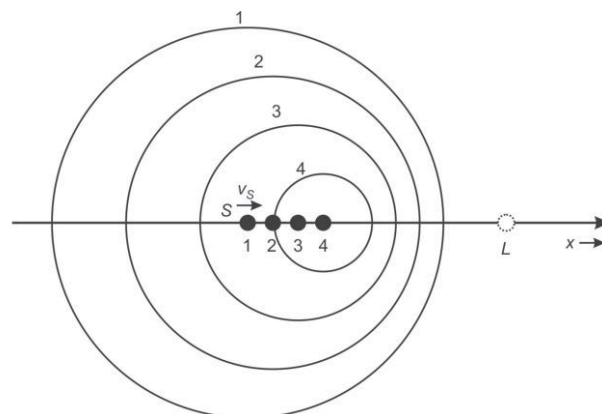


Fig. 3.7: Spherical wavefronts of the sound waves emitted by a moving source, S . The source is moving towards a stationary listener, L . Note that the separation between wavefronts is less in between S and L compared to the opposite side.

Let us now consider the situation when source S , producing sound of frequency f and wavelength λ is moving with speed $v_s (< v)$ towards a stationary listener L . Let S be at positions marked 1, 2, 3, 4, and so on while it moves towards L (Fig. 3.7). The spherical wavefronts emitted by the moving source at the instants when it is at these positions are shown by correspondingly marked circles. (These circles are drawn by keeping positions marked 1, 2, 3, 4 and so on as centres.) Note that the wavefronts emitted by the moving source S are relatively more closely spaced in the direction of motion of the source and are relatively more widely separated on the opposite side. Since wavelength is equal to the distance between two successive wavefronts, its value will be shorter than the normal for the listener L towards whom S is moving. **That is, the motion of the source towards the stationary listener changes the wavelength of the sound wave emitted by it.** Since these wavefronts move with constant speed, the change in wavelength causes a change in frequency. In the instant case, wavelength between the source and the listener decreases. Therefore, the frequency detected by L will be higher than the frequency of sound emitted by S .

You may now ask: **What happens if S is moving away from L ?** This situation is obtained if L is located to the left of the source S in Fig. 3.7. For L at any location to the left of the source which is moving towards right, the separation between any two consecutive wavefronts is larger than when S is stationary. This implies that the wavelength will be larger than the normal situation and hence frequency detected by the listener will be lower than the frequency of sound emitted by S .

The expression for the apparent frequency f' when the source is moving and the listener is stationary is:

$$f' = f \left(\frac{v}{v \mp v_s} \right) \quad (3.26)$$

where the negative sign in denominator signifies the case where the source moves towards the listener and the positive sign corresponds to the source moving away from the listener.

Thus, we find that the apparent frequency of sound changes whenever there is a relative motion between the source and the listener. **However, you must note that the reasons for the change in frequency in the two cases discussed above are different.** When the source is stationary and listener is in motion, the change in frequency is caused due to the change in the number of wavefronts (or the number of complete wavelengths) received by the listener. On the other hand, when the listener is stationary and the source is moving, change in frequency is caused due to the change in the wavelength of the sound.

Doppler effect for sound waves finds many applications. Do you know that Doppler shift in ultra-sound waves reflected from moving body tissues allows measurement of blood flow? It is commonly used by obstetricians to detect foetal heart-beat. Do you know how it arises? As the heart muscle pulsates, the frequency of reflected ultra-sound waves is different from the frequency of incident waves. Similarly, a sonar makes use of the Doppler effect in determining the velocity of a submarine relative to a ship.

The electromagnetic waves also exhibit Doppler effect. In aircraft navigation, radar works by measuring the Doppler shift of high frequency radio waves reflected from moving aeroplanes. The Doppler shift of star-light allows us to study stellar motion. If the stars were stationary, the spectrum of light from a star will have spectral lines corresponding to, say hydrogen, at the same frequency (or wavelength) as in the hydrogen spectrum obtained on the earth. However, if the star is moving away from us (that is, the earth), the hydrogen spectral lines in the spectrum of star will be shifted towards the red-end of the spectrum due to Doppler effect. And, on the other hand, if the stars are approaching us, the spectral lines will be shifted towards the blue-end of the spectrum.

When the spectrum produced by light from stars is examined, it is observed that these lines are slightly shifted, as compared to the corresponding lines from the same elements on the earth, this shift is generally towards the **red-end**, which implies that wavelength of light emitted by the star increases. This observation is interpreted as if the star is receding, that is, the star is moving away from us. This find supports the expanding universe theory of the evolution of the universe.

Before proceeding further, let us recapitulate what you have learnt in this section.

3.6 SUMMARY

Concept	Description
<i>Production of sound</i>	■ Sound is a longitudinal wave produced by vibrating object. Sound waves comprises successive regions of compressions and rarefactions. Like any mechanical wave, sound needs a material medium for its propagation..
<i>Perception of sound</i>	■ Longitudinal waves having frequency in the range 20 Hz to 20 kHz are perceived as sound by human beings.
<i>Threshold of hearing</i>	■ It is the minimum value of the intensity of sound at the position of listener for being heard by human beings.

Mathematical representation of sound wave

- The 1-D sound waves travelling in the positive x-direction is represented by anyone of the two following equivalent expressions:

$$\psi(x, t) = a \sin(kx - \omega t)$$

$$\psi(x, t) = a \sin(\omega t - kx)$$

The 1-D sound waves propagating in the negative x-direction is represented as

$$\psi(x, t) = a \sin(kx + \omega t)$$

$$\psi(x, t) = a \sin(\omega t + kx)$$

where $\psi(x, t)$ is the longitudinal displacement of the particles of the medium along the direction of wave propagation.

Speed of sound waves

- In a gaseous medium, the speed of sound is given by

$$v = \sqrt{\frac{\gamma p}{\rho}} = \sqrt{\frac{\gamma k_B T}{\rho}}$$

In liquid medium, the speed of sound is

$$v = \sqrt{\frac{B}{\rho}}$$

In solids, the speed of sound is

$$v = \sqrt{\frac{Y}{\rho}}$$

Forced vibration and resonance

- Sound propagates in a medium due to forced vibrations of the particles of the surrounding medium by the vibrating object generating sound. In the steady-state, the frequency of oscillations of a forced oscillator is equal to the frequency of the driving force. The steady-state displacement of the forced oscillator is given as

$$x(t) = a(\omega) \cos(\omega t - \theta)$$

where the amplitude $a(\omega)$ is given as

$$a(\omega) = \frac{F_0}{m[(\omega_0^2 - \omega^2)^2 + 4b^2\omega^2]^{1/2}}$$

When the frequency of the driving force is equal to the natural frequency of the driven system, the amplitude of oscillations increases manifold and this phenomenon is called **resonance**.

Intensity of sound

- Intensity I of sound is defined as the rate of energy transfer by the sound wave per unit area normal to the direction of wave propagation:

$$I = \frac{P}{A} = \frac{P}{4\pi r^2}$$

$$\text{and } I \propto \frac{1}{r^2} \quad \text{and} \quad a \propto \frac{1}{r}$$

where a is amplitude and r is the distance from the source of sound.

Intensity level

- The minimum intensity of the sound wave which a human being can hear is called Threshold of Hearing (TOH). Its value is 10^{-12} Wm^{-2} .

Intensity of sound is measured/expressed using a logarithmic scale called decibel (dB). It is a relative scale in which intensity of sound is expressed relative to the threshold of hearing I_0 . The intensity level, β of sound of intensity I is given as

$$\beta = (10 \text{ dB}) \log_{10} \left(\frac{I}{I_0} \right)$$

Musical notes

- From the perception point of view, a musical note is characterised by pitch, loudness and quality which, respectively, depend on the frequency, intensity and waveform of the sound wave.

Musical scale

- Musical scales are based on the concept of octave. An octave is defined as a set of two frequencies wherein the value of higher frequency is two times the value of the lower frequency.

The diatonic musical scale consists of a series of eight notes with the frequency of the last and the first note in the ratio 2:1.

The frequency intervals between the notes of the musical scale are not uniform. However, the sensation of pitch produced by these notes is of equal interval.

The ratio between two successive notes of a musical scale is called **musical interval**.

In the **tempered scale**, the octave is divided into 12 equal musical intervals. The musical interval between two consecutive notes of tempered scale is 1.05946.

Acoustics of building

- In a closed space like room or auditorium, the sensation of hearing is affected by phenomena such as reverberation, echo, focussing, echelon effect, loudness, resonance and noise. The process of minimising the effect of these factors is studied under acoustics of buildings.

Reverberation refers to persistence of sound even after the production of sound has stopped. A reasonable reverberation time is desirable for appreciating the richness of musical sound.

3.7 TERMINAL QUESTIONS

1. An object vibrating with frequency of 20 Hz produces 1-D sound waves in air propagating in positive x-direction. The maximum displacement of air particles from their respective mean positions is 0.03 m and the distance between two consecutive compressions is 0.08 m. Write the mathematical expression representing this sound wave.
2. Determine the frequencies of the fundamental mode and the next two harmonics that can be set up on a sitar string of length 1.2 m. Take the speed of waves on the string to be $3.8 \times 10^3 \text{ ms}^{-1}$.
3. The intensity level of sound at a place is 40 dB. Calculate the intensity of sound at that place.
4. Determine the sound intensity level if the intensity changes by a factor of a) 4, b) 10, c) 10^4 .
5. The frequency of the keynote (or the tonic) in a diatonic musical scale is 264 Hz. Calculate the frequencies of the various notes of the scale.

3.8 SOLUTIONS AND ANSWERS

Self-Assessment Questions

1. a) True, b) False, c) False, d) True
2. The equation for 1-D sound waves travelling in the negative x-direction is $\psi(x, t) = a \sin(kx + \omega t)$

As per the problem, $a = 0.5 \text{ cm}$; $T = 0.4 \text{ s}$ and $\lambda = 32.0 \text{ cm}$.

$$\text{So, } k = \frac{2\pi}{\lambda} = \frac{2\pi}{32.0 \text{ cm}} = 0.2 \text{ cm}^{-1}$$

$$\text{and } \omega = 2\pi f = \frac{2\pi}{T} = \frac{2\pi}{0.4 \text{ s}} = 15.7 \text{ s}^{-1}$$

Thus, the desired equation of the wave is

$$\psi(x, t) = 0.5 \sin(0.2x + 15.7t) \text{ cm}$$

3. The speed of sound in solid is given as

$$v = \sqrt{Y/\rho}$$

As per problem, $v = 5 \times 10^3 \text{ ms}^{-1}$ and $\rho = 1500 \text{ kg m}^{-3}$. So,

$$5 \times 10^3 \text{ ms}^{-1} = \sqrt{\frac{Y}{1500 \text{ kg m}^{-3}}}$$

$$\text{or } Y = (5 \times 10^3 \text{ ms}^{-1}) \times (1500 \text{ kg m}^{-3}) = 3.8 \times 10^{10} \text{ Nm}^{-2}$$

4. The fundamental frequency of a string, such as string of a guitar, fixed at both ends is given as

$$f_1 = \frac{v}{2L}$$

We have, $L = 2.0 \text{ m}$ and $v = 400 \text{ ms}^{-1}$. So, $f_1 = \frac{400 \text{ ms}^{-1}}{2 \times (2.0 \text{ m})} = 100 \text{ Hz}$.

Frequency of second overtone, $f_3 = 3 \times f_1 = 300 \text{ Hz}$

5. The intensity level, β in decibel is given as [Eq. (3.22)]:

$$\beta = (10 \text{ dB}) \log_{10}(I/I_0)$$

Since, $\beta = 80 \text{ dB}$, we have

$$\log_{10}(I/I_0) = 8.0$$

$$\text{or } I = I_0 \times (10)^8 = (10^{-12} \text{ Wm}^{-2}) \times (1.0 \times 10^8) = 1.0 \times 10^{-4} \text{ Wm}^{-2}.$$

Terminal Questions

1. We know that 1-D sound wave propagating in the positive x-direction is given as

$$\psi(x, t) = a \sin(kx - \omega t)$$

So, to write the mathematical expression for the wave, we need to determine amplitude, a , wave number k and angular frequency ω . As per the problem, the maximum displacement of the particles of the medium is 0.03 m . So, amplitude, $a = 0.03 \text{ m}$.

The distance between two compressions is equal to the wavelength of the wave. So,

$$\lambda = 0.08 \text{ m} \Rightarrow k = \frac{2\pi}{\lambda} = \frac{2 \times 3.14}{0.08 \text{ m}^{-1}} = 78.5 \text{ m}^{-1}$$

Frequency of the wave is equal to the frequency of the vibrating object producing the wave. So,

$$f = 20 \text{ s}^{-1} \Rightarrow \omega = 2\pi f = 2 \times (3.14) \times (20 \text{ s}^{-1}) = 125.6 \text{ s}^{-1}$$

So, the mathematical expression of the sound wave is

$$\psi(x, t) = (0.03) \sin[(78.5)x - (125.6)t] \text{ m}$$

2. We know that the normal mode frequencies of the string fixed at both ends is given as

$$f_n = \frac{nv}{2L}$$

We have, $L = 1.2 \text{ m}$ and $v = 3.9 \times 10^3 \text{ ms}^{-1}$. So, the frequency of the fundamental mode ($n = 1$) is

$$f_1 = \frac{3.9 \times 10^3 \text{ ms}^{-1}}{2 \times 1.2 \text{ m}} = 1625 \text{ s}^{-1}$$

Frequencies f_2 and f_3 of the next two harmonics, namely second and third harmonics are

$$f_2 = 2f_1 = 2 \times 1625 \text{ s}^{-1} = 3250 \text{ s}^{-1}$$

$$f_3 = 3f_1 = 3 \times 1625 \text{ s}^{-1} = 4875 \text{ s}^{-1}$$

3. We know that the intensity level, β of sound is given as

$$\beta = (10 \text{ dB}) \log_{10}(I/I_0)$$

The threshold of hearing, I_0 is 10^{-12} Wm^{-2} and $\beta = 40 \text{ dB}$. So,

$$40 \text{ dB} = (10 \text{ dB}) \log_{10}(I/I_0)$$

$$\frac{I}{I_0} = 10^4 \Rightarrow I = (10^{-12} \text{ Wm}^{-2}) \times (10^4) = 10^{-8} \text{ Wm}^{-2}.$$

4. Let the initial intensity level of sound is

$$\beta_1 = (10 \text{ dB}) \log_{10}(I/I_0)$$

where I is the intensity of sound.

- a) When the intensity is changed by factor of 4, the intensity level is

$$\beta_2 = (10 \text{ dB}) \log_{10}(4I/I_0)$$

So, change in intensity level is

$$\begin{aligned} \beta_2 - \beta_1 &= (10 \text{ dB}) \left[\log \frac{4I}{I_0} - \log \frac{I}{I_0} \right] \\ &= (10 \text{ dB}) \log \left[\frac{(4I/I_0)}{(I/I_0)} \right] = (10 \text{ dB}) \log(4) = 6 \text{ dB} \end{aligned}$$

- b) When the intensity is increased by a factor of 10, the intensity level is

$$\beta_3 = (10 \text{ dB}) \log_{10}(10I/I_0)$$

So, change in intensity level

$$\beta_3 - \beta_1 = (10 \text{ dB}) \log_{10} 10 = 10 \text{ dB}$$

- c) When the intensity is increased by a factor of 10^4 , the intensity level is

$$\beta_4 = (10 \text{ dB}) \log_{10} \left(\frac{I \times 10^4}{I_0} \right)$$

So, change in intensity level

$$\beta_4 - \beta_1 = (10 \text{ dB}) \times 4 \times \log(10) = 40 \text{ dB}$$

5. The frequency of various notes in the diatonic scale having keynote 264 Hz is calculated using the frequency ratios (Table 3.2):

Note	Keynote (sa)	re	ga	ma
Frequency (in Hz)	264 Hz	264×1.125 = 297	264×1.250 = 330	264×1.333 = 352

pa	dha	ni	sa
264×1.500 = 396	264×1.667 = 440	264×1.875 = 495	264×2 = 528



UNIT 4

ELECTROMAGNETIC WAVE

Electromagnetic waves are present everywhere in nature and used extensively in all modern technology like FM radio, radars, wireless communication and microwave ovens.

(Source of picture: wikimedia.org)

Structure

- | | | | |
|-----|--|-----|-----------------------|
| 4.1 | Introduction | 4.4 | Summary |
| | Expected Learning Outcomes | 4.5 | Terminal Questions |
| 4.2 | Wave Equation for Electromagnetic Waves | 4.6 | Solutions and Answers |
| | Transverse Nature of Electromagnetic Waves | | |
| 4.3 | Huygens Principle | | |
| | Wavefronts and their Properties | | |

STUDY GUIDE

In this unit, you will study about electromagnetic waves. The study of electromagnetic waves in a course primarily on optics should not surprise you because light is an electromagnetic wave. You have learnt about electromagnetic waves in Unit 16 and 17 of second semester course BPHCT-133 entitled Electricity and Magnetism. So, we advise you to revise these units. Like mechanical waves, electromagnetic waves are described by the wave equation, which is obtained on the basis of Maxwell's equations. Since you will use vector quantities such as electric field $\vec{E}$ and magnetic field $\vec{B}$ in this unit, you should also revise vector algebra and vector calculus from Block 1 of BPHCT-133. While studying this unit, you should focus on how it was predicted that light is an electromagnetic wave, what are the differences and similarities between electromagnetic waves and mechanical waves, and how the concepts of wavefronts and rays help us understand the propagation of light. You should work through the derivations while studying the unit to master the mathematical treatment. Also, you should try to solve SAQs and TQs before looking at their solutions at the end of the unit to check your understanding of the concepts discussed in the unit.

“Maxwell's equations have had a greater impact on human history than any ten presidents.”

Carl Sagan

4.1 INTRODUCTION

In Units 1 and 2 of this course, you have learnt about basic features of wave motion and superposition of waves. You now know that the waves are characterised by physical quantities like amplitude, frequency and wavelength. You have also learnt that the wave equation describes the propagation of a wave. For example, the wave equation for transverse wave on a string describes the propagation of 'disturbance' (displacement of the elements of the string) along the string. In this unit, you will learn how to apply concepts related to wave motion to study electromagnetic waves.

Modern scientific understanding about the nature of light was developed in the process of answering the question: What is light? Is light made up of particles or there are light waves? In the nineteenth century, both the particle (corpuscle) and wave theories of light were being put forward by scientists. Though the corpuscular theory, supported by Newton, could explain phenomena like rectilinear propagation of light and formation of shadows, it failed to explain observations related to interference, diffraction and polarisation of light. These phenomena could be explained on the basis of the wave model of light proposed by Huygens. However, the real breakthrough in understanding the nature of light came from the work of Maxwell, which put the wave theory of light on a sound theoretical pedestal. According to the electromagnetic theory proposed by Maxwell, light is an electromagnetic wave. In this unit, you will learn what an electromagnetic wave is, what prompted Maxwell to predict that light is an electromagnetic wave and what are the differences and similarities between electromagnetic waves and mechanical waves.

In Sec. 4.2, you will learn how to derive wave equation for electric field and magnetic field from Maxwell's equations, which describe electric and magnetic fields. These wave equations imply that electric field and magnetic field propagate in space as waves. In this section, you will learn that the speed with which the electromagnetic wave travels in free space is equal to the speed of light. This prompted Maxwell to predict that light is an electromagnetic wave. You will learn about some of the properties of electromagnetic waves which are similar to experimentally observed properties of light. In Sec. 4.3, you will learn Huygens principle which gives a satisfactory explanation of the optical phenomena like interference, diffraction and polarisation. You will discover that the concept of wavefront is at the core of this principle.

In the next unit, you will learn about polarisation of light which provides experimental evidence for the transverse nature of light waves.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ derive the wave equation for electric field and magnetic field from Maxwell's equations;
- ❖ show that electromagnetic waves are transverse waves;
- ❖ explain the concept of wavefronts; and
- ❖ state Huygens principle and use it to discuss propagation of light.

4.2 WAVE EQUATION FOR ELECTROMAGNETIC WAVES

In Unit 1, you have learnt that in mechanical waves such as transverse waves on a string or longitudinal sound waves in air, the 'disturbance' in the medium propagates like a wave. In case of the transverse waves on a string, the 'disturbance' in the medium is the displacement of the elements of the string and in case of longitudinal sound waves in air, it is the displacement of the air particles. These disturbances are caused due to periodic forced oscillations of the particles of the medium. You have also studied that the wave equation governs the propagation of waves. The wave equation relates the speed of propagation of the waves with the physical properties of the medium in which they propagate. For example, the speed of transverse waves on a string depends on the tension in the string and mass per unit length of the string.

The electromagnetic waves are different from mechanical waves because these originate from oscillating charges. In this section, we seek answers of such questions regarding electromagnetic waves as: Which physical quantity is varying in the propagation of electromagnetic wave? What is the wave equation for electromagnetic waves which governs its propagation? Does the wave equation enable us to determine the speed of electromagnetic waves? Let us find the answers to these questions.

You have learnt Maxwell's equations in Unit 16 of second semester course entitled Electricity and Magnetism (BPHCT-133). Recall that Maxwell's equations in differential form are given as

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \quad (\text{Gauss's law for electrostatics}) \quad (4.1)$$

$$\vec{\nabla} \cdot \vec{B} = 0 \quad (\text{Gauss's law for magnetism}) \quad (4.2)$$

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \quad (\text{Faraday's law for electromagnetic induction}) \quad (4.3)$$

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{J} + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \quad (\text{Modified Ampere's law}) \quad (4.4)$$

where $\vec{E}$, $\vec{B}$, ρ and $\vec{J}$ denote the electric field, magnetic field, free charge density and the conduction current density, respectively.

From Eqs. (4.3) and (4.4), it is evident that time-varying magnetic field cannot exist without a variable electric field and a time-varying electric field cannot exist without a variable magnetic field. **In other words, Maxwell's equations unified electric and magnetic phenomena.** In addition, as you will learn in this section, Maxwell's equations predict (i) the existence of electromagnetic waves and (ii) light is an electromagnetic wave. Thus, Maxwell's equations also unified optical phenomenon with electric and magnetic phenomena.

Let us now derive wave equation for electromagnetic waves from Maxwell's equations. For keeping the mathematical treatment simple, we consider Maxwell's equations in vacuum (charge free and current free space) so that $\rho = 0$ and $\vec{J} = 0$. Then, Eqs. (4.1 to 4.4) reduce to

From Unit 16 of the course BPHCT-133, you will recall that Maxwell's equation is a set of equations representing the fundamental laws of electricity and magnetism.

Gauss's law for electricity [Eq. (4.1)] gives electric field due to electric charge. **Gauss's law for magnetism** [Eq. (4.2)] says that magnetic flux through a closed surface is zero; that is, unlike electric charge, magnetic charge (or monopoles) does not exist. **Faraday's law** [Eq. (4.3)] says that a changing magnetic field is accompanied by an electric field. **Ampere's law modified by Maxwell** [Eq. (4.4)] by adding the second term on the RHS says that changing electric field is accompanied by a magnetic field. As such, Ampere's law, without the modification introduced by Maxwell, gives magnetic field due to steady current (having current density, $\vec{J}$). **This set of four equations, called Maxwell's equations, describe electric and magnetic phenomena.**

$$\vec{\nabla} \cdot \vec{E} = 0 \quad (4.5)$$

$$\vec{\nabla} \cdot \vec{B} = 0 \quad (4.6)$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (4.7)$$

$$\vec{\nabla} \times \vec{B} = \epsilon_0 \mu_0 \frac{\partial \vec{E}}{\partial t} \quad (4.8)$$

where μ_0 and ϵ_0 are the magnetic permeability and permittivity of free space. Note that Eqs. (4.7) and (4.8) are coupled and first order partial differential equations. These equations can be uncoupled with the help of Eqs. (4.5) and (4.6) and wave equations of $\vec{E}$ and $\vec{B}$ fields can be obtained.

Now, to obtain the wave equation for electric field $\vec{E}$ in vacuum, we first take the curl of Eq. (4.7):

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = \vec{\nabla} \times \left(-\frac{\partial \vec{B}}{\partial t} \right) = -\frac{\partial}{\partial t} (\vec{\nabla} \times \vec{B}) \quad (4.9)$$

To simplify the left hand side of Eq. (4.9), we use the following vector identity for any vector field $\vec{A}$:

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \vec{\nabla} (\vec{\nabla} \cdot \vec{A}) - \nabla^2 \vec{A} \quad (4.10)$$

Using the above identity, we can write the LHS of Eq. (4.9) as

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = \vec{\nabla} (\vec{\nabla} \cdot \vec{E}) - \nabla^2 \vec{E}$$

From Eq. (4.5), we have $\vec{\nabla} \cdot \vec{E} = 0$. Thus, we find that Eq. (4.9) reduces to

$$-\nabla^2 \vec{E} = -\frac{\partial}{\partial t} (\vec{\nabla} \times \vec{B})$$

On substituting the value of $\vec{\nabla} \times \vec{B}$ from Eq. (4.8) in the above expression, we get

$$\nabla^2 \vec{E} = \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} \quad (4.11)$$

Eq. (4.11) is a second order partial differential equation for the time varying electric field $\vec{E}$.

Let us now compare Eq. (4.11) with the wave equation for mechanical wave given as [Eq. (1.12), Unit 1]

$$\frac{\partial^2 y}{\partial x^2} = \left(\frac{1}{v^2} \right) \frac{\partial^2 y}{\partial t^2} \quad (4.12)$$

We find that these two equations are similar: you will get Eq. (4.11) if you replace $y(x,t)$ by $\vec{E}$ and $(1/v^2)$ by $(\mu_0 \epsilon_0)$ in Eq. (4.12). So, we can say that Eq. (4.11) is the wave equation for $\vec{E}$ field and it predicts that $\vec{E}$ field propagates like wave in space. You may ask: How this time varying $\vec{E}$ field is generated at the first place? We will address this question shortly.

Maxwell's equations also lead to a wave equation for magnetic field $\vec{B}$. To find the wave equation for $\vec{B}$, let us take the curl of Eq. (4.8):

$$\vec{\nabla} \times \vec{\nabla} \times \vec{B} = \epsilon_0 \mu_0 \frac{\partial}{\partial t} (\vec{\nabla} \times \vec{E}) \quad (4.13)$$

The 3-D wave equation has the form

$$\nabla^2 \psi = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2}$$

where $\psi(x, y, z, t)$ is the physical quantity which propagates wavelike with speed v .

Again, using the vector identity given by Eq. (4.10), we can write LHS of Eq. (4.13) as

$$\vec{\nabla} \times \vec{\nabla} \times \vec{B} = \vec{\nabla}(\vec{\nabla} \cdot \vec{B}) - \nabla^2 \vec{B}$$

Since Eq. (4.6) tells us that $\vec{\nabla} \cdot \vec{B} = 0$, we have

$$\vec{\nabla} \times \vec{\nabla} \times \vec{B} = -\nabla^2 \vec{B}$$

Substituting this value of LHS in Eq. (4.13), we get

$$-\nabla^2 \vec{B} = \epsilon_0 \mu_0 \frac{\partial}{\partial t} (\vec{\nabla} \times \vec{E})$$

Substituting the value of $(\vec{\nabla} \times \vec{E})$ from Eq. (4.7) in above equation, we get

$$\nabla^2 \vec{B} = \mu_0 \epsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2} \quad (4.14)$$

So, we find that Maxwell's equations in vacuum (charge free and current free space) give wave equations for time varying electric and magnetic fields. This implies that time varying electric and magnetic fields propagate like waves in space.

At this stage, you may ask: What is the physical process involved in the generation and propagation of electromagnetic waves? For a qualitative understanding of the underlying process, note that one of the Maxwell's equation [Eq. (4.3)] tells us that a time-varying $\vec{B}$ field generates a $\vec{E}$ field. Since Eq. (4.3) involves a vector cross-product, the generated $\vec{E}$ field is perpendicular to the direction in which $\vec{B}$ field changes. Similarly, another Maxwell's equation [Eq. (4.4)] tells us that a time-varying $\vec{E}$ field generates a $\vec{B}$ field. Again, since Eq. (4.4) involves a vector cross-product, the generated $\vec{B}$ field will be perpendicular to the direction in which $\vec{E}$ field changes.

With the above understanding about the relation between time varying electric and magnetic fields, let us consider an electric charge. When the charge is at rest, it has a constant $\vec{E}$ field extending radially in all directions. If the charge is now somehow made to undergo accelerated motion (such as oscillatory motion), the $\vec{E}$ field of the charge changes. The time varying $\vec{E}$ field set up by the oscillating charge is experienced by a test charge placed at some distance from it. So, the time varying $\vec{E}$ field propagates in space with some finite speed and also generates a time-varying $\vec{B}$ field represented by Eq. (4.4). This time-varying $\vec{B}$ field, in turn, generates a time-varying $\vec{E}$ field [Eq. (4.3)]. This process continues; one changing field generates the other which is also changing. These coupled $\vec{E}$ and $\vec{B}$ fields were conceptualised by Maxwell as a single entity called electromagnetic field which propagates in space like a wave. The electromagnetic field, once generated, is self-sustaining because time-varying $\vec{E}$ and $\vec{B}$ fields could continuously generate each other and propagate in space. So, wave equations given by Eqs. (4.11) and (4.14) for $\vec{E}$ field and $\vec{B}$ field, respectively, refer to the electric and magnetic components of the single entity called electromagnetic field which propagates in space as electromagnetic wave.

As you have learnt in Unit 1, the wave equation enables us to determine the speed of the wave. So, you may like to know the speed with which

Wave equation is a relation between space and time variations of some physical quantity. A physical quantity, for which a wave equation exists, moves like a wave. In case of one-dimensional wave on a string, the physical quantity is the displacement of the elements of the string and in case of electromagnetic waves, it is electric field $\vec{E}$ and magnetic field $\vec{B}$.

electromagnetic wave propagates in space. If we compare Eqs. (4.11) or (4.14) with Eq. (4.12), we find that the speed of electromagnetic wave is given as

$$v = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \quad (4.15)$$

Eq. (4.15) shows that the speed of an electromagnetic wave depends only on μ_0 and ϵ_0 which are characteristics of the medium. This suggests that all electromagnetic waves should, irrespective of their frequencies or amplitudes propagate with this speed in **free space**. We can easily calculate the magnitude of v by substituting the values of $\epsilon_0 (= 8.85 \times 10^{-12} \text{C}^2 \text{N}^{-1} \text{m}^{-2})$ and $\mu_0 (= 4\pi \times 10^{-7} \text{Ns}^2 \text{C}^{-2})$ in Eq. (4.15):

$$\begin{aligned} v &= \frac{1}{[(8.85 \times 10^{-12} \text{C}^2 \text{N}^{-1} \text{m}^{-2}) \times (4 \times 10^{-7} \text{Ns}^2 \text{C}^{-2})]^{1/2}} \\ &= 2.99794 \times 10^8 \text{ms}^{-1} \end{aligned}$$

This is precisely the speed of electromagnetic waves in vacuum. It is worthwhile to mention here that using the then best known value of ϵ_0 , Maxwell found that electromagnetic waves should travel at a speed of $3.1074 \times 10^8 \text{ms}^{-1}$. This was very close to the speed of light measured by Fizeau ($3.14858 \times 10^8 \text{ms}^{-1}$). **Based on these results, Maxwell proposed that light is an electromagnetic wave and electromagnetic theory of light was born.** In his own words:

This velocity is so nearly that of light, that it seems we have strong reason to believe that light itself is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws.

We cannot help but wonder at such pure gold having come out of his researches on electric and magnetic phenomena. It was a classic example of the unification of knowledge towards which science is ever striving. **With this one calculation, Maxwell brought the entire science of optics under the umbrella of electromagnetism.** Its significance is profound because it identifies light with structures consisting of electric and magnetic fields travelling freely through free space.

Also note that the speed of electromagnetic waves is determined by two constants namely ϵ_0 (permittivity) and μ_0 (permeability) which appear in Coulomb's law and Biot-Savart law. These laws, and these constants have no relation whatsoever with light. But, according to Maxwell's electromagnetic theory, these are related to the speed of light.

Heinrich Hertz conducted a series of experiments to produce and detect electromagnetic waves at a considerable distance. He used an alternating current (ac) source to drive an electric-dipole antenna as transmitting antenna. The frequency of the transmitted electromagnetic wave was the same as the frequency of the ac source. The receiving antenna was connected to a circuit tuned to the same frequency. By measuring the wavelength and frequency of electromagnetic waves, Hertz calculated their speed. He found it to be precisely equal to the speed of light. He also demonstrated that electromagnetic waves also exhibit properties like reflection, refraction,

As you know, the normal path of the development of the science is from phenomena or experimental observation to theory. But, in case of electromagnetic waves, it was the other way round. The existence of electromagnetic waves was demonstrated experimentally by Heinrich Hertz much after its theoretical prediction by Maxwell.

polarisation, etc. and hence provided conclusive experimental evidence that **light is an electromagnetic wave**.

So far, you have learnt to obtain wave equations for $\vec{E}$ and $\vec{B}$ fields using Maxwell's equations and discovered that light is an electromagnetic wave because their speeds in vacuum are same. You have also learnt about the experiments conducted by Hertz to detect electromagnetic waves. The polarisation of light is one phenomenon which shows that light is a transverse wave. Thus, if light is indeed an electromagnetic wave, the electromagnetic wave must be a transverse wave. Is it so? Let us find out.

4.2.1 Transverse Nature of Electromagnetic Waves

In Unit 1, you learnt that only those mathematical functions represent a travelling wave which satisfies the corresponding wave equation. For example, the expression for displacement, $y(x, t) = a \sin(kx - \omega t)$ represents a 1-D wave travelling in the positive x-direction because it satisfies the 1-D wave equation given by Eq. (4.12). This requirement is satisfied for electromagnetic waves also by the mathematical function representing $\vec{E}$ and $\vec{B}$ fields.

But, in case of electromagnetic waves, an additional condition is imposed on any function representing the $\vec{E}$ and $\vec{B}$ fields. Thus, the functions representing $\vec{E}$ and $\vec{B}$ fields must satisfy wave equations [Eqs. (4.11) and (4.14)] as well as Maxwell's equations [Eqs. (4.5) to (4.8)]. This means that in solving the wave equation for electromagnetic waves, we must ensure that the solutions also satisfy Maxwell's equation. Only then they would represent an electromagnetic wave.

The above conditions on the electric and magnetic fields enable us to explain the transverse nature of electromagnetic waves. Let us see how.

Let the solutions of wave equations [Eqs. (4.11) and (4.14)] be:

$$\vec{E} = \vec{E}_0 \sin(kz - \omega t) \quad (4.16)$$

$$\vec{B} = \vec{B}_0 \sin(kz - \omega t) \quad (4.17)$$

Eqs. (4.16) and (4.17) represent the electric and magnetic fields associated with electromagnetic waves travelling in the positive z-direction with wave number k and angular frequency ω . $\vec{E}_0$ and $\vec{B}_0$ are the amplitudes of electric and magnetic fields, respectively. The wave number, k is equal to $2\pi/\lambda$, where λ is the wavelength of the electromagnetic wave and the angular frequency, ω is equal to $2\pi f$ where f is the frequency of the wave.

Before proceeding further, study the following example to get an idea about the wave parameters of an electromagnetic wave.

Note that Eqs. (4.16) and (4.17) give $\vec{E}$ and $\vec{B}$ fields as sinusoidally varying functions of space and time. These functions are similar to Eq. (1.15) of Unit 1 which represents a transverse wave on a string.

EXAMPLE 4.1: PARAMETERS OF ELECTROMAGNETIC WAVES

The electric field associated with an electromagnetic wave propagating in vacuum is given by:

$$\vec{E} = \{ (28 \text{ NC}^{-1}) \sin[(1.8 \text{ m}^{-1})z + (5.4 \times 10^8 \text{ s}^{-1})t] \} \hat{y}$$

Determine wavelength and frequency of the wave and its direction of propagation.

SOLUTION ■ From the given expression for $\vec{E}$, we note that the wave number, k of the wave is

$$k = 1.8 \text{ m}^{-1}$$

$$\text{So, } \frac{2\pi}{\lambda} = 1.8 \text{ m}^{-1} \Rightarrow \lambda = \frac{2 \times 3.14}{1.8 \text{ m}^{-1}} = 3.48 \text{ m}$$

The angular frequency, ω of the wave is

$$\omega = 5.4 \times 10^8 \text{ s}^{-1}$$

$$\text{or } 2\pi f = 5.4 \times 10^8 \text{ s}^{-1} \Rightarrow f = \frac{5.4 \times 10^8 \text{ s}^{-1}}{2 \times 3.14} = 8.59 \times 10^7 \text{ s}^{-1}$$

The wave is propagating along the negative z -direction.

Now, you may like to answer an SAQ.

SAQ 1 – Calculating the parameters of electromagnetic wave

A harmonic electromagnetic wave of wavelength 450 nm is propagating in positive z -direction in vacuum. What is the frequency of the wave? Also calculate the angular frequency ω and wave number k of the wave.

Now, to understand the transverse nature of electromagnetic wave, recall from Unit 1 that in a transverse wave, the displacement of the particles of the medium is perpendicular to the direction of propagation of the wave. Thus, if electromagnetic wave is to be transverse in nature, the $\vec{E}$ and $\vec{B}$ field variations must be perpendicular to the direction of propagation (z -direction) of the wave. So, if we show now that this is indeed the case, we can conclude that the electromagnetic waves are transverse in nature.

To that end, let us first examine whether Eqs. (4.16) and (4.17) satisfy the following two Maxwell's equations [Eqs. (4.5) and (4.6)] in vacuum:

$$\vec{\nabla} \cdot \vec{E} = 0 \text{ and } \vec{\nabla} \cdot \vec{B} = 0$$

If we substitute Eq. (4.16) in the LHS of the above equations for $\vec{E}$ field, we get

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \frac{\partial}{\partial x}(E_{0x}) \sin(kz - \omega t) + \frac{\partial}{\partial y}(E_{0y}) \sin(kz - \omega t) \\ &\quad + \frac{\partial}{\partial z}(E_{0z}) \sin(kz - \omega t) \end{aligned}$$

If the equation $\vec{\nabla} \cdot \vec{E} = 0$ is to be satisfied by $\vec{E}$ field given by Eq. (4.16), all partial derivatives in the above expression should be zero. But, the partial derivative with respect to z will be zero if and only if E_{0z} is zero. It is so because the partial derivative of the sine function with respect to z is non-zero. Thus, we find that $\vec{E}$ given by Eq. (4.16) can represent the electric field associated with electromagnetic wave if and only if

$$E_{0z} = (\vec{E}_0)_z = 0 \quad (4.18)$$

Similarly, you can show that $\vec{B}$ given by Eq. (4.17) can represent the magnetic field associated with the electromagnetic wave if and only if

$$B_{0z} = (\vec{B}_0)_z = 0 \quad (4.19)$$

SAQ 2 – Magnetic field along the direction of electromagnetic wave propagation

Show that the component of the magnetic field associated with an electromagnetic wave along the direction of propagation of the wave is zero.

So, Eqs. (4.18) and (4.19) show that the components of the electric field and magnetic field, associated with the electromagnetic wave, along the direction of propagation of the wave are zero. **This implies that the direction of propagation of the electromagnetic wave is perpendicular to directions of the electric and magnetic fields that constitute it.** Thus, we conclude that electromagnetic waves are transverse in nature.

At this stage, you may ask: Do the solutions of wave equations given by Eqs. (4.16) and (4.17) satisfy the remaining two Maxwell's equations [Eqs. (4.7) and (4.8)]? To get an answer to this question, we need to substitute Eqs. (4.16) and (4.17) in Eqs. (4.7) and (4.8), respectively. When we do that [refer to Sec. 17.2, Unit 17, BPHCT-133], we find that $\vec{E}$ and $\vec{B}$ fields will satisfy Maxwell's equation if and only if these are related to each other as under:

$$\vec{B} = \frac{k}{\omega} (\hat{k} \times \vec{E})$$

or $\hat{k} \times \vec{E} = c \vec{B}$

because the velocity of light, $c = \frac{\omega}{k}$.

So, we can write the relation between the magnitudes of $\vec{E}$ and $\vec{B}$ as

$$\frac{|\vec{E}|}{|\vec{B}|} = c = \frac{\omega}{k}$$

or $E_0 = \frac{\omega}{k} B_0$ (4.20)

Further, owing to the transverse nature of the electromagnetic wave, we can write the electric and magnetic fields associated with an electromagnetic wave propagating in z-direction as

$$\vec{E} = \hat{y} E_0 \sin(kz - \omega t) \quad (4.21)$$

$$\vec{B} = \hat{x} B_0 \sin(kz - \omega t) \quad (4.22)$$

where E_0 and B_0 are constants. Note that we have taken the electric and magnetic fields along y and x directions, respectively because the electromagnetic wave is propagating along z-direction and it is a transverse wave.

The relation between E_0 and B_0 given by Eq. (4.20) signifies the following:

- Recall from Unit 1 that (ω/k) is equal to speed of the wave; for electromagnetic wave speed is c , the speed of light.
- Since the magnitudes of $\vec{E}$ and $\vec{B}$ field vectors differ only by a real number, c , the electric and magnetic field vectors should be in-phase. This

Eq. (4.20) indicates that the magnitude of electric field exceeds that of magnetic field in electromagnetic waves. (However, this situation prevails for non-relativistic motion of the charge.) So, for most **terrestrial situations, electromagnetic waves are essentially electrical phenomenon in so far as their interaction with matter is concerned.**

means that, if $\vec{E}$ becomes zero (or maximum) at some instant, $\vec{B}$ must also necessarily be zero (or maximum) at that instant, and so on. This also shows that neither electric nor magnetic wave can exist without the other. An electric field varying in time sets up a space-time varying magnetic field, which, in turn produces an electric field varying in space and time, and so on. We cannot separate them. This mutually supporting role results in the generation of electromagnetic waves. The pictorial representation of fields of an electromagnetic wave (propagating along the positive z -direction) is shown in Fig. 4.1. Note that electric and magnetic fields are oriented at right angles to one another and to the direction of wave motion.

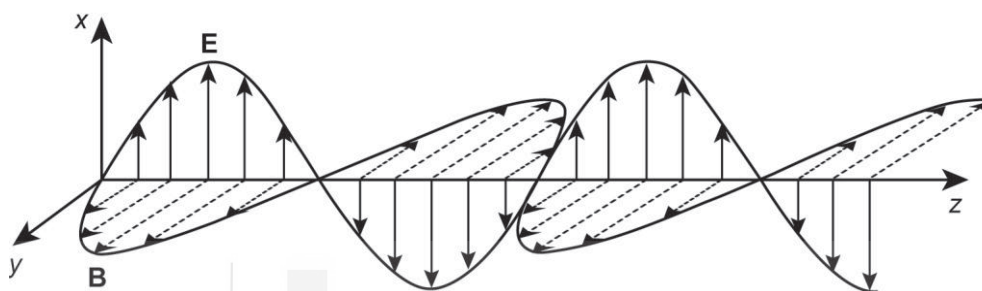


Fig. 4.1: The electric and magnetic fields associated with a plane electromagnetic wave.

Before proceeding further, answer an SAQ.

SAQ 3 – Nature of electromagnetic wave

- Compare sound waves and light waves and list two differences and two similarities between them.
- If the amplitude of the electric field associated with an electromagnetic wave is doubled, by what factor will the amplitude of the associated magnetic field change?

So far, you have studied how Maxwell's equations predict the existence of electromagnetic waves, how he concluded that light was an electromagnetic wave. You have also learnt about the transverse nature of electromagnetic wave. You now know that light is an electromagnetic wave. Though the origin of electromagnetic waves is different from those of mechanical waves such as transverse waves on a string and sound waves in air, it has all the characteristics of a wave. The electromagnetic waves are characterised by wave parameters such as amplitude, wavelength, frequency and phase. These waves obey the superposition principle and transport energy (electromagnetic energy) in space. Maxwell's electromagnetic theory gives a satisfactory explanation of most of the optical phenomena such as interference, diffraction and polarisation. You will learn Maxwell's electromagnetic theory in further detail if you pursue higher studies in physics.

However, much before Maxwell proposed the electromagnetic theory of light, successful attempts were made by physicists like Huygens and Fresnel to develop a wave model of light. In this course, to keep the discussion simpler, we shall use the wave model of light based on Huygens principle to understand the phenomena of interference and diffraction.

Huygens was the most notable physicist who contributed to the development of wave model using the concepts of secondary wavelets and wavefront. You will now learn Huygens principle. But, before that, let us recall the important points of this section.

Recap

- The wave equation for the $\vec{E}$ and $\vec{B}$ field vectors are

$$\nabla^2 \vec{E} = \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}$$

$$\nabla^2 \vec{B} = \mu_0 \epsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2}$$

- The speed of electromagnetic wave in vacuum is equal to the speed of light. This fact prompted Maxwell to predict that light is an electromagnetic wave.
- The electromagnetic waves are transverse in nature.
- The magnitude of the $\vec{E}$ and $\vec{B}$ fields associated with electromagnetic wave in free space is related to each other as

$$E_0 = cB_0 = \frac{\omega}{k} B_0$$

4.3 HUYGENS PRINCIPLE

Huygens principle provides a basis to understand the propagation of waves. It applies to mechanical as well as electromagnetic waves. **According to this principle,**

every point on a wavefront behaves like a source of spherical secondary wavelets such that the wavefront at some later time is the envelop of these wavelets.

So, there are two central concepts in this principle: **wavefront** and **secondary wavelets**.

To understand Huygens principle and its application to the propagation of waves, you first need to know what a wavefront and secondary wavelet are.

4.3.1 Wavefronts and their Properties

To understand what a wavefront is, consider waves generated on a water surface when we drop a pebble (Fig. 4.2). The disturbance created by the pebble on the water surface propagates as wave. The wave on the water surface is a two-dimensional wave and it propagates in the form of concentric circles having the centre at the point where the dropped pebble touches the water surface. All the water particles on a given circle are in the same phase because they are all at the same distance from the centre (source of the wave). In other words, these particles are said to be in same phase because their state of motion is same (refer to Sec. 1.2 of Unit 1 to recall the meaning of the state of motion). **The locus of all such points which are in the same phase is called a wavefront.** So, the circles representing 2-D waves on a water surface are called the wavefronts of the wave.

Can you guess what would be the shape of wavefronts of 3-D waves generated by a point source such as a point source of sound waves? The

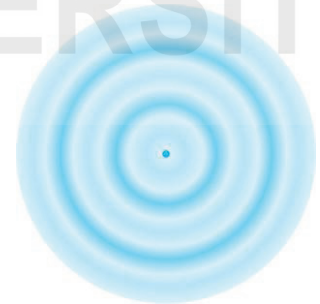


Fig. 4.2: Waves on a water surface.

wavefronts will be spherical centred at the source. Recall from Sec. 1.4 of Unit 1 that if the point source of wave is at a sufficiently large distance, the spherical waves can be approximated as plane waves and the wavefronts of the plane waves are planes perpendicular to the direction of propagation of waves. The wavefronts of spherical and plane waves are depicted in Fig. 4.3.

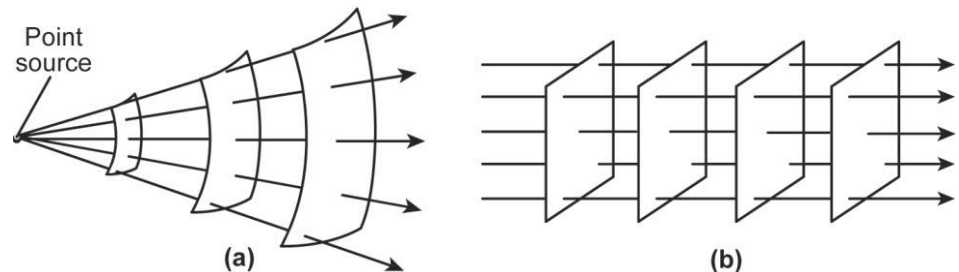


Fig. 4.3: a) Spherical; b) plane wavefronts.

Huygens principle also states that any point on a wavefront serves as a source of waves and these point sources generate spherical waves. These point sources on a wavefront are called **secondary sources** of waves and the waves generated by them are called **secondary wavelets**. The spherical wavelets propagate with the same speed as that of the wave.

Now, let us apply Huygens principle and the concepts of wavefronts and secondary wavelets to describe the propagation of electromagnetic wave. Consider a point source, S which is generating 3-D waves. Suppose AB represents a part of the spherical wavefront of the wave at some time, say t_1 . Now, to construct the wavefront of the wave at some later time, say $t_2 (> t_1)$, assume that points a, b, c, d , and e on the wavefront AB are the secondary sources which generate spherical secondary wavelets. That is, spherical waves are being generated by each of these secondary sources.

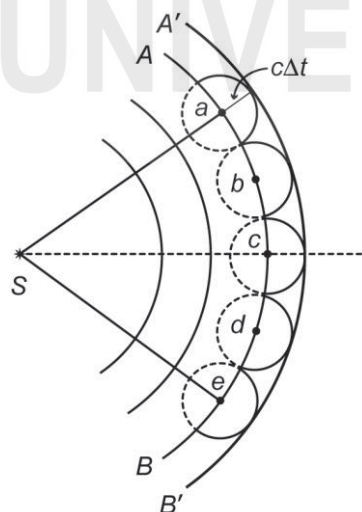


Fig. 4.4: Constructing wavefronts using Huygens principle.

The secondary wavelets generated by point sources a, b, c, d and e propagate away from their respective sources with the same velocity, c as that of the electromagnetic wave. In the time interval, $\Delta t (= t_2 - t_1)$ the spherical secondary wavelets will travel a distance equal to $c\Delta t$. So, we can construct the wavefronts of each of these secondary wavelets by drawing a circle of radius $c\Delta t$ centred at each point sources a, b, c, d and e , as shown in Fig. 4.4.

The envelop $A'B'$ of these circles in the forward direction is the new wavefront of the wave generated by the source S . The straight lines originating from the source and perpendicular to the wavefront is called **ray**. A ray of light gives the direction of propagation of light. **Note that the wavefront and ray are, respectively, imaginary surface and imaginary line constructed on the basis of Huygens principle.**

With the help of the method described above for constructing wavefronts for a spherical wave using Huygens principle, you can construct wavefronts for a plane wave. Note that wavefronts of a plane wave are planes perpendicular to the direction of propagation of the wave.

Properties of Wavefronts

The wavefronts have following properties:

1. Wavefront is an imaginary surface constructed by joining all the points having the same phase at a given instant of time.
2. The shape of the wavefront depends on the nature of wave emanating from a source.
3. Wavefronts propagate with the speed of the wave.
4. Wavefronts are always normal to the rays.
5. Wavefronts do not propagate in the backward direction.

These properties are satisfied by the wavefronts associated with light waves. Although Huygens principle provided wave model based explanations for phenomena like reflection and refraction of light, it could not give satisfactory explanation for phenomena like rectilinear propagation of light and polarisation of light. Later on, Huygens principle was modified by Fresnel by adding the concept of superposition or interference of waves. The modified Huygens principle, called Huygens-Fresnel principle, could successfully account for the experimental observations related to rectilinear propagation and diffraction of light. You will learnt more about this principle and how it account for diffraction of light later in this course.

Let us now sum up what you have learnt in this unit.

4.4 SUMMARY

Concept	Description
Wave equation	■ The wave equations for electric and magnetic fields, obtained on the basis of Maxwell's equations are: $\nabla^2 \vec{E} = \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}$ and $\nabla^2 \vec{B} = \mu_0 \epsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2}$ These wave equations indicate that the time varying $\vec{E}$ and $\vec{B}$ fields constitute an electromagnetic wave. The speed of the electromagnetic wave in vacuum is given by $\sqrt{1/(\mu_0 \epsilon_0)}$ which is equal to the speed of light. Thus, light is an electromagnetic wave.

Transverse nature of electromagnetic waves

- The mathematical expressions for $\vec{E}$ and $\vec{B}$ field constituting an electromagnetic wave must satisfy the respective wave equations as well as Maxwell's equations. These conditions are met only when the $\vec{E}$ and $\vec{B}$ fields are perpendicular to the direction of propagation of electromagnetic wave. That is, electromagnetic waves are transverse in nature.
- For an electromagnetic wave propagating along the positive z-direction, the electric and magnetic fields can be expressed as

$$\vec{E} = \hat{y} E_0 \sin(kz - \omega t)$$

$$\text{and } \vec{B} = \hat{x} B_0 \sin(kz - \omega t)$$

The magnitudes of $\vec{E}$ field and $\vec{B}$ field constituting an electromagnetic wave in vacuum are related by the relation

$$E_0 = cB_0 = \frac{\omega}{k} B_0$$

Huygens principle

- The wave model of light is based on the Huygens principle. According to this principle, every point on a wavefront behaves like a source of spherical secondary wavelets and the wavefront at some later time is the envelop of these wavelets.
- Wavefront is the locus of all points which are in the same phase.
- Wavefronts propagate in forward direction with the velocity of the wave.

4.5 TERMINAL QUESTIONS

1. Write the expression for the $\vec{E}$ and $\vec{B}$ fields of plane electromagnetic wave propagating in the positive z-direction in vacuum such that its electric field is in the positive x-direction. The amplitude of the magnetic field associated with the wave is 300 nT and the frequency of the wave is 8.5 GHz.
2. The wavelength of an electromagnetic wave is 400 nm. The wave is travelling along the positive y-direction and its $\vec{E}$ field is along the z-direction. Determine the angular frequency ω and wave number k of the wave. If the electric field amplitude is 500 Vm^{-1} , calculate the amplitude of the magnetic field.
3. The $\vec{E}$ field of a plane electromagnetic wave travelling in vacuum is given by

$$\vec{E} = \hat{x} [230 \sin(1.20 \times 10^7 z - 3.6 \times 10^{15} t)] \text{ Vm}^{-1}$$

Determine the frequency and wavelength of the wave. Does the electromagnetic wave correspond to visible light?

4. The amplitude of electric field associated with an electromagnetic wave is 110 Vm^{-1} and the frequency of the wave is 60 MHz. Determine the amplitude of the associated magnetic field, and angular frequency, wave number and wavelength of the wave.

4.6 SOLUTIONS AND ANSWERS

Self-Assessment Questions

1. As per the problem, the wavelength λ , of the electromagnetic wave is

$$\lambda = 450 \text{ nm} = 450 \times 10^{-9} \text{ m}$$

The electromagnetic wave propagates in vacuum with the speed of light, $c = 3 \times 10^8 \text{ ms}^{-1}$. And, we know that the speed of electromagnetic wave is given by

$$c = \lambda f \Rightarrow f = \frac{c}{\lambda} = \frac{3 \times 10^8 \text{ ms}^{-1}}{450 \times 10^{-9} \text{ m}} = 6.66 \times 10^{14} \text{ s}^{-1}$$

Angular frequency, $\omega = 2\pi f = 2 \times 3.14 \times 6.66 \times 10^{14} \text{ s}^{-1} = 4.1 \times 10^{15} \text{ s}^{-1}$

Wave number, $k = \frac{2\pi}{\lambda} = \frac{2 \times 3.14}{450 \times 10^{-9} \text{ m}} = 1.4 \times 10^7 \text{ m}^{-1}$

2. The expression for the magnetic field $\vec{B}$ associated with the electromagnetic wave propagating along the positive z-direction is given as

$$\vec{B} = \vec{B}_0 \sin(kz - \omega t) \quad (\text{i})$$

This expression for $\vec{B}$ must satisfy Maxwell's equation,

$$\vec{\nabla} \cdot \vec{B} = 0 \quad (\text{ii})$$

Substituting the value of $\vec{B}$ from Eq. (i) in Eq. (ii), we can write the LHS as

$$\begin{aligned} & \left(\hat{x} \frac{\partial}{\partial x} + \hat{y} \frac{\partial}{\partial y} + \hat{z} \frac{\partial}{\partial z} \right) \cdot (\hat{x} B_{0x} + \hat{y} B_{0y} + \hat{z} B_{0z}) \sin(kz - \omega t) \\ &= B_{0x} \frac{\partial}{\partial x} [\sin(kz - \omega t)] + B_{0y} \frac{\partial}{\partial y} [\sin(kz - \omega t)] \\ & \quad + B_{0z} \frac{\partial}{\partial z} [\sin(kz - \omega t)] \end{aligned}$$

From the above, we note that Maxwell's equation given by Eq. (ii) will be satisfied if and only if

$$B_{0z} = 0$$

The above expression for the z component of the $\vec{B}$ field of an electromagnetic wave indicates that the z-component of the magnetic field is zero. Thus, $\vec{B}$ field is perpendicular to the direction (z-direction) of

propagation of the electromagnetic wave. This is possible only if electromagnetic wave is transverse.

3. a) Sound waves are responsible for hearing whereas light waves (electromagnetic waves) are responsible for seeing. The two differences between these two waves are:
 - i) Sound is a mechanical wave which requires a material medium for its propagation whereas light, being electromagnetic wave, can propagate in vacuum as well.
 - ii) Sound waves are longitudinal whereas light wave are transverse in nature.

The similarities between sound waves and light waves are:

- i) Both the sound and light waves are characterised by the wave parameters such as amplitude, wavelength, frequency and phase.
 - ii) The propagation of both sound waves and light waves is governed by wave equations.
- b) The amplitudes of electric field $\vec{E}$ and magnetic field $\vec{B}$ associated with electromagnetic wave are related to each other as

$$E_0 = cB_0$$

So, if the amplitude E_0 of the $\vec{E}$ field is doubled, the amplitude of the $\vec{B}$ field will also be doubled.

Terminal Questions

1. The $\vec{E}$ and $\vec{B}$ fields associated with the electromagnetic wave propagating in z-direction are given as

$$\vec{E} = \vec{E}_0 \sin(kz - \omega t) = \hat{x} E_0 \sin(kz - \omega t)$$

$$\vec{B} = \vec{B}_0 \sin(kz - \omega t) = \hat{y} B_0 \sin(kz - \omega t)$$

As per the problem, amplitude B_0 of the $\vec{B}$ field is 300 nT, and frequency of the wave is 8.5 GHz. So,

$$\omega = 2\pi f = 2 \times 3.14 \times 8.5 \times 10^9 \text{ Hz} = 5.3 \times 10^{10} \text{ s}^{-1}$$

Since the wave is travelling in vacuum, its speed $c = 3 \times 10^8 \text{ ms}^{-1}$. So,

$$\lambda = \frac{c}{f} = \frac{3 \times 10^8 \text{ ms}^{-1}}{8.5 \times 10^9 \text{ s}^{-1}} = 0.03 \text{ m}$$

So,

$$k = \frac{2\pi}{\lambda} = \frac{2 \times 3.14}{0.03 \text{ m}} = 2.1 \times 10^{-2} \text{ m}^{-1}$$

The magnitude of electric field

$$E_0 = c B_0 = 90 \text{ Vm}^{-1}$$

So, the expression for $\vec{E}$ and $\vec{B}$ fields are

$$\vec{E} = \hat{x} \{ (90 \text{ Vm}^{-1}) \sin[(2.1 \times 10^{-2} \text{ m}^{-1})z - (5.3 \times 10^{10} \text{ s}^{-1})t] \}$$

$$\vec{B} = \hat{y} \{ (300 \text{ nT}) \sin[(2.1 \times 10^{-2} \text{ m}^{-1})z - (5.3 \times 10^{10} \text{ s}^{-1})t] \}$$

2. As per the problem, wavelength λ of the electromagnetic wave is 400 nm.
So, its frequency, f is

$$f = \frac{c}{\lambda} = \frac{3 \times 10^8 \text{ ms}^{-1}}{400 \times 10^{-9} \text{ m}} = 7.5 \times 10^{14} \text{ s}^{-1}$$

So, the angular frequency of the wave is

$$\omega = 2\pi f = 2 \times 3.14 \times 7.5 \times 10^{14} \text{ s}^{-1} = 4.7 \times 10^{15} \text{ s}^{-1}$$

And, the wave number, k is

$$k = \frac{2\pi}{\lambda} = \frac{2 \times 3.14}{400 \times 10^{-9} \text{ m}} = 1.6 \times 10^{-7} \text{ m}^{-1}$$

The amplitude of $\vec{E}$ and $\vec{B}$ fields are related as

$$B_0 = \frac{E_0}{c} = \frac{500 \text{ Vm}^{-1}}{3 \times 10^8 \text{ ms}^{-1}} = 1.67 \times 10^{-6} \text{ T}$$

3. As per the problem,

$$\vec{E} = \hat{x} [230 \sin(1.20 \times 10^7 z - 3.6 \times 10^{15} t)] \text{ Vm}^{-1}$$

The frequency of the wave is

$$f = \frac{\omega}{2\pi} = \frac{3.6 \times 10^{15} \text{ s}^{-1}}{2 \times 3.14} = 5.7 \times 10^{14} \text{ s}^{-1}$$

And, the wavelength is

$$\lambda = \frac{2\pi}{k} = \frac{2 \times 3.14}{1.2 \times 10^7 \text{ m}^{-1}} = 5.2 \times 10^{-7} \text{ m}$$

Since the wavelength, λ is $5.2 \times 10^{-7} \text{ m} = 520 \times 10^{-9} \text{ m} = 520 \text{ nm}$, the electromagnetic wave is in the range of wavelength for visible light.

The expression for the associated magnetic field is

$$\vec{B} = \hat{y} [B_0 \sin(1.20 \times 10^7 z - 3.6 \times 10^{15} t)] \text{ T}$$

$$\text{Since } B_0 = \frac{E_0}{c} = \frac{230 \text{ Vm}^{-1}}{3 \times 10^8 \text{ ms}^{-1}} = 7.6 \times 10^{-7} \text{ T}$$

$$\text{So, } \vec{B} = \hat{y} [(7.6 \times 10^{-7}) \sin(1.20 \times 10^7 z - 3.6 \times 10^{15} t)] \text{ T}$$

4. The amplitude of the magnetic field associated with electromagnetic wave is

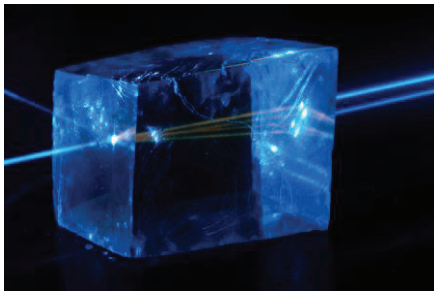
$$B_0 = \frac{E_0}{c} = \frac{110 \text{Vm}^{-1}}{3 \times 10^8 \text{ms}^{-1}} = 360 \text{nT}$$

$$\text{Angular frequency, } \omega = 2\pi f = 2 \times 3.14 \times (60 \times 10^6) \text{s}^{-1} = 3.77 \times 10^8 \text{s}^{-1}$$

$$\text{Wave number, } k = \omega/c = \frac{3.77 \times 10^8 \text{s}^{-1}}{3 \times 10^8 \text{ms}^{-1}} = 1.25 \text{m}^{-1}$$

$$\text{Wavelength, } \lambda = \frac{2\pi}{k} = \frac{2 \times 3.14}{1.25 \text{m}^{-1}} = 5.02 \text{m}$$





UNIT 5

POLARISATION OF LIGHT

We see two images of an object when it is observed through certain crystals. This is caused due to birefringence or double refraction. Such crystals are called birefringent crystals and are used for producing polarised light.

(Source of picture: wikimedia.org)

Structure

- | | |
|---|--|
| <p>5.1 Introduction
Expected Learning Outcomes</p> <p>5.2 What is Polarisation?</p> <p>5.3 Simple States of Polarised Light
Linear Polarisation
Circular Polarisation
Elliptical Polarisation</p> | <p>5.4 Principles of Producing Linearly Polarised Light
Ideal Polariser: Malus's Law
Polarisation by Reflection: Brewster's Law
Polarisation by Double Refraction
Selective Absorption: Dichroism</p> <p>5.5 Wave Plates: Circular and Elliptical Polarisers</p> <p>5.6 Summary</p> <p>5.7 Terminal Questions</p> <p>5.8 Solutions and Answers</p> |
|---|--|

STUDY GUIDE

In this unit, you will study polarisation of light. Polarisation of light is an important phenomenon because it is an experimental evidence of transverse nature of light. In Unit 4, you have learnt that light waves are transverse in nature because it is an electromagnetic wave. This means that the electric and magnetic fields associated with light waves are perpendicular to its direction of propagation. The transverse nature of the mechanical waves on a string is easy to visualise. But, we cannot 'see' the transverse nature of light waves because it is the variation of $\vec{E}$ and $\vec{B}$ fields in space and time. The transverse nature and hence polarisation of light is inferred on the basis of experiments. If you revise the propagation of transverse wave on a string from Unit 1, it will help you understand polarisation and transverse nature of light better. You should also revise vector algebra studied in second semester course BPHCT-133 because we will use expressions of electric field and magnetic field in vector form. In addition, you need to focus on the geometry of various optical components needed for producing polarised light. Try to solve SAQs and TQs yourself first before looking at their solutions and answers.

"Physics is the belief that a simple and consistent description of nature is possible."

Niels Bohr

5.1 INTRODUCTION

In the previous four units of this course, you have studied about mechanical waves and electromagnetic waves. You now have fairly good understanding of wave motion and how do we describe waves graphically as well as mathematically. In Unit 4, you learnt that light is also a wave phenomenon; light is an electromagnetic wave. Therefore, we can apply the concepts related to wave motion such as superposition of waves for the analysis of optical phenomena such as interference and diffraction.

In Unit 4, you also learnt that Maxwell's equations and the wave equations for electric field ($\vec{E}$) and magnetic field ($\vec{B}$) led us to conclude that electromagnetic waves (including light) are transverse in nature. This means that the $\vec{E}$ and $\vec{B}$ fields associated with the electromagnetic waves are perpendicular to its direction of propagation. **The experimentally observed phenomenon of polarisation of light requires that light should be a transverse wave.** So, Maxwell's electromagnetic theory of light did explain this experimental observation. In this unit, you will study polarisation of light applying the concepts of wave motion and electromagnetic waves studied so far.

You must have seen people using antiglare goggles as also antiglare windshields for their cars. Do you know that polarisation of light has something to do with these? Polarisation of light also plays a vital role in designing sky light filters for cameras and numerous optical instruments including the polarising microscope and polarimeter. All these optical devices and equipment are based on polarisation of light.

We begin this unit with an explanation as to what we mean by polarisation and polarised waves in general in Sec. 5.2. You will learn what distinguishes the polarised light from unpolarised light. The polarisation of light is understood on the basis of the principle of superposition of waves. In Sec. 5.3, you will learn that, depending upon the phase relation between the two superposing linearly polarised light waves, we can have three distinct states of polarisations. These states of polarisation are called linear, circular and elliptical. You will also learn that linear and circular polarisations are the special cases of elliptical polarisation. In Sec. 5.4, you will learn Malus's law which defines an ideal polariser and Brewster's law which prescribes the condition under which we can obtain linearly polarised light due to reflection. These laws enable us to devise methods and optical components for producing linearly polarised light. In addition, in this section, you will also learn about the properties such as double refraction (or birefringence) and selective absorption of certain materials which are used for producing polarised light. In Sec. 5.5, you will learn about optical components called half wave plate and quarter wave plate which are used for producing polarised light.

With this unit, we come to the end of Block 1 of the course. In the next Block, which comprises three units, you will learn about interference of light.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ explain what is linearly, circularly and elliptically polarised state of light;
- ❖ describe how light can be polarised by reflection;

- ❖ solve simple problems based on Malus's law and Brewster's law;
- ❖ explain how optical birefringence is used in production of polarised light;
- ❖ describe the construction and working of Nicol prism;
- ❖ explain the production of linearly polarised light due to dichroism; and
- ❖ understand the working of half wave plate and quarter wave plate.

5.2 WHAT IS POLARISATION?

You have learnt in Unit 4 that light is an electromagnetic wave and the electromagnetic waves are transverse in nature. This means that the electric and magnetic fields associated with the light wave are perpendicular to the direction of propagation. **When the electric (or magnetic) field variations in a light wave are in a fixed plane, we say that the light is polarised.**

To visualise the polarisation phenomenon, let us consider a transverse wave on a string. Refer to Fig. 5.1 which depicts a transverse wave travelling along a string. From Fig. 5.1a, you will note that the string vibrates only in the vertical plane. That is, vibrations of the particles of the string are confined to just **one** single plane. **Such a wave is said to be linearly (or plane) polarised.**

How would you classify waves shown in Fig. 5.1b and c? The wave shown in Fig. 5.1b is also plane polarised since vibrations are confined to the horizontal plane. But, the wave in Fig. 5.1c is unpolarised because simultaneous vibrations in more than one plane are present. However, it can be polarised by placing a slit in its path as in Fig. 5.1d. When the first slit (slit 1) is oriented vertically, horizontal vibrations are cut off. This means that only vertical vibrations are allowed to pass so that the wave is linearly or plane polarised. What happens when a horizontal slit (slit 2) is placed beyond the vertical slit (slit 1) in the path of propagation of the wave? In that situation, both the horizontal as well as vertical components (of the incident wave) will be blocked. And the wave amplitude will reduce to zero as shown in Fig. 5.1d beyond slit 2.

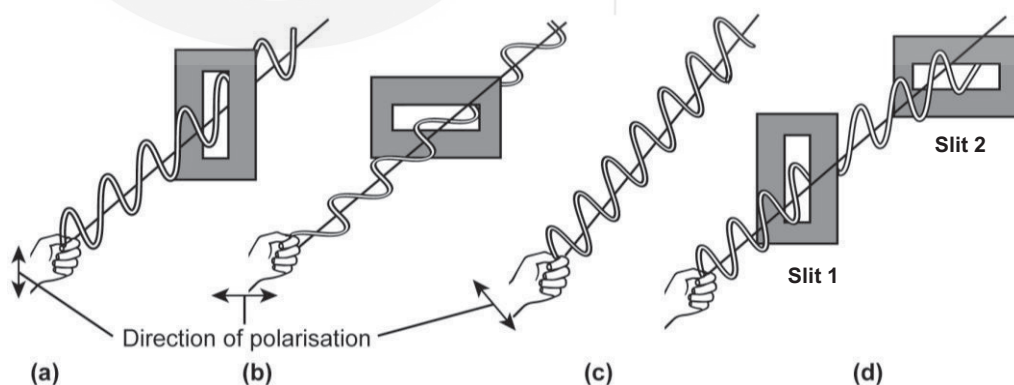


Fig. 5.1: a) A vertically plane polarised transverse wave on a string; b) a horizontally plane polarised transverse wave; c) an unpolarised wave; d) the wave in (c) becomes plane polarised after passing through slit 1 and the wave amplitude reduces to zero if another slit (slit 2), oriented perpendicular to slit 1 is introduced.

The polarisation is a property which applies only to transverse waves.

Can you guess why is it so? Why longitudinal waves cannot be polarised? In a transverse wave, the oscillations of the particles of the medium can be in any

direction in the plane perpendicular to the direction of wave propagation. When the oscillation are confined to only one direction, say by means of placing a slit in the path of the wave (Fig. 5.1d), the wave is said to be plane polarised. In contrast, in case of longitudinal waves, such as sound waves, the direction of oscillations of the particles of the medium is always along the direction to wave propagation. So, longitudinal waves do not exhibit polarisation.

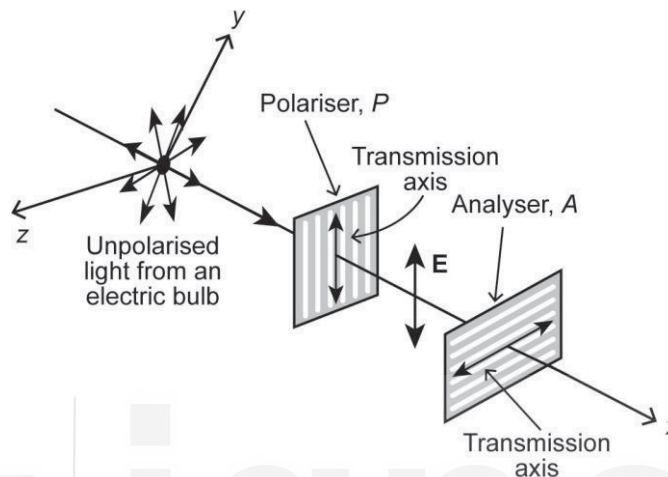


Fig. 5.2: Schematic diagram of the apparatus for observing polarisation of light.

Let us now consider what do we mean by polarisation of visible light. Refer to Fig. 5.2 which depicts a typical experimental arrangement to observe polarisation. The light from a source, say an electric bulb, is made to pass through a **polariser** P , which works for light waves in the same way as slit 1 in Fig. 5.1 works for mechanical wave. The light waves emerging from the bulb is unpolarised. It is observed that the intensity of light emerging out of P reduces to about 50%. It is also observed that rotating P in its own plane introduces no further change in light intensity.

Now, if a second identical polariser called **analyser**, A is introduced in the path of light so that its transmission axis is parallel to P , the intensity of light from the bulb remains unaffected. But, rotating A in its own plane has a dramatic effect! For 90° rotation with respect to the transmission axis of P , the light is nearly cut-off.

You may ask: **How do we explain this experimental observation in terms of electromagnetic theory?** As you learnt in Unit 4, electromagnetic nature of light demands complete description of associated electric (and magnetic) field vectors and the way it varies with respect to the direction of propagation. For the arrangement shown in Fig. 5.2, the electric field vector of the light emitted by the source can have all orientations in the yz -plane. Such light is called **unpolarised light**; that is, the $\vec{E}$ field vector keeps changing its orientation in a random manner in the plane perpendicular to the direction of wave propagation. The light wave propagates as such till it reaches the polariser P , which allows essentially unhindered passage of $\vec{E}$ field vector oriented parallel to its **transmission axis**. If the transmission axis is along y -axis, the electric field along y -direction (E_y) passes through it unaffected. In addition, the y -components of electric field vectors inclined to y -axis can also pass through P . Thus, after passing through the polariser P , the electric vectors oriented only along y -axis will be present. **When electric vector**

A **polariser** is an optical device which converts a beam of unpolarised light into a beam of polarised light. A polariser polarises light waves by permitting only those electric field vectors, associated with the light waves, to pass through which are parallel to its (polariser's) **transmission axis**.

oscillates along a straight line in a plane perpendicular to the direction of propagation, the light is said to be linearly or plane polarised. The plane polarised light wave further travels to the analyser A , which is identical to P . When A is at 90° with respect to P , it can allow only the z -components of the $\vec{E}$ field to pass. Since only y -components of $\vec{E}$ field are present in the wave incident on A , no light is transmitted by A .

So, on the basis of the above qualitative description of polarisation, we may conclude that:

1. No polarisation of longitudinal waves occurs as the vibrations of the particles of the medium are along the direction of propagation of the waves only.
2. The transverse nature of light is responsible for its polarisation.

You may now like to know: **Do natural (or ordinary) light sources emit polarised light?** Answer to this question is 'yes' as well as 'no'! You know that emission of light involves a large number of randomly oriented atomic (or molecular) emitters. Every individual excited atom radiates polarised waves for about 10^{-8} s. These waves form a resultant wave of given polarisation which persists for the lifetime of the excited atom. At the same time, other atoms (molecules) also emit light waves, whose resultant states of polarisation may be quite different. Because of this randomness, every orientation of electric vector in space is equally probable in the light emitted by ordinary sources. That is, electric vectors associated with light waves from an ordinary source are oriented in all directions in a plane perpendicular to the direction of propagation and thus there is a completely unpredictable change in the overall polarisation. Moreover, due to such rapid changes, individual resultant polarisation states become almost indiscernible. The light is then said to be **unpolarised**.

The visible light emitted by ordinary sources does not correspond to either of these extremes. The oscillations of the associated electric field vectors are neither completely regular nor completely irregular. That is, light from any source is **partially polarised**. We ascribe a degree of polarisation to partially polarised light. The degree of polarisation is one for completely polarised light and zero for unpolarised light.

The next logical step perhaps would be to know various types of polarised light. Let us learn about this aspect now.

5.3 SIMPLE STATES OF POLARISED LIGHT

You now know that in electromagnetic theory, light propagation is depicted as evolution of electric field vector in a plane perpendicular to the direction of transmission. For unpolarised light, spatial variation of electric field at any given time is more or less irregular. For **plane** or **linearly polarised light**, the tip of electric vector oscillates up and down in a straight line in the same plane. The space variation of $\vec{E}$ field for linearly polarised light wave is shown in Fig. 5.3a. Note that the orientation of $\vec{E}$ field is fixed (along the y -direction) although its magnitude and sign vary with time. Thus, in this case, the electric field resides in what is known as **plane of polarisation** or the **plane of vibration** (shown by dotted line in Fig. 5.3a) The reference plane shows the path followed by the tip of the electric vector as time passes. The tip of $\vec{E}$ field

In a right handed coordinate system, if a right handed screw is turned so that it rotates the x -axis towards the y -axis, the direction of advance of the screw represents the positive z -axis.

vector executes one full cycle as one full wavelength passes through the reference plane.

In addition to linear polarisation, there are two other states of polarisation, namely **circular polarisation** and **elliptical polarisation**. The path followed by the tip of $\vec{E}$, as seen in the reference plane as time passes, for these states of polarisation are shown in Fig. 5.3b and c respectively.

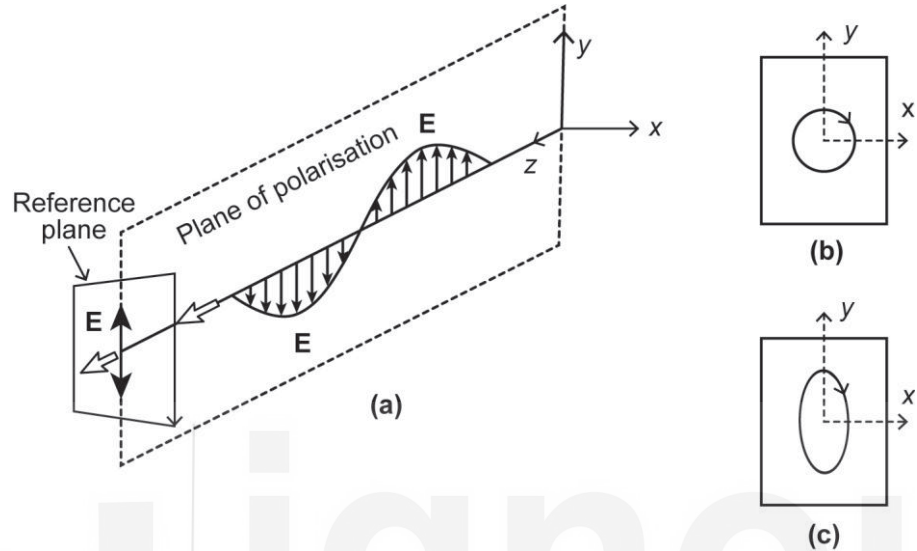


Fig. 5.3: Spatial variation of electric field vector for a) linearly polarised light; note that, in the reference plane, the bold face arrow indicates the path taken by the tip of the electric field vector as time passes; b) and c) show the path taken by the tip of the electric field vector for circularly and elliptically polarised light.

The yz-plane (or $x = 0$ plane) in Fig. 5.3 is the plane of polarisation of the wave. We can identify other states of polarisation **by looking at the trajectories of the tip of the electric field vector** as the wave pass through the reference plane. You should always look at the reference plane from the side away from the source (looking back at the source) for the definitions to be unique.

You should note that the addition of a positive ϕ in the phase ($kz - \omega t$) means $\vec{E}_2(z, t)$ lags $\vec{E}_1(z, t)$ for $\phi > 0$ and vice-versa.

The states of polarisation namely linear, circular and elliptical are the outcome of the superposition of two light waves whose electric field vectors are mutually perpendicular and the waves have same frequency and direction of propagation. Let us now mathematically analyse how such superposition gives rise to linearly, circularly or elliptically polarised light.

5.3.1 Linear Polarisation

Suppose that two light waves are propagating along the z-direction. Let their electric field vectors be mutually perpendicular, i.e. we choose these along the x and y-axes. So, from Unit 4, you may recall that the electric field vectors associated with these light waves can be represented as

$$\vec{E}_1(z, t) = \hat{x} E_{01} \cos(kz - \omega t) \quad (5.1)$$

$$\text{and } \vec{E}_2(z, t) = \hat{y} E_{02} \cos(kz - \omega t + \phi) \quad (5.2)$$

where $\hat{x}$ and $\hat{y}$ are unit vectors along the x and y-axes, respectively. E_{01} and E_{02} are the maximum values of $\vec{E}_1$ and $\vec{E}_2$, respectively, and ϕ is the phase difference between the two waves.

From Unit 2, you may recall that the nature of the resultant wave due to the superposition of these two waves is determined by the phase difference between them and the value of the ratio E_{02}/E_{01} . Mathematically, we can write electric field vector of the resultant wave due to the superposition of these two waves as

$$\begin{aligned} \vec{E}(z, t) &= \vec{E}_1(z, t) + \vec{E}_2(z, t) \\ &= \hat{x} E_{01} \cos(kz - \omega t) + \hat{y} E_{02} \cos(kz - \omega t + \phi) \end{aligned} \quad (5.3)$$

Let us first take the simplest case where the phase difference, ϕ is zero or an integral multiple of $\pm 2\pi$. That is, when in-phase waves are superposed, Eq. (5.3) takes the form

$$\vec{E}(z,t) = (\hat{x}E_{01} + \hat{y}E_{02})\cos(kz - \omega t) \quad (5.4)$$

From Eq. (5.4), we note that the amplitude of the resultant wave is $\sqrt{E_{01}^2 + E_{02}^2}$ and the electric field oscillations in the reference frame make an angle $\theta = \tan^{-1}(E_{02}/E_{01})$ with the x-axis.

For the special case of in-phase waves of equal amplitude ($E_{01} = E_{02} = E_0$), the resultant wave has amplitude equal to $\sqrt{2}E_0$ and the associated electric vector is oriented at 45° with the x-axis (see the margin remark). So, we may conclude that **when two in-phase linearly polarised light waves are superposed, the resultant wave has fixed orientation as well as amplitude.** That is, it too is linearly polarised, as depicted in Fig. 5.4a. In the plane of observation, you will see a single resultant $\vec{E}$ oscillating sinusoidally in time along an inclined line (Fig. 5.4b). The tip of the $\vec{E}$ -field progresses through one complete cycle as the wave advances along the z-axis through one wavelength.

If we reverse this process, we can say that **any linearly polarised light can be visualised as a combination of two linearly polarised light with planes of polarisation parallel to $x = 0$ and $y = 0$ planes.** (This is similar to resolving a vector in a plane along two mutually perpendicular directions.) In the subsequent sections, you will use this result frequently.

When the amplitudes of superposing waves are equal, i.e.

$E_{01} = E_{02} = E_0$, the amplitude of the resultant electric field $\vec{E}$ is [Eq. (5.4)]:

$$\begin{aligned} &= \sqrt{(E_0)^2 + (E_0)^2} \\ &= \sqrt{2}E_0 \end{aligned}$$

and

$$\begin{aligned} \tan\theta &= \frac{E_{02}}{E_{01}} = \frac{E_0}{E_0} = 1 \\ &= \tan 45^\circ \end{aligned}$$

or $\theta = 45^\circ$

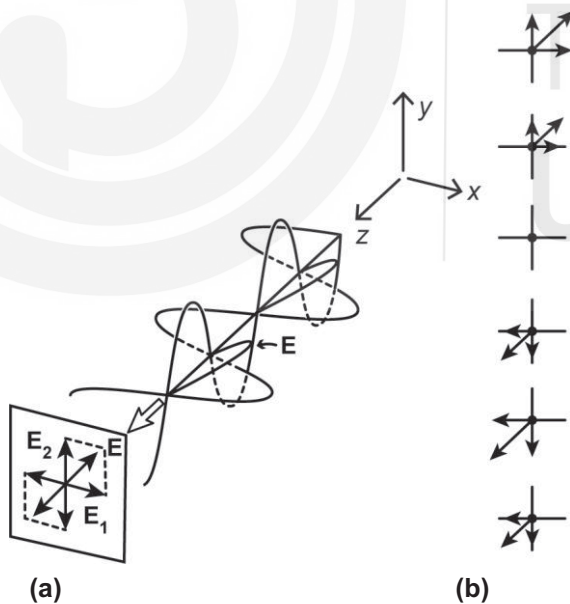


Fig. 5.4: a) Schematic representation of a linearly polarised light wave due to superposition of two linearly polarised light waves; b) the orientation and movement of the $\vec{E}$ -field of the resultant light wave as time passes.

If the phase difference between the two superposing orthogonal linearly polarised light waves is an odd integral multiple of $\pm \pi$, the resultant wave will again be linearly polarised. You should convince yourself that the electric field vector of the resultant wave in this case is given as

$$\vec{E}(z,t) = (\hat{x}E_{01} - \hat{y}E_{02})\cos(kz - \omega t) \quad (5.5)$$

What will be the orientation of the resultant electric vector in the reference plane for this case? To know the answer of this question, work-out the following SAQ.

SAQ 1 - Linearly polarised light

Depict the orientation of electric field vector defined by Eq. (5.5) in the reference plane.

$$\cos(-\theta) = \cos \theta$$

$$\cos\left(\frac{\pi}{2} - \theta\right) = \sin \theta$$

5.3.2 Circular Polarisation

We now investigate the nature of the resultant light wave arising due to superposition of two orthogonal plane polarised light waves whose amplitudes are equal ($E_{01} = E_{02} = E_0$) but phases differ by $\pi/2$, i.e. their relative phase difference $\phi = 2n\pi - \frac{\pi}{2}$, $n = 0, \pm 1, \pm 2, \dots$. For $n = 0$, we can rewrite Eqs. (5.1) and (5.2) as

$$\vec{E}_1(z, t) = \hat{x} E_0 \cos(kz - \omega t) \quad (5.6a)$$

$$\vec{E}_2(z, t) = \hat{y} E_0 \sin(kz - \omega t) \quad (5.6b)$$

The resultant wave due to the superposition of the above linearly polarised light waves is given by

$$\vec{E}(z, t) = E_0 [\hat{x} \cos(kz - \omega t) + \hat{y} \sin(kz - \omega t)] \quad (5.7)$$

In Eq. (5.7), you may note that the scalar amplitude of the electric field $\vec{E}$ of the resultant wave is constant ($= E_0$) but its orientation varies with time. To convince yourself about the value of the scalar amplitude of the resultant electric field $\vec{E}$ as given by Eq. (5.7), go through the following example.

EXAMPLE 5.1: AMPLITUDE OF THE RESULTANT ELECTRIC FIELD

Show that the scalar amplitude of the electric field of the resultant wave given by Eq. (5.7) is equal to E_0 .

SOLUTION ■ From vector algebra, you know that the magnitude of a vector $\vec{A}$ is given as

$$|\vec{A}| = (\vec{A} \cdot \vec{A})^{1/2}$$

So, from Eq. (5.7), we can write that

$$\begin{aligned} |\vec{E}| &= (\vec{E} \cdot \vec{E})^{1/2} \\ &= [E_0 [\hat{x} \cos(kx - \omega t) + \hat{y} \sin(kz - \omega t)] \cdot E_0 [\hat{x} \cos(kx - \omega t) \\ &\quad + \hat{y} \sin(kz - \omega t)]]^{1/2} \\ &= (E_0^2)^{1/2} [\cos^2(kx - \omega t) + \sin^2(kz - \omega t)]^{1/2} \end{aligned}$$

because $\hat{x} \cdot \hat{x} = 1$ and $\hat{x} \cdot \hat{y} = 0 = \hat{y} \cdot \hat{x}$

So, $|\vec{E}| = E_0$

To determine the trajectory along which the tip of electric field $\vec{E}$ of the resultant wave moves, we can readily combine the scalar parts of Eqs. (5.6a) and (5.6b) to yield

$$\left(\frac{E_1}{E_0}\right)^2 + \left(\frac{E_2}{E_0}\right)^2 = 1 \quad (5.8)$$

which is the **equation of a circle**. Thus, we may conclude that the orientation of resultant electric vector changes continuously and its tip moves along a circle as the wave propagates (time passes). In the reference plane, therefore, we will observe a dot (tip of the electric field vector) moving along a circle.

This means that $\vec{E}$ is not restricted to a single plane. The question now arises: What is the direction of rotation? Obviously there are two possibilities: clockwise and counter-clockwise. To know which of these is relevant here, we tabulate $\vec{E}$ at different space points at a given time say, $t = 0$, using Eq. (5.7). The tabulated values of $\vec{E}$ are as under:

Location in space	$z = 0$	$z = \frac{\lambda}{8}$	$z = \frac{\lambda}{4}$	$z = \frac{3\lambda}{8}$	$z = \frac{\lambda}{2}$	$z = \frac{5\lambda}{8}$	$z = \frac{3\lambda}{4}$	$z = \frac{7\lambda}{8}$	$z = \lambda$
Resultant Electric field, $\vec{E}$	$\hat{x}E_0$	$\frac{\hat{x} + \hat{y}}{\sqrt{2}}E_0$	$\hat{y}E_0$	$\frac{\hat{y} - \hat{x}}{\sqrt{2}}E_0$	$-\hat{x}E_0$	$\frac{\hat{x} + \hat{y}}{\sqrt{2}}E_0$	$-\hat{y}E_0$	$\frac{\hat{x} - \hat{y}}{2}E_0$	$\hat{x}E_0$

The orientations of electric field $\vec{E}$ at different values of z are depicted in Fig. 5.5. If you position yourself in the reference plane and observe the evolution of $\vec{E}$ from $z = \lambda$ to $z = 0$ (backward towards source), you will find that the **tip of $\vec{E}$ rotates clockwise**. Such a light wave is said to be **right-circularly polarised** or **right circular light**. The electric field makes one complete rotation as the wave advances through one wavelength.

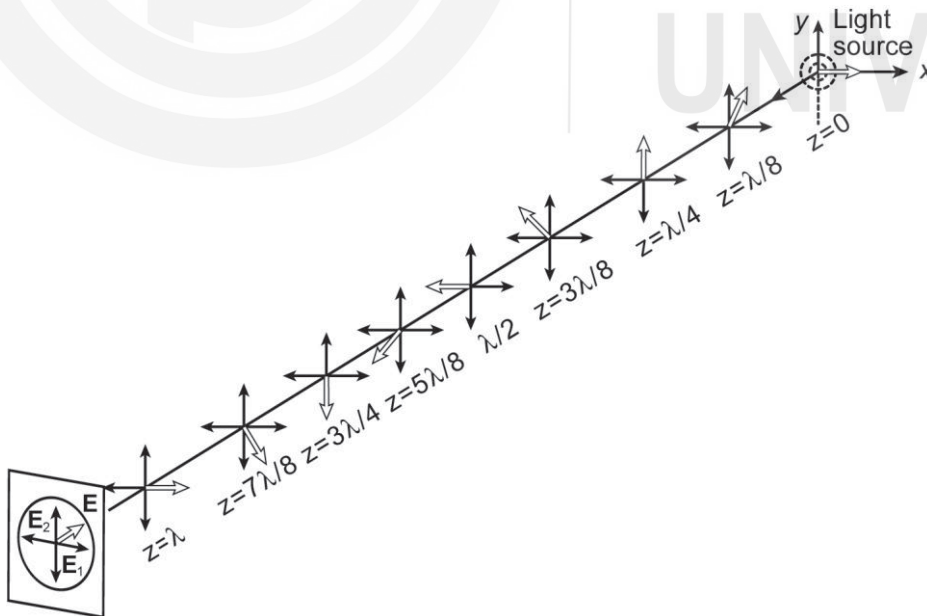


Fig. 5.5: Rotation of the electric field vector in a right-circularly polarised light wave.

For phase difference $\phi = 2n\pi + \frac{\pi}{2}$ with $n = 0, \pm 1, \pm 2, \dots$, Eq. (5.7) becomes

$$\vec{E}(z, t) = E_0 [\hat{x} \cos(kz - \omega t) - \hat{y} \sin(kz - \omega t)] \quad (5.9)$$

SAQ 2 - Electric field vector of resultant light wave

Using Eqs. (5.1) and (5.2) and $\phi = 2n\pi + \frac{\pi}{2}$; $n = 0, \pm 1, \pm 2, \dots$, derive Eq. (5.9).

Eq. (5.9) shows that the $\vec{E}$ vector rotates counter-clockwise in the reference frame. (Before proceeding further you should convince yourself about the validity of this statement by tabulating the values of $\vec{E}$ at $t = 0$ for different space points, and drawing a figure like Fig. 5.5) Such a light wave is called **left-circular polarised light** or **left-circular light**.

Can you now guess as to what will happen if two oppositely polarised circular waves of equal amplitude are superposed? To know the outcome of such superposition, you should add Eqs. (5.7) and (5.9). Doing that, you will obtain the expression for the electric field of the resultant light wave as given below:

$$\vec{E} = 2E_0 \hat{x} \cos(kz - \omega t) \quad (5.10)$$

Eq. (5.10) is similar to Eq. (5.1) which represents a linearly polarised light wave. Thus, we conclude that **superposition of two oppositely polarised circular light waves (of same amplitude) results in a linearly or plane polarised light wave**.

5.3.3 Elliptical Polarisation

Let us now consider the most general case by relaxing the conditions on amplitude and phase difference between the two superposing orthogonal linearly polarised light waves. Physically, we expect that beside its rotation, even the magnitude of resultant electric field vector will change. This means that the tip of the resultant $\vec{E}$ should trace out an ellipse in the reference plane as the wave propagates.

To analyse this mathematically, we write the scalar part of Eq. (5.2) in expanded form:

$$\frac{E_2}{E_{02}} = \cos(kz - \omega t) \cos \phi - \sin(kz - \omega t) \sin \phi$$

On substituting the value of $\cos(kz - \omega t)$ from Eq. (5.1), we get

$$\frac{E_2}{E_{02}} = \frac{E_1}{E_{01}} \cos \phi - \sin(kz - \omega t) \sin \phi$$

$$\text{or} \quad \frac{E_2}{E_{02}} - \frac{E_1}{E_{01}} \cos \phi = -\sin(kz - \omega t) \sin \phi \quad (5.11)$$

Further, the scalar part of Eq. (5.1) can be written as

$$\frac{E_1}{E_{01}} = \cos(kz - \omega t)$$

So, we can write

$$\sin(kz - \omega t) = \left[1 - (E_1/E_{01})^2 \right]^{1/2}$$

because $\sin x = \sqrt{1 - \cos^2 x}$

Thus, substituting the above expression for $\sin(kz - \omega t)$ in Eq. (5.11), we get

$$\frac{E_2}{E_{02}} - \frac{E_1}{E_{01}} \cos \phi = -\left[1 - (E_1/E_{01})^2\right]^{\frac{1}{2}} \sin \phi$$

Squaring both sides of the above expression and rearranging the terms, we get

$$\left(\frac{E_2}{E_{02}} - \frac{E_1}{E_{01}} \cos \phi\right)^2 = \left[1 - (E_1/E_{01})^2\right] \sin^2 \phi$$

$$\text{or} \quad \left(\frac{E_2}{E_{02}}\right)^2 + \left(\frac{E_1}{E_{01}}\right)^2 - 2\left(\frac{E_1}{E_{01}}\right)\left(\frac{E_2}{E_{02}}\right) \cos \phi = \sin^2 \phi \quad (5.12)$$

Eq. (5.12) defines an ellipse whose principal axis is inclined with the (E_1, E_2) coordinate system (Fig. 5.6). The angle of inclination, say α , is given by

$$\tan 2\alpha = \frac{2E_{01}E_{02} \cos \phi}{E_{01}^2 - E_{02}^2} \quad (5.13)$$

For $\alpha = 0$ or equivalently $\phi = \pm\pi/2, \pm3\pi/2, \dots$, Eq. (5.12) reduces to

$$\left(\frac{E_1}{E_{01}}\right)^2 + \left(\frac{E_2}{E_{02}}\right)^2 = 1 \quad (5.14)$$

Eq. (5.14) defines an ellipse whose principal axes are aligned with the coordinate axes. So, we discover that elliptical polarisation results from the superposition of two linearly polarised light waves of unequal amplitude and arbitrary phase difference between them. It is interesting to note that the linear and circular polarisations, discussed in the earlier sections, are special cases of elliptical polarisation. You should solve the following SAQ to show that it is indeed the case.

SAQ 3 - Linear and circular polarisation from elliptical polarisation

Starting with Eq. (5.12), show that linear and circular polarisation states are special cases of elliptical polarisation.

Before proceeding further, let us recall the important points discussed so far.

- **Polarisation** of light is due to its **transverse nature**.
- When the electric field vector associated with a light wave oscillates along a straight line in a plane perpendicular to the direction of propagation of light, it is said to be **linearly** or **plane polarised light**.
- When two linearly polarised light waves whose electric field vectors are mutually perpendicular are superposed, we obtain **linearly**, **circularly** and **elliptically** polarised light depending upon the phase difference between them and the relation between their amplitudes.
- Linear and circular polarisations are the **special cases** of elliptical polarisation.

Recap

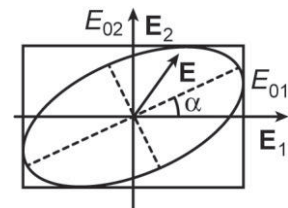


Fig. 5.6: Schematic diagram of elliptically polarised light wherein the tip of the resultant electric field vector traces an ellipse.

Now that you know what is polarised light and what are the different states of polarisation, the next logical step is to know the principles and techniques to

obtain polarised light. First, let us discuss the principles of producing linearly polarised light.

5.4 PRINCIPLES OF PRODUCING LINEARLY POLARISED LIGHT

The most important optical component in any polarised light producing arrangement is a **polariser**. Polariser changes the unpolarised natural light into polarised light. **An ideal linear polariser is one which reduces the intensity of an incident unpolarised light beam by exactly 50 percent.** A linear polariser produces linearly or plane polarised light.

When unpolarised light is incident on an ideal polariser, the outgoing light is in a define polarisation state (P - state) with an orientation parallel to the transmission axis of the polariser. That is, the polariser somehow discards all except one particular polarisation state. You would like to know: **How do we determine whether or not a device is a linear polariser?** The law which provides us necessary tool is Malus's law. Let us learn about it now.

5.4.1 Ideal Polariser: Malus's Law

Refer to Fig. 5.7. Unpolarised natural light indicated by $\vec{E}$ field in all direction in the reference plane R_1 , is incident on an ideal polariser, P whose transmission axis, shown by dotted line, makes an angle θ with the y -axis. For this arrangement, only the electric field parallel to the transmission axis of the polariser will be transmitted as shown in reference plane R_2 . This polarised light is incident on an identical ideal polariser, called analyser, A whose transmission axis is vertical; that is, along the y -axis. Suppose that there is no absorption of light in the material of either the polariser or analyser.

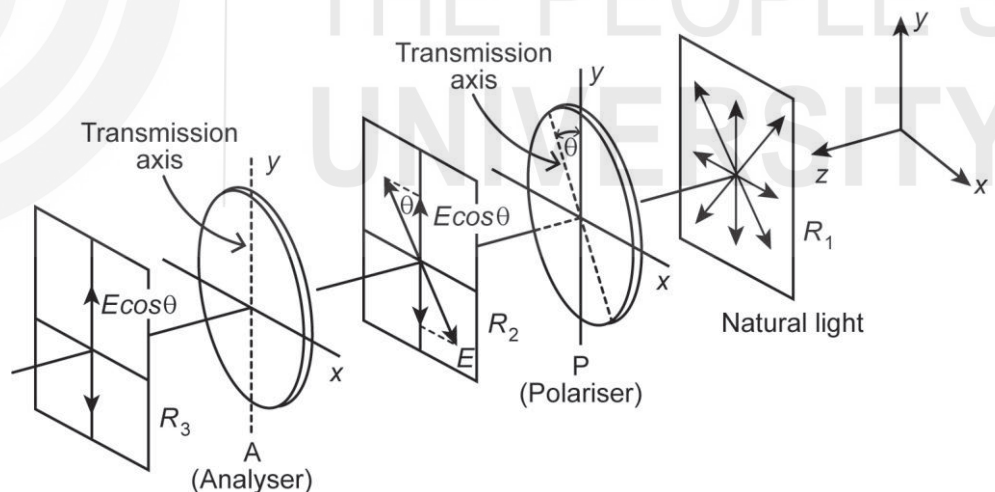


Fig. 5.7: A linear polariser.

Under these conditions, if E is the amplitude of the electric field $\vec{E}$ transmitted by the polariser, only its component $E \cos\theta$ parallel to the transmission axis of the analyser would pass through the analyser. Thus, the intensity of the polarised light reaching the detector (reference plane, R_3) is given by

$$I(\theta) = I(0)\cos^2 \theta \quad (5.15)$$

where θ is the angle between the transmission axes of the polariser, P and the analyser, A . The maximum intensity $I(0)$ occurs when the transmission axis of the polariser and the analyser are parallel.

Eq. (5.15) constitutes what is known as **Malus's law**. Malus law can be used to check whether an optical device is an ideal linear polariser or not. Study the following example.

EXAMPLE 5.2: IDEAL POLARISER

A beam of vertically polarised light is incident on a linear polariser. It is observed that the intensity of the light emerging out of the polariser is 25 percent of the incident intensity when the angle between the transmission axis of the polariser and the vertical is 60° . Is it an ideal polariser?

SOLUTION ■ Let us assume that the linear polariser is an ideal polariser. It should, therefore, obey Malus's law:

$$I(\theta) = I(0)\cos^2 \theta$$

As per the problem, the angle between the vertical (polarisation direction of the vertically polarised light) and the transmission axis of polariser is 60° .

So,

$$I(60^\circ) = I(0)\cos^2(60^\circ) = \frac{I(0)}{4}$$

So, the intensity of the emergent light is one-fourth or 25% of the intensity, $I(0)$ of the incident light as given in the problem. Thus, the linear polariser is an ideal polariser.

Now, before proceeding further, you may like to solve an SAQ.

SAQ 4 - Application of Malus's law

Unpolarised light is incident on two polarising sheets placed one over another. What must be the angle between their transmission axes if the intensity of light transmitted finally is one-third of the intensity of the incident light? Assume that each polarising sheet acts as an ideal polariser.

So far, we have discussed the criteria to decide whether a polariser is an ideal linear polariser or not. Polarised light can be produced using a variety of **physical mechanisms**: reflection, birefringence (or double refraction), scattering and dichroism (or selective absorption). You will now learn about some of these in detail.

5.4.2 Polarisation by Reflection: Brewster's Law

Reflection of light from a dielectric like plastic or glass is one of the most common methods of obtaining polarised light. You may have noticed the glare across a window pane or the sheen on the surface of a billiard ball or book jacket. It is due to reflection at the surface and the reflected light is partially polarised. However, to understand why reflected light is partially polarised, we will consider laboratory situations.

Suppose that an unpolarised light wave is incident on an interface between two different media at an angle θ_i as shown in Fig. 5.8a. Note that we have represented unpolarised light by the symbol $\longleftrightarrow$ which indicates two

The phenomenon of polarisation of light due to reflection was studied by Malus. One evening, he was examining a calcite crystal while standing at the window of his house in Luxembourg. The image of the Sun was reflected towards him from the windows of Luxembourg Palace. When he looked at the image through the calcite crystal, he was amused at disappearance of one of the double images as he rotated the crystal.

orthogonal components of the associated electric field. The dot (•) indicates electric field component perpendicular to the plane of incidence and arrow ($\leftrightarrow$) indicates electric field component parallel to the plane of incidence (that is, plane of the paper). This representation of unpolarised light is due to the fact that we can represent an incoming unpolarised light as made up of two orthogonal, equal amplitude linearly polarised light having their respective electric field vectors parallel and perpendicular to plane of incidence.

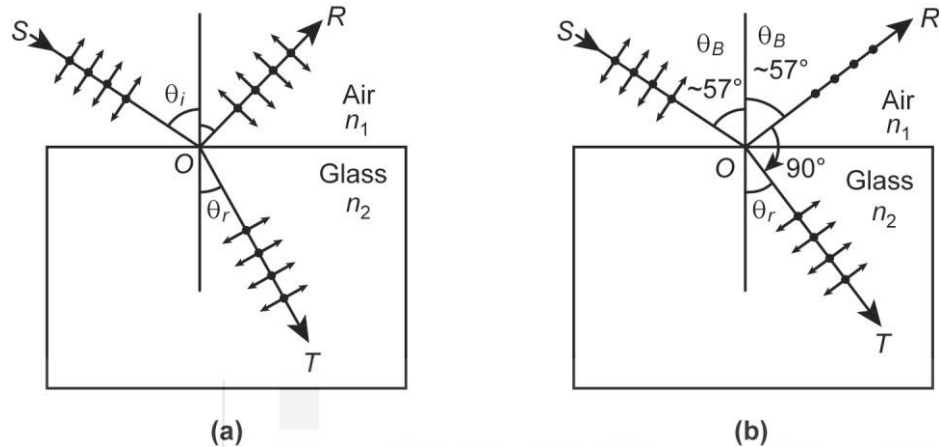


Fig. 5.8: a) Unpolarised light beam incident at angle θ_i at the interface of two optically different media; b) when the incidence angle is equal to Brewster's angle, the reflected light is linearly or plane polarised.

When the unpolarised light is incident at a particular angle, the reflected light is linearly polarised and the transmitted light is partially polarised as shown in Fig. 5.8b. This incidence angle, θ_B is called **Brewster angle**. The Brewster angle is also called **polarising angle**. It is also observed that when the angle of incidence is equal to Brewster angle, the reflected and transmitted rays are perpendicular to each other. This provides us with one of the most convenient methods for producing polarised light. These observations can be understood on the basis of Snell's law of refraction (Fig. 5.8b):

$$n_1 \sin \theta_B = n_2 \sin \theta_r$$

where n_1 and n_2 are the refractive indices of the media at whose interface light undergoes reflection. Since $\theta_r = \frac{\pi}{2} - \theta_B$, it readily follows that

$$n_1 \sin \theta_B = n_2 \cos \theta_B$$

$$\text{or} \quad \tan \theta_B = \frac{n_2}{n_1} \quad (5.16)$$

That is, the tangent of Brewster angle is equal to the ratio of the refractive indices of the media at whose interface incident light is reflected. When the incident beam is in air ($n_1 = 1$) and the transmitting medium is glass ($n_2 = 1.5$), the Brewster angle is nearly 56° . For air-water interface, like the surface of a pond or a lake θ_B is 53° . This means that when the sun is 37° above the horizontal, the light reflected by a calm pond or lake should be completely linearly polarised.

We, however, encounter some problems in utilizing this phenomenon to construct an effective polariser on account of two reasons:

- The reflected beam, although completely polarised, is weak.
- The transmitted beam, although strong, is only partially polarised.

These shortcomings are overcome by using reflection from a **pile of plates**. You can fabricate such a device with glass plates for the visible light, silver chloride plates for the infrared, and quartz plates for the ultraviolet region. It is an easy matter to construct a crude arrangement of this sort with a dozen or so microscope slides (Fig. 5.9).

Before proceeding further, you may like to solve an SAQ.

SAQ 5 - Polarisation by reflection

A plate of flint glass (refractive index 1.67) is immersed in water (refractive index 1.33). Calculate the Brewster angle for internal as well as external reflection at an interface.

Having studied how reflection of light can be used to produce polarised light, you may be tempted to know whether or not the phenomenon of refraction can also be used for the same? Refraction of light in isotropic crystals like NaCl or non-crystalline substances like glass, water or air does not lead to polarisation of light. However, there are crystalline substances such as **calcite** and **cellophane** which are optically anisotropic. That is, their optical properties are not the same in all directions. When light passes through such crystalline substances, we observe what is known as **double refraction** or **birefringence**. Let us now learn how the property of birefringence exhibited by anisotropic crystals is used to produce polarised light.

5.4.3 Polarisation by Double Refraction

To understand the meaning of double refraction or birefringence, mark a **black dot** on a piece of paper and observe it through a glass plate. You will see only one dot. Now, observe the dot through a calcite crystal. You will be surprised at the remarkable observation: instead of one, **two grey dots** appear, as shown in Fig. 5.10. Further, rotation of the crystal will cause one of the dots to remain stationary while the other appears to move along a circle about it. Similarly, if you place a calcite crystal on your book, you will see two images of each letter.

You may ask: Why does this happen? It is because the calcite crystal splits the incident light beam into two beams. This phenomenon of splitting of a light beam into two is known as **double refraction** or **birefringence**. Materials exhibiting this property are said to be **birefringent**. We bring you the excitement of Bartholinus, who discovered birefringence, in his words:

Greatly prized by all men is diamond, and many are the joys which similar treasures bring, such as precious stones and pearls ... but he, who, on the other hand, prefers the knowledge of unusual phenomena to these delights, he will, I hope, have no less joy in a new sort of body, namely, a transparent crystal, recently brought to us from Iceland, which perhaps is one of the greatest wonders that nature has produced. As my investigation of this crystal proceeded there showed itself a wonderful and extraordinary phenomenon: objects which are looked at through the crystal do not show, as in the case of other transparent bodies, a single refracted image, but they appear double.

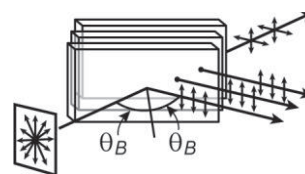


Fig. 5.9: Polarisation of light by a pile of plates.

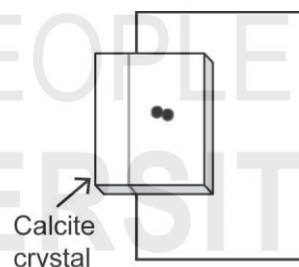


Fig. 5.10: Double refraction of a light beam by calcite crystal.

In some of the books, you may find that ordinary and extraordinary rays are being denoted by bold letters **O** and **E**. We have used small letters (*o*- and *e*-) to avoid confusion with the notation for the electric field.

Before we discuss polarisation of light due double refraction, let us first familiarise ourselves with some of the concepts related to this phenomenon.

The two refracted beams into which incident light splits have different angles of refraction. The distinguishing feature of these two refracted light beams is that one of these obeys Snell's law. It is called the **ordinary ray (o-ray)** in accordance with the nomenclature given by Bartholinus. The other beam does not obey Snell's law and is called the **extraordinary ray (e-ray)**. That is, a birefringent crystal displays two distinct indices of refraction.

Another important concept is that of **optic axis**, which signifies some special directions in a birefringent crystal along which two refractive indices are equal. This implies that both o- and e-rays in the direction of optic axis propagate with same velocity. You may now like to know: Does optic axis refer to any particular line through the crystal? No, **the optic axis refers to a direction**.

To understand the concept of optic axis, refer to Fig. 5.11 which shows a calcite crystal, also known as Iceland spar (CaCO_3). It is transparent to visible light and generally used in the shape of rhombohedra bounded by six parallelograms with angles 102° and 78° (Fig. 5.11). At corners A and H, the angle of each of the three faces is obtuse and these points are known as blunt corners. **A line passing through one of the blunt corners and is equally inclined to all the three edges meeting at that point gives the direction of optic axis.** Also, any line parallel to this line is also known as optic axis. This means that for any given point in the crystal, an optic axis may be drawn which will be parallel to that for any other point. For example, AA' and broken lines parallel to AA' shown in Fig. 5.11 are the optic axes for a calcite crystal.

When unpolarised light is incident perpendicular to these special directions called optic axis, both the o- and e-rays travel in the same direction with different velocities.

In addition to optic axis, two more geometrical concepts are useful in the context of double refraction by birefringent crystals like calcite. One is the concept of **principal section** which is a plane containing optic axis and perpendicular to the opposite faces of the crystal. Since there are six faces in the crystal, for every point there are three principal sections. A principal section always cuts the surface of calcite crystal in parallelogram with angles 109° and 71° . Another concept is that of **principal plane**. The principal plane for o-ray is a plane containing optic axis and the o-ray, and the principal plane for e-ray is the plane containing optic axis and e-ray.

Birefringent crystals which possess only one optic axis are called **uniaxial crystals**. And, crystals having two optic axes are called **biaxial crystals**. Calcite, quartz and ice are examples of uniaxial crystals and mica is a biaxial crystal. Most of the polarisation devices are made of uniaxial crystals.

Further, the uniaxial crystal for which the refractive index for o-ray (n_o) is more than the refractive index for the e-ray (n_e) is called **negative uniaxial crystal**. On the other hand, if $n_e > n_o$, we have a **positive uniaxial crystal**. The difference, $\Delta n = |n_e - n_o|$ is a measure of birefringence. The values of n_o and n_e for same uniaxial crystals are given in Table 5.1.

With this background knowledge of the geometrical concepts associated with birefringent crystals, we are now in a position to discuss how unpolarised light incident on a uniaxial crystal gets polarised? We know that when unpolarised

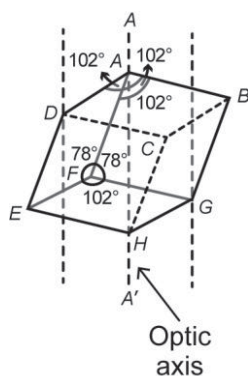


Fig. 5.11: Calcite crystal. The line AA' shows the direction of the optic axis.

Table 5.1: Refractive indices of some uniaxial birefringent crystals for light of wavelength 5893 Å

Crystal	n_o	n_e
Tourmaline	1.669	1.638
Calcite	1.6584	1.4864
Quartz	1.5443	1.5534
Sodium Nitrate	1.5854	1.3368
Ice	1.309	1.313

light beam enters a calcite crystal, it splits into the o- and e-rays. Now, refer to Fig. 5.12a which shows a calcite crystal. $ABCD$ is one of the principal sections of the crystal which contains the optic axis and is normal to the faces $BECF$ and $AHDG$. When unpolarised light is incident on the crystal, it splits into e-ray and o-ray: the electric field of e-ray vibrates in the plane containing the optic axis and the electric field of o-ray vibrates perpendicular to it, as shown in Fig. 5.12b. We may, therefore, conclude that due to double refraction, the unpolarised light beam splits into two components which are **plane polarised**. This is the basic mechanism of producing polarised light using a birefringent crystal.

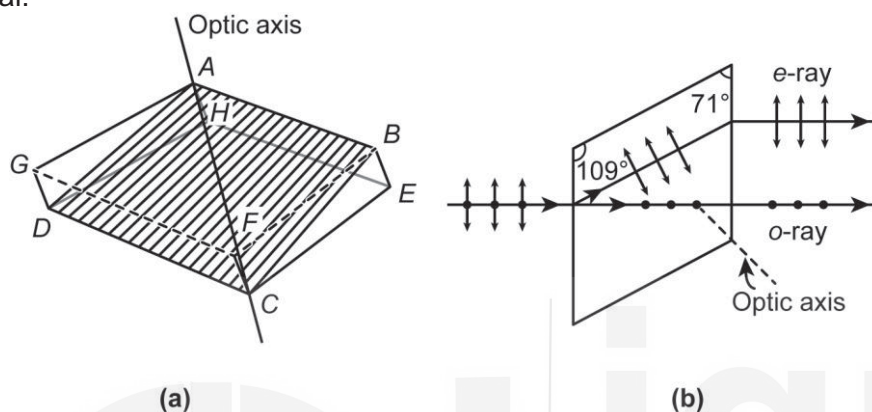


Fig. 5.12: a) $ABCD$ is one of the principal sections of the calcite crystal;
b) unpolarised light beam passing through a principal section of the calcite crystal.

Huygens explained many aspects of double refraction in calcite on the basis of wave theory. To appreciate Huygens explanation, refer to Fig. 5.13. Since the o-ray obeys Snell's law, it propagates with uniform velocity in all directions in the crystal. As a result, the wave surfaces corresponding to o-ray are spherical. However, the e-rays propagate with different velocities in different directions in the crystal and hence the resulting wave surface is an ellipsoid of revolution, i.e., a spheroid. Further, to reconcile with the fact that both the o- and e-rays travel with the same velocity along the optic axis, both the wave surfaces were assumed to touch each other at the two extremities of the optic axis. These features are depicted in Fig. 5.13. You may now like to know the nature of wave surfaces for o- and e-waves in a positive uniaxial crystals. This is subject matter of TQ 1.

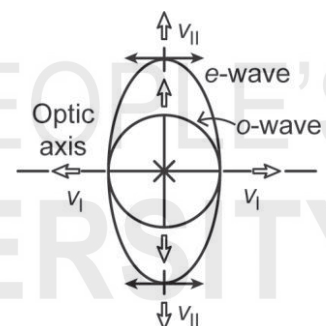


Fig. 5.13: o- and e-wave surfaces in negative uniaxial crystal (calcite).

Now, before proceeding further, let us recapitulate the important points discussed so far in this section.

- An **ideal polariser** reduces the intensity of unpolarised light by 50 percent. Malus's law provides a basis to determine whether a polariser is an ideal polariser or not.
- When a beam of unpolarised light is incident at **Brewster's angle** on a reflecting surface, the reflected beam is plane polarised.
- **Birefringence** is a phenomenon in which a beam of light is split into two beams. It is a property of the material and such materials are called birefringent materials.
- The o-rays obey **Snell's law** of refraction whereas e-ray do not.
- **Optic axis is a direction** in a birefringent crystal along which o- and e-rays propagate with same velocity.

Recap

From the above discussion it follows that, due to double refraction, an unpolarised light wave splits into *o*- and *e*-components with their $\vec{E}$ -vibrations perpendicular to each other. This splitting of unpolarised light into two linearly polarised light gives us an opportunity to produce linearly polarised light by selective absorption of one of the polarisation-states. This is readily done by a device, called Nicol prism, which removes the *o*-ray through total internal reflection. It was designed by William Nicol in 1828. You will learn about it now.

Nicol Prism

Nicol prism is made from a naturally occurring crystal of calcite. To understand its construction, refer to Fig. 5.14a which shows a crystal *ADEGFHBC* whose length is three times its breadth. *ABCD* is one of the principal sections with angle *ABC* equal to 71° . The end faces *AB* and *CD* are cut in such a way that they make angle 68° and 112° instead of 71° and 109° as shown in Fig. 5.14b. The resulting crystal is then cut into two pieces along a plane passing through the two blunt corners and perpendicular to the principal section. The cut surfaces are polished to optical flatness and then cemented together with a layer of a non-refrigent material such as Canada Balsam.

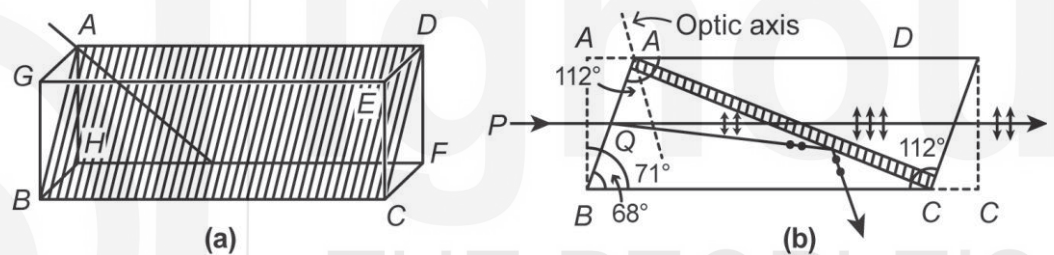


Fig. 5.14: a) Nicol prism; b) principal section *ABCD* of the Nicol prism.

You may ask: Why is Canada Balsam used as cementing material? Well, we will address this question in a short while.

To understand the working of Nicol prism, let a ray *PQ* of unpolarised light be incident on its face *AB*. This ray splits into *o*- and *e*-rays. Now, to make the light emerging from Nicol prism linearly polarised, the *o*-ray is removed by the process of total internal reflection.

When the *o*-ray is incident on the cut surface, it is travelling from optically denser calcite to rarer Canada Balsam medium. For sodium light, refractive index of Canada Balsam is 1.55, which is midway between refractive indices for *o*-ray ($n_o = 1.66$) and the *e*-ray ($n_e = 1.48$) in calcite. Thus, it is an optically rarer medium with respect to *o*-ray and denser for *e*-ray. Further, the critical angle for total internal reflection of *o*-ray is $\sin^{-1}(1.55/1.66) = 69^\circ$.

So, if the *o*-ray is incident on Canada Balsam surface at critical angle for total internal reflection, it will be removed from emergent ray. The angle of incidence on Canada Balsam depends on the angle which *AB* makes with blunt edge *BC* and also on the breadth and length ratio of the crystal. For this reason, the length of the crystal is chosen thrice its breadth and natural angle of 70° of the calcite crystal is reduced to 68° . The reduction in angle ensures that *o*-ray is incident on Canada Balsam at an angle greater than critical angle (69°) so that it is totally internally reflected. The emergent light will, therefore, be made up only of plane polarised *e*-ray.

Though Nicol prism is a useful device for obtaining plane polarised light, it has the following limitations:

1. It can be used for polarisation of visible light only.
2. e-ray also can get totally reflected by the Canada Balsam surface if it is travelling along the optic axis. Why? It is so because in this situation, the refractive index for e-ray will be same as for o-ray (i.e. greater than the refractive index for Canada Balsam).

With time, a number of modifications have been incorporated in the basic design of the Nicol prism to overcome some of these limitations. However, we will not go into these details.

So far you have studied about production of linearly polarised light by reflection and double refraction. Other methods employed to produce linearly polarised light are selective absorption (or dichroism) and scattering. We will here discuss only dichroism and that too in brief.

5.4.4 Selective Absorption: Dichroism

As you know, unpolarised light wave can be regarded as made up of two orthogonal, linearly polarised light waves. Many naturally occurring and man made materials have the property of selective absorption of one of these two (o- and e-) rays; the other passes through without much attenuation. This property of selective absorption is known as **dichroism**. Materials exhibiting this property are said to be **dichroic materials**. The net result of passing an unpolarised light through dichroic material is the production of linearly polarised light beam. A particularly simple dichroic device is the so-called Wire-Grid polariser. You will learn about it now.

The Wire-Grid Polariser

The wire-grid polariser consists of a grid of parallel conducting wires, as shown in Fig. 5.15. Suppose that unpolarised light is incident on the grid from the right. It can be considered as made up of two orthogonal P -states: P_x and P_y in the reference plane R_z . The y -component of the electric field drives the electrons of each wire and generates a current. It produces (Joule) heating of the wire. The net result is that energy is transferred from the field to the wire grid. In addition, electrons accelerating along the y -direction radiate in the forward as well as backward directions. The incident wave tends to be cancelled by the wave re-radiated in the backward direction. As a result, transmission of y -component of field is almost blocked.

However, the x -component of field is essentially unaltered as it propagates through the grid and the light coming out of the wire-grid is linearly polarised. The wire-grid polariser almost completely attenuates the light due to P_y polarisation state (or the y -component) when the spacing between the wires is less than or equal to the wavelength of the incident wave. You must realise that this restriction is rather stringent for the fabrication of a wire-grid polariser for visible light ($\lambda \sim 5 \times 10^{-7} \text{ m}$).

An easy way out of this difficulty in the fabrication of the grid polariser is to employ long chain polymer molecules made up of atoms which provide high electrical conductivity along the length of the chain. These chains of polymer

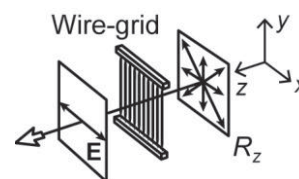


Fig. 5.15: The wire-grid polariser.

molecules behave similar to the wires in the wire-grid polariser. The alignment of these chains are almost parallel to each other. Because of high electrical conductivity, the electric vector of unpolarised light parallel to the chain gets absorbed. And the *P*-state perpendicular to these chains passes through. These chemically synthesized polarisers are fabricated in the form of plastic sheets and are known as **polaroids**. Since the spacing between these molecular chains in a polaroid is small compared to the optical wavelength, such polaroids are extremely effective in producing linearly polarised light.

Dichroic Crystals

Some naturally occurring crystalline materials are inherently dichroic due to anisotropy in their structure. One of the best known dichroic materials is **tourmaline**, a precious stone often used in jewellery. Tourmalines are essentially boron silicates of differing chemical composition. The component of $\vec{E}$ field perpendicular to the principal axis is strongly absorbed by the sample. Thicker the crystal, more complete will be the absorption. A plate cut from a tourmaline crystal parallel to its optic axis acts as a linear polariser. This is illustrated in Fig. 5.16.

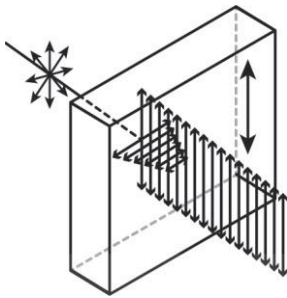


Fig. 5.16: Tourmaline crystal polariser.

We shall now consider a class of optical elements known as wave plates which serve to change the polarisation state of the incident light wave. A wave plate introduces a phase lag between the two *P*-states by a predetermined amount. That is, the relative phase difference between the two emerging components is different from its initial value. This concept can be used to convert a given polarisation state into any other and in so doing it is possible even to produce circular or elliptic polarisation as well. This is the subject matter of the next section.

5.5 WAVE PLATES: CIRCULAR AND ELLIPTICAL POLARISERS

Consider a plane wave incident on a calcite crystal. It splits into *o*- and *e*-waves. Since calcite is a negative uniaxial crystal, $n_o > n_e$ and $v_{\parallel}$ (velocity of *e*-wave) $>$ $v_{\perp}$ (velocity of *o*-wave) implying that the *e*-ray travels faster than the *o*-ray. After traversing the calcite crystal of thickness d , the path difference between them is given by

$$\Delta = d(n_o - n_e)$$

and the relative phase difference between *o*- and *e*-rays is

$$\delta = \frac{2\pi}{\lambda} \Delta = \frac{2\pi}{\lambda} (n_o - n_e) d \quad (5.18)$$

Though, while entering the crystal, both the components were in phase.

The **state of polarisation** of emerging light depends on phase difference δ , apart from the amplitudes of incoming orthogonal field components. Let us now consider some specific cases:

- When the phase difference, $\delta = 2m\pi$; where m is an integer, the relative path difference is $m\lambda$. A device which induces a path difference between the two orthogonal field vibrations is integral multiples of λ is called the **full-wave plate**. It introduces no observable effect on the polarisation of the incident beam. That is, the field vibrations of the emergent light will be identical with the field vibrations of the incident light.

In case of positive uniaxial crystal, $n_e > n_o$ and hence the path difference will be $d(n_e - n_o)$. In fact the general expression for the path difference is $d|n_o - n_e|$.

ii) When $\delta = (2m + 1)\pi$, the relative path difference will be $\left(m + \frac{1}{2}\right)\lambda$. Such

crystals are called **half-wave plates**.

iii) When $\delta = (2m + 1)\frac{\pi}{2}$, the relative path difference will be $\left(m + \frac{1}{2}\right)\frac{\lambda}{2}$. Such

a birefringent sheet is called **quarter-wave plate**.

When linearly polarised light traverses a quarter-wave plate, the emergent light will, in general, be elliptical.

Thus, we may conclude that the **path difference between the o- and e-waves in a birefringent device depends on its thickness**.

You should now solve the following SAQ.

SAQ 6 - Thickness of a wave-plate

Calculate the thickness of a quarter wave-plate for light of wavelength 5890 Å. The refractive indices for o- and e-rays are 1.55 and 1.54 respectively.

We now summarise what you have learnt in this unit.

5.6 SUMMARY

Concept	Description
Polarised light	Visible light can be linearly, circularly or elliptically polarised. All these polarisation states arise due to superposition of two linearly (or plane) polarised light waves characterised by different amplitudes and phases.
States of polarisation	The electric field vectors of two linearly polarised light beams propagating along z-axis can be represented as $\vec{E}_1(z, t) = \hat{x} E_{01} \cos(kz - \omega t)$ $\vec{E}_2(z, t) = \hat{y} E_{02} \cos(kz - \omega t + \phi)$ where E_{01} and E_{02} are the amplitudes of the two waves and ϕ is the phase difference between them. Superposition of these two polarised waves will result in Linearly polarised light if $\phi = 0$ or an integral multiple of $\pm 2\pi$ Circularly polarised light if $\phi = \pi/2$ and $E_{01} = E_{02}$ Elliptically polarised light if $\phi = \pi/2$ and $E_{01} \neq E_{02}$
Ideal linear polariser	According to Malus's law, when the transmission axes of polariser and the analyser are at an angle θ , the intensity of the polarised light reaching the detector is given by $I(\theta) = I(0)\cos^2 \theta$ where, $I(0)$ is the intensity of the polarised light when $\theta = 0$.
Brewster's law	When natural light strikes an interface at Brewster's angle, $\theta_B = \tan^{-1}(n_2 / n_1)$, where n_1 and n_2 are the refractive indices of medium of incidence and transmission, the reflected light is linearly polarised. Brewster's angle is also called polarising angle.

- Birefringence** ■ When light is incident on a calcite crystal, it splits into two. The phenomenon is known as double refraction or birefringence. These two refracted beams are known as *o*- and *e*-rays. Snell's law holds for *o*-rays (ordinary rays) and does not hold for *e*-ray (extraordinary ray).
- Optic axis** ■ In a birefringent material, the *o*- and the *e*-rays travel in the same direction with same velocity along the optic axis. However, in a direction perpendicular to the optic axis, they travel with different velocities. The electric field vibrations for *o*- and the *e*-rays are mutually perpendicular.
- Nicol prism** ■ The phenomenon of double refraction is used for producing linearly polarised light. Nicol prism works on this principle. In the Nicol prism, the *o*-ray undergoes total internal reflection at the interface and the transmitted beam consists of only electric field vibrations corresponding to *e*-ray and hence the transmitted beam is linearly polarised.
- Dichroism** ■ Selective absorption (or dichroism) of the electric field component with particular orientations by a dichroic material can also be used for producing linearly polarised light. Tourmaline is an example of dichroic material.
- Wave-plates** ■ For a calcite crystal of thickness d , the path difference between *o*- and *e*-rays is given by $\Delta = d|n_o - n_e|$.

The corresponding phase difference is

$$\delta = \frac{2\pi}{\lambda} \Delta = \frac{2\pi}{\lambda} d(|n_o - n_e|)$$

When the phase difference $\delta = 2m\pi$ where m is an integer, the relative path difference between the *o*- and *e*-rays will be $m\lambda$. Such crystals are called full-wave plate. When $\delta = (2m + 1)\pi$, path difference will be $\lambda/2$ and such crystals act as half-wave plate. And when $\delta = (2m + 1)\frac{\lambda}{2}$, path difference will be $\lambda/4$ (for $m = 0$) and such crystals are called quarter-wave plate.

5.7 TERMINAL QUESTIONS

1. In sub-section 5.4.3, you studied about propagation of *o*- and *e*-waves in a negative uniaxial crystal (calcite). Draw a diagram and describe the propagation of *o*- and *e*-waves in a positive uniaxial crystal (quartz) for normal incidence.
2. A beam of light is incident on a polariser-analyser set up. When the transmission axes of the polariser and analyser are parallel, the intensity of light emerging from the analyser is 0.6 Wm^{-2} . What will the intensity be when the angle between the transmission axes of the polariser and analyser is 45° ?
3. A block of glass ($n = 1.5$) is immersed in water ($n = 1.3$). a) Calculate the polarising angle for a beam of light incident from glass onto the glass-water surface. b) Calculate the polarising angle for a beam of light incident from the water onto the glass-water surface.
4. For a certain crystal, $n_o = 1.54$ and $n_e = 1.55$ for light of wavelength $6 \times 10^{-7} \text{ m}$. Calculate the least thickness of a quarter-wave plate made from the crystal for use with light of this wavelength.

5.8 SOLUTIONS AND ANSWERS

Self-Assessment Questions

1. The plane of vibration of the electric vector defined by Eq. (5.5) is rotated with respect to that shown in the Fig. 5.4. This is signified by the negative sign before $\hat{y}$ in the parentheses and is depicted in Fig. 5.17.
2. The electric fields of two orthogonal, linearly polarised light waves of equal amplitudes propagating along z-direction is

$$\vec{E}_1(z, t) = \hat{x} E_0 \cos(kz - \omega t)$$

$$\vec{E}_2(z, t) = \hat{y} E_0 \cos(kz - \omega t + \phi)$$

When they superpose, the electric field of the resultant light wave is

$$\begin{aligned} \vec{E}(z, t) &= \vec{E}_1(z, t) + \vec{E}_2(z, t) = E_0 \left[\hat{x} \cos(kz - \omega t) + \hat{y} \cos(kz - \omega t + \frac{\pi}{2}) \right] \\ &= E_0 [\hat{x} \cos(kz - \omega t) - \hat{y} \sin(kz - \omega t)] \end{aligned}$$

3. We know from Eq. (5.12) that

$$\left(\frac{E_2}{E_{02}} \right)^2 + \left(\frac{E_1}{E_{01}} \right)^2 - 2 \left(\frac{E_2}{E_{02}} \right) \left(\frac{E_1}{E_{01}} \right) \cos \phi = \sin^2 \phi \quad (i)$$

If we choose $\phi = 0$ in (i), we get

$$\left(\frac{E_2}{E_{02}} \right)^2 + \left(\frac{E_1}{E_{01}} \right)^2 - 2 \left(\frac{E_2}{E_{02}} \right) \left(\frac{E_1}{E_{01}} \right) = 0$$

which can be written in a compact form:

$$\left(\frac{E_2}{E_{02}} - \frac{E_1}{E_{01}} \right)^2 = 0$$

$$\text{or} \quad E_2 = \frac{E_{02}}{E_{01}} E_1 \quad (ii)$$

This defines a straight line ($y = mx$) with slope E_{02} / E_{01} . In other words, elliptically polarised light reduces to linearly polarised light for $\phi = n\pi (n = 0, \pm 1, \pm 2, \dots)$.

When $\phi = \pi/2$ and $E_{01} = E_{02} = E_0$, Eq. (5.12) reduces to

$$\left(\frac{E_2}{E_0} \right)^2 + \left(\frac{E_1}{E_0} \right)^2 = 1$$

which defines a circle ($x^2 + y^2 = a^2$) of radius E_0 .

4. Since both polarising sheets are ideal, the intensity of the incident unpolarised beam, I , will reduce to half after passing through one of them as shown in the Fig. 5.18. After passing through the second polarising sheet, we are told that the intensity reduces to one third of original value.

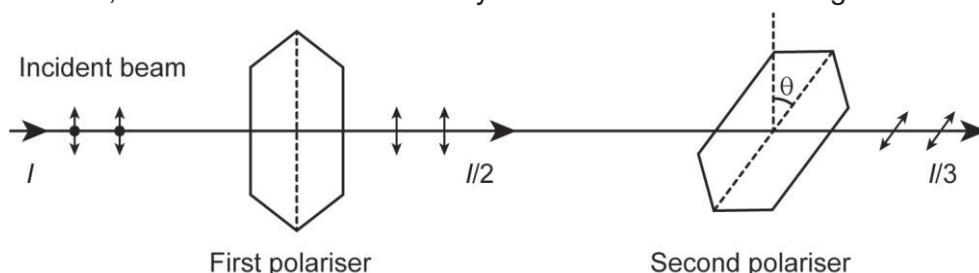


Fig. 5.18: Unpolarised light beam of intensity I passing through two polarisers.

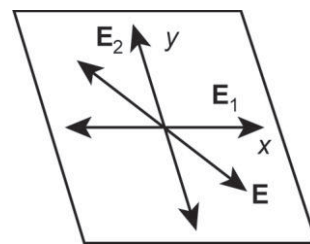


Fig. 5.17

From Malus's law we know that

$$I(\theta) = I(0) \cos^2 \theta$$

Here $I(\theta) = I/3$ and $I(0) = 1/2$. Therefore

$$\cos^2 \theta = (2/3) = 0.666$$

$$\text{or } \theta = \cos^{-1}(.666)^{1/2} = 35.3^\circ$$

That is, the angle between the transmission axes of two polarisers is about 35° .

5. For external reflection

$$\tan \theta_B = \frac{n_2}{n_1} = \frac{1.67}{1.33}$$

$$\Rightarrow \theta_B = \tan^{-1}\left(\frac{1.67}{1.33}\right) \Rightarrow \theta_B = 51.47^\circ$$

For internal reflection

$$\tan \theta_B = \frac{n_2}{n_1} = \frac{1.33}{1.67}$$

$$\theta_B = 38.53^\circ$$

6. The path difference produced between the o- and e-rays of a birefringent crystal of thickness d is $\Delta = d(|n_o - n_e|)$. And, corresponding relative phase difference is given by

$$\delta = \frac{2\pi}{\lambda} \Delta = \frac{2\pi}{\lambda} d(|n_o - n_e|)$$

The phase difference produced by a quarter-wave plate, $\delta = \pi/2$

On comparing the above expressions for the phase difference, we have

$$d = \frac{\lambda}{4(n_o - n_e)} = \frac{5890 \text{ \AA}}{4 \times (1.55 - 1.54)} = \frac{5.89 \times 10^{-7} \text{ m}}{0.0432} = 1.36 \times 10^{-5} \text{ m}$$

Terminal Questions

1. In case of negative uniaxial crystal (calcite), e-ray travels faster than the o-ray and hence $n_o > n_e$. Therefore, when a light beam is incident normally upon a calcite crystal, whose optic axis is parallel to the refracting surface and lies in the plane of incidence, o-wave has a spherical wavefront and the e-wave has an spheroidal wavefront.

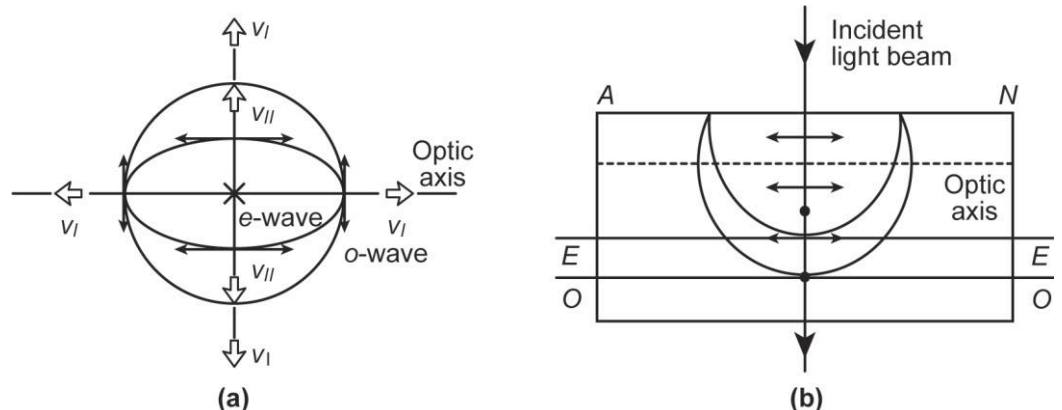


Fig. 5.19: a) o- and e-wave surfaces in a positive uniaxial crystal (quartz); b) propagation of o- and e-waves in quartz.

In case of positive uniaxial crystals like quartz, the *e*-ray travels slower than the *o*-ray. Therefore, the spherical wavefront corresponding to *o*-ray will be outside the spheroidal wavefront corresponding to *e*-ray (Fig. 5.19a). Since the optical properties of birefringent crystal are symmetrical with respect to its optic axis, the axis of revolution of the spheroid must coincide with the optic axis of the crystal. When a light beam falls on a positive uniaxial crystal, with its optic axis in the plane of incidence and parallel to the refracting surface, the wavefront for *o*- and *e*-waves are shown in Fig. 5.19b.

In the above mentioned case, *EE* and *OO* are the refracted wavefronts for *e*- and *o*-rays, respectively, at the same instant of time. They are parallel to each other and travel in the same direction which is perpendicular to the refracting surface *AN*. These two wavefronts, however, will travel with different velocities. As a result, a path difference will be introduced between the *o*- and the *e*-rays on emergence, but there is no separation between the two beams. In principle, we can construct quarter-wave plate, half-wave plate etc. using positive uniaxial crystal as well.

2. According to Malus's law, the intensity of light passing through a polariser-analyser set up is given as

$$I(\theta) = I(0) \cos^2 \theta$$

where θ is the angle between the transmission axes of the polariser and analyser. As per the problem, when $\theta = 0$, intensity is 0.65 Wm^{-2} . So, $I(0) = 0.65 \text{ Wm}^{-2}$.

When the angle $\theta = 45^\circ$, we have

$$I(45^\circ) = I(0) \cos^2 \theta = (0.65 \text{ Wm}^{-2}) \times \cos^2(45^\circ) = 0.3 \text{ Wm}^{-2}$$

3. Polarising angle, θ_B is given as

$$\tan \theta_B = \frac{n_2}{n_1}$$

- a) For glass to water, $n_1 = 1.5$ and $n_2 = 1.3$.

$$\text{So, } \theta_B = \tan^{-1}(1.3/1.5) = 40.6^\circ$$

- b) For water to glass, $n_1 = 1.3$ and $n_2 = 1.5$.

$$\text{So, } \theta_B = \tan^{-1}(1.5/1.3) = 48.9^\circ$$

4. In a birefringent crystal of thickness d , the path difference between the *o*- and *e*-rays is $d|n_e - n_o|$. As per the problem, $n_e > n_o$. So we can write $\Delta = d(n_e - n_o)$ and the corresponding phase difference between the *o*- and *e*-rays is

$$\delta = \frac{2\pi}{\lambda} d(n_e - n_o)$$

Since a quarter wave-plate gives rise to a phase difference of $\pi/2$, we can write

$$\frac{\pi}{2} = \frac{2\pi}{\lambda} d(n_e - n_o) \Rightarrow d = \frac{\lambda}{4(n_e - n_o)}$$

We have

$$n_o = 1.54, n_e = 1.55 \text{ and } \lambda = 6 \times 10^{-7} \text{ m.}$$

Hence,

$$d = \frac{6 \times 10^{-7} \text{ m}}{4 \times (1.55 - 1.54)} = \frac{6 \times 10^{-7} \text{ m}}{0.0364} = 1.65 \times 10^{-5} \text{ m.}$$

That is, the quarter-wave plate should be $1.65 \times 10^{-5} \text{ m}$ thick.

